DOCTORADO EN PLANEACIÓN ESTRATÉGICA Y DIRECCIÓN DE TECNOLOGÍA

Universidad Popular Autónoma del Estado de Puebla
Centro Interdisciplinario de Posgrados
Investigación y Consultoría
Departamento de Ingeniería
Doctorado en Planeación Estratégica y Dirección
de Tecnología

Análisis de Información Aplicando Tecnología de Aprendizaje Automático como Soporte en la Toma de Decisiones. Una Ventaja Competitiva para las IES: Caso UPPuebla

Tesis que para obtener el Grado de Doctor
en Planeación Estratégica y Dirección de Tecnología

Presenta

M.C. Argelia Berenice Urbina Nájera

Puebla, México Febrero, 2015

Universidad Popular Autónoma del Estado de Puebla
Centro Interdisciplinario de Posgrados
Investigación y Consultoría
Departamento de Ingeniería
Doctorado en Planeación Estratégica
y Dirección de Tecnología

Se aprueba la Tesis:

Análisis de información aplicando tecnología de aprendizaje automático como soporte en la toma de decisiones. Una ventaja competitiva para las IES: Caso UPPuebla

MC. Argelia Berenice Urbina Nájera

Comité Asesor

Dr. Jorge de la Calleja Mora
Director de Tesis

Dr. Carlos A. Vega Lebrún
Asesor

Dra. Beatriz Pico González
Asesora

Puebla, México.

30 de Enero de 2015



Tesis para obtener el grado de Doctor en
Planeación Estratégica y Dirección de Tecnología

Análisis de Información Aplicando Tecnología de Aprendizaje Automático como Soporte en la Toma de Decisiones. Una Ventaja Competitiva para las IES: Caso UPPuebla

Presenta: **Argelia Berenice Urbina Nájera**



Febrero, 2015

ÍNDICE

CAPÍTULO III: METOTOLOGÍA

CAPÍTULO IV: MÉTODOS PARA MEJORAR LA TOMA DE DECISIONES

CAPÍTULO V: ANÁLISIS DE LOS RESULTADOS

CONCLUSIONES Y TRABAJO FUTURO

ANEXOS

Resumen

El objetivo general de este trabajo de investigación es analizar información para mejorar la toma de decisiones y crear una ventaja competitiva en las Instituciones Públicas de Educación Superior (IPES) a través del desarrollo de métodos computacionales empleando algoritmos de aprendizaje automático, a fin de favorecer la inteligencia de negocios en las organizaciones. Para la recolección de datos se diseñó y validó un instrumento llamado H-A, que permite evaluar en escala Likert las competencias de comunicación, autodirección, interpersonal, intrapersonal, digital, autonomía y afinidades personales de líderes y colaboradores. En el procesamiento de los datos se empleó la herramienta Weka; en la cual se utilizaron tres algoritmos. El algoritmo *k-medias* se aplicó para formar grupos de trabajo basados en sus competencias. Los algoritmos selección de atributos, a través de evaluadores y métodos de búsqueda, para identificar las competencias más importantes de un grupo de individuos. El algoritmo árboles de decisión para caracterizar a líderes y colaboradores a través del árbol obtenido y encontrar la característica más importante.

La experimentación con el algoritmo *k-medias* se realizó mediante dos enfoques. Por un lado, considerando el lado didáctico-pedagógico y por el otro el lado, un enfoque empresarial. Para el primer enfoque, el conjunto de datos se basó en información obtenida de profesores y estudiantes de la Universidad Politécnica de Puebla; mientras que para el segundo enfoque, se consideraron a las primeras diez empresas más importantes del País. En ambos enfoques se usó la distancia Euclídea y Manhattan. La propuesta derivada de esta experimentación radica en un proceso para integrar individuos a equipos de trabajo "similares" a fin de lograr con esto que el equipo sea eficiente y por consiguiente, lograr mayor productividad en la empresa; en este sentido, todas las competencias fueron importantes, así como las afinidades personales. Al mismo tiempo, este algoritmo permitió

agrupar a líderes basadas en sus competencias y afinidades personales; al identificarlas de acuerdo a la puntuación obtenida en la escala Likert, en donde sobresalen las competencias: interpersonal, autodirección, digital y autonomía; dejando a un lado las afinidades personales.

Con los algoritmos selección de atributos, se aplicaron cinco evaluadores y dos métodos de búsqueda. Tras la experimentación se pudieron reconocer las competencias más importantes que deben clasificar al capital humano en líder o colaborador, según sus competencias y afinidades. Éstos algoritmos, en contraposición con el algoritmo árboles de decisión, sí consideraron importante las afinidades personales entre cada individuo listándola como la característica principal, seguida de las competencias: autodirección, interpersonal, intrapersonal, autonomía, digital y comunicación.

El algoritmo árboles de decisión, permitió caracterizar al capital humano, particularmente en líder o colaborador en función de sus competencias y afinidades; así como encontrar la característica más importante entre ellos. No obstante, el árbol generado solamente consideró que un líder debe tener puntajes superiores a 4 en las competencias: comunicación (la más importante), autodirección y habilidad digital; mientras que las características para un colaborador es que su puntaje sea menor a 4 en las mismas competencias que para el líder.

Finalmente, la aplicación de la tecnología de aprendizaje automático genera inteligencia de negocios al aplicar herramientas tecnológicas que contribuyan a analizar, procesar y depurar información de manera rápida, sencilla y confiable. En este sentido, la aplicación de los algoritmos k-medias, selección de atributos y árboles de decisión, permiten tomar decisiones sin sesgo; pues ofrecen de manera gráfica la información procesada que facilita el proceso, particularmente relacionado con el capital humano.

Abstract

The overall objective of this research is to analyze information to improve decision making and create a competitive advantage in the Public Institutions of Higher Education (PIHE) through the development of computational methods using machine learning algorithms in order to promote business intelligence in organizations. To data collection was designed and validated an instrument called HA, which allows Likert scale assess communication skills, self-direction, interpersonal, intrapersonal, digital, autonomy and personal affinities to leaders and collaborators. In the data processing it used the Weka tool; where three algorithms are used. The k-means algorithm was applied to form working groups based on their competencies. The attribute selection algorithm through evaluators and search methods to identify the most important skills of a group of individuals and finally, the decision tree algorithm to characterize leaders and employees through the tree obtained. Experiments with k-means algorithm are realized by two approaches.

On one hand, considering the didactic-pedagogical side and on the other side, business approach. For the first approach, the data set was based on information from teachers and students of the Polytechnic University of Puebla; while for the second approach, they were considered the top ten largest companies in the country. In both approaches the Manhattan and Euclidean distance was used. The proposal resulting from this experiment is a process to integrate individuals in work teams "similar" to do the team efficient and therefore achieve greater productivity in the company; in this sense, all skills were important including the personal affinities. At the same time, this algorithm allowed group leaders based on their skills and personal affinities; to identify them according to their score on the Likert scale, with outstanding skills: interpersonal, self-direction, digital and autonomy; leaving aside personal affinities.

With attribute selection algorithm five evaluators and two search methods were applied. After experimentation could recognize the key skills required to classify the human capital in leader or partner, according to their skills and affinities. This algorithm, in contrast to the decision tree algorithm, itself considered important personal affinities between each individual listed as the principal, followed by skills: self-direction, interpersonal, intrapersonal, autonomy, and digital media; according to their order of importance....

The decision tree algorithm was used to characterize human capital, particularly in leader or partner depending on their skills and affinities. However, the tree generated only considered that a leader must have skills scores above four: communication (the most important), self-direction and digital skills; while features for a partner is that your score is less than 4 in the same powers to the leader.

Finally, the application of machine learning technology creates business intelligence to apply technological tools that help to analyze process and debug information quickly, easily and reliably. In this sense, the application of k-means, feature selection and decision trees, algorithms allow taking decisions without bias; they offer graphically processed information that facilitates the process, particularly related to human capital.

Capítulo I
Introducción

INTRODUCCIÓN

Este capítulo tiene como objetivo presentar los antecedentes que dan origen a esta investigación. A partir de ello, se describe la problemática, el objetivo general y objetivos específicos, así como la justificación, los alcances e impacto que se espera tras la realización de este trabajo de investigación.

1.1. Antecedentes

Actualmente, se vive una época de cambios vertiginosos y complejos sin precedentes, en donde la capacidad de aprendizaje y análisis de los individuos y de las organizaciones se ha convertido en un tema cada vez más significativo y crítico.

Debido a los grandes volúmenes de información que se generan diariamente y a su movilidad, se gestan problemas para almacenarla y procesarla. A pesar de todos los avances tecnológicos, aún persisten incógnitas sobre qué se hace con toda esta información, quién y cómo se analiza.

Aunado a esto, existe un déficit de talento humano con habilidades analíticas que favorezcan a que los directivos tomen decisiones oportunamente fundamentadas en este análisis de información y no en su buen juicio.

Sprenger (2010) y Arias Delgado (2011) resumen las características excepcionales del cerebro humano en: *la capacidad para detectar patrones y efectuar aproximaciones, la capacidad de varios tipos de memoria, la capacidad de autocorregirse y aprender desde la experiencia por medio del análisis de datos externos y autoreflexión y finalmente, una capacidad infinita de crear.* A pesar de contar con estas maravillosas habilidades, la mente humana es imperfecta y tiene sesgos subconsientes.

La toma de decisiones es un proceso complejo y de múltiples dimensiones, que no puede ser restringido a un único ámbito, en un solo tiempo ni ser generado por un único actor; en otras palabras, consiste en seleccionar una opción, de una serie de alternativas disponibles, a fin de resolver un determinado problema o bien, para enfrentar un problema potencial que se puede presentar en un determinado momento (Dirección General de Desarrollo de la Gestión e Innovación Educativa-SEP, 2013) y (Hellriegel & Slocum, 2009).

Es por ello, que reconociendo los límites de la mente humana muchos directivos se cuestionan ¿Cómo encontrar, analizar y evaluar más posibilidades en menos tiempo? ¿Cómo hacer elecciones más asertivas en los negocios? Es aquí cuando toman importancia las nuevas posibilidades directivas de la inteligencia de negocios, cuyo valor es complementar la intuición incidiendo de manera objetiva en la mejora de la toma de decisiones y en la mejora del desempeño de los directivos y del personal clave en las empresas (Torres Pérez, Sánchez García, & Ramírez Gutiérrez, 2014).

En términos prácticos, la inteligencia de negocios (BI por sus siglas en inglés *Business Intelligence*) "*es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda mejorar el proceso de toma de decisiones en los negocios*" (Sinnexus, 2012). Incluye a los procesos que buscan proveer información para facilitar la toma de decisiones, con base en el uso de los datos generados (o adquiridos).

En este sentido, la minería de datos contribuye a generar esta inteligencia de negocios al explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto (Witten & Frank, 2005). Básicamente, ayuda a comprender el contenido de un conjunto de datos y para este fin, usa prácticas estadísticas y emplea algoritmos computacionales, particularmente de la inteligencia artificial.

Es así, como la inteligencia de negocios actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva; proporcionando información privilegiada para responder a los problemas de negocio como: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control

financiero, optimización de costos, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, entre otros (Sinnexus, 2012).

1.2. Problemática

En la actualidad, la sociedad genera grandes cantidades de información continuamente; por ejemplo cuando se realiza una compra, cuando se hace una reservación de hotel o avión, cuando se registra la entrada al trabajo, cuando se hace algún trámite gubernamental o educativo, o bien, cuando se paga algún servicio público. Produciendo con todo ello, un conjunto de datos enorme que permite controlar, mejorar, administrar, examinar, investigar, planificar, predecir, someter, negociar o tomar decisiones de cualquier ámbito según el dominio de interés (Ferrell, Hirt, & Ferrel, 2010).

Generalmente el análisis de este cúmulo de información se realiza de forma manual, o bien, empleando técnicas como: lluvia de ideas, decisión por consenso, negociación colectiva, modelo de preferencias subjetivas, entre otros(Bensoussan & Fleisher, 2013); lo que deriva en un proceso tedioso, sesgado y que puede consumir mucho tiempo.

Además de que en la actualidad, a pesar de contar con tecnología de vanguardia los directivos siguen preguntándose: ¿Cómo encontrar, analizar y evaluar más posibilidades en menos tiempo? ¿Cómo hacer elecciones más asertivas en los negocios?

Se identifican tecnologías que favorecen el análisis, tratamiento, transmisión y manejo de manera flexible de estas grandes cantidades de información, destinadas a: 1) Potenciar el procesamiento que supone el cálculo y tratamiento automático de datos según reglas preestablecidas, 2) Transmitir datos y/o imágenes que posibiliten a un número mayor de individuos o máquinas disponer inmediatamente de una enorme cantidad de información, 3) Control de flujos de información que radica en conectar numerosos dispositivos de procesamiento de información y finalmente, 4) Manipulación de conocimientos que corresponda a los sistemas que puedan aprender de la experiencia (Bustillo Porro, 2009).

Una de estas tecnologías es el aprendizaje automático (disciplina que integra a la inteligencia artificial) que se ha aplicado en la solución de problemas complejos en diversas áreas como: **Medicina**(Cruz & Wishant, 2007), (Kadhim A-Shayea, 2011), (Er, Yumusak, & Temurtas, 2010), (Heckerling, Canaris, Flach, Tape, Wigton, & Gerber, 2007) y (Pérez, De La Calleja, Medina, & Benitez, 2012); en **Astronomía** (De La Calleja, Benitez, Medina, & Fuentes, 2011) y (De La Calleja, Huerta, & Fuentes, 2010); en **Negocios**(Salles J., 2011), (Tirenni, Kaiser, & Herrmann, 2007) y (Alhah, Abu Hammad, Samhouri, & Al-Ghandoor, 2011); en **Robótica** (Conforth & Meng, 2008) y (Engedy, 2009)y en **Visión por computadora**(Isik, Leibo, & Poggio, 2012) y (Yokono & Poggio, 2009).

En este trabajo, se considera a la **inteligencia de negocios** (***BI*** por sus siglas en inglés) desde un punto de vista pragmático, y asociándolo directamente con las tecnologías de la información. Y considerando la definición de varios autores sobre la **BI** "*como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada en información estructurada, para su explotación directa o* ***para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio*****".**

Después de una revisión minuciosa de la literatura, se aprecia que la aplicación del aprendizaje automático en los negocios y toma de decisiones es aún poco estudiado; teniendo con ello un área de oportunidad para aplicar algoritmos de aprendizaje automático que puedan contribuir a una sistematización de la información para mejorar la toma de decisiones y crear una ventaja competitiva sobre la inteligencia de negocios (IN).

1.3. Justificación

La toma de decisiones es una tarea circunscrita a la función directiva de la empresa, organización, instituciones públicas o privadas. Tomar decisiones es un proceso sistemático compuesto de elementos claramente definidos que implica identificar y seleccionar un curso de acción para hacer frente a un problema concreto o bien, para aprovechar

oportunidades que se presenten (Rico García & Sacristán Navarro, 2012) y (Drucker, Hammond, Raiffa, & Argyris, 2001).

Como se mencionó anteriormente, la aplicación del aprendizaje automático se ha enfocado en aspectos particulares de la ciencia, dejando un poco de lado a los negocios y la toma de decisiones. Por ello surge el interés de proponer métodos haciendo que los procesos en la toma de decisiones se ejecuten automáticamente o semiautomáticamente a fin de agilizar la gestión que ello demande.

Para ello, se aplicará la tecnología de aprendizaje automático con el objetivo de innovar en la inteligencia de negocios y con ello, contribuir en la mejora de los procesos en la toma de decisiones.

1.4. Objetivos

En esta sección se detalla el objetivo general de esta investigación, mismo que conduce a generar objetivos específicos.

1.4.1 Objetivo general

Analizar información para mejorar la toma de decisiones y crear una ventaja competitiva en las IES a través del desarrollo de métodos computacionales empleando algoritmos de aprendizaje automático

1.4.2 Objetivos específicos

- Extraer información automáticamente para mejorar la toma de decisiones a través de aprendizaje automático logrando asertividad en el proceso

- Integrar equipos de trabajo de acuerdo a sus habilidades y afinidades personales aplicando el algoritmo k-medias
- Identificar a grupos de líderes a través de sus competencias y afinidades personales aplicando el algoritmo k-medias para la mejora del desempeño de los directivos y del personal clave en las empresas
- Identificar automáticamente las características más relevantes del capital humano para asignar con efectividad actividades acordes a su perfil usando selección automática de atributos
- Generar un árbol de decisión que permita clasificar al capital humano en líderes o colaboradores basándose en sus competencias y afinidades
- Desarrollar una propuesta estratégica para mejorar la toma de decisiones en la clasificación y selección del capital humano

1.5. Alcances

El alcance de esta investigación es experimental[1] (Díaz Narváez, 2009) pues tras la revisión de la literatura reveló que la aplicación del aprendizaje automático en la inteligencia de negocios es poco estudiada y con ello se establecen antecedentes para su aplicación en esta área. Al mismo tiempo, tiene un alcance descriptivo[2](Prieto Herrera, 2013) ya que se busca especificar las competencias y afinidades particulares de un grupo de personas que permiten enriquecer la productividad cuando se agrupan como equipo de trabajo.

En la realización de este proyecto existe una limitante particular relacionada con la obtención del conjunto de datos y la experimentación; pues no se tiene convenio alguno con una empresa que tenga más de 30 directivos y más de 100 colaboradores, por lo que la experimentación se realizará con la información obtenida de profesores y estudiantes de la Universidad Politécnica de Puebla (UPPuebla), dado que cuenta con más de 30 profesores de tiempo completo y más de 1500 estudiantes.

[1] Es aquella que se efectúa sobre un tema u objeto desconocido o poco estudiado, por lo que sus resultados constituyen una visión aproximada de dicho objeto, es decir, un nivel superficial de conocimiento.
[2] Consiste, fundamentalmente, en caracterizar un fenómeno o situación concreta indicando sus rasgos más peculiares o diferenciadores. Su meta no se limita a la recolección de datos, sino a la predicción e identificación de las relaciones que existen entre dos o más variables.

De esta limitación, se deriva un par más relacionado a la resistencia o desinterés de los involucrados (Estudiantes, Tutores, Director de Programa Académico, Administrativos) en facilitar la información necesaria para la experimentación, reduciendo con ello la posibilidad de tener un conjunto de datos grande.

1.6. Impacto esperado

La relevancia que se espera al proponer el uso de la tecnología de aprendizaje automático en el proceso de la toma de decisiones en la función directiva y en general en la inteligencia de negocios, es ofrecer otra forma para formar grupos de trabajo en función de sus competencias y afinidades semejantes y por consiguiente afianzar el desempeño de los mismos, en beneficio de la productividad empresarial.

Se espera que se formen equipos de trabajo homogéneos en cuanto a competencias, habilidades, afinidades y conocimientos, a fin de potenciar sus capacidades individuales.

Al mismo tiempo, se espera impactar en el ámbito social, ambiental, empleo, económico y sistema de innovación. En el capítulo V se detallarán cada uno de ellos.

Alternativamente, la aplicación de esta tecnología puede adaptarse a problemas relacionados con las instituciones de educación en donde también la creación de grupos de personas o asignaciones de personas para realizar actividades en conjunto determinan el mejor desempeño y por ende aumento de la productividad.

Referencias del capitulo

Alhah, S., Abu Hammad, A., Samhouri, M., & Al-Ghandoor, A. (2011). Modeling stock market exchange prices using artificial neural networks: a study of amman stock exchange. *Jordan Journal of Mechanical and Industrial Engineering, 5*(5), 439:446.

Arias Delgado, L. P. (2011). *Módulo: Cerebro y Aprendizaje.* Chile: Fundación universitaria del Área Andina.

Bensoussan, B. E., & Fleisher, C. S. (2013). *Analysis without paralysis: 12 tools to make better strategic decision.*USA: Pearson Education.

Bustillo Porro, V. (2009). Nuevas Tecnologías de la información: Herramientas para la educación. *Teoría de la Educación, 3*(10), http://campus.usal.es/~teoriaeducacion/rev_numero_06/n6_art_bustillo.htm. Obtenido de Ediciones Universidad de Salamanca: http://campus.usal.es/~teoriaeducacion/rev_numero_06/n6_art_bustillo.htm

Conforth, M., & Meng, Y. (2008). An Artificial Neural Network Based Learning Method for Mobile Robot Localization . *Robotics, Automation and Control* (pp. 494-504). Viena, Austria: I-Tech.

Coordinación de Universidades Politecnicas. (2012). *Modelo Educativo*. De Coordinación de Universidades Politecnicas: http://politecnicas.sep.gob.mx/

Cruz, J. A., & Wishant, D. S. (2007). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics, 2*.

De La Calleja, J., Benitez, A., Medina, M. A., & Fuentes, O. (2011). Machine learning from imbalanced data sets for astronomical object classification. *SoCPaR* (pp. 435-439). Dalian, China: IEEE Xplore Digital Library.

De La Calleja, J., Huerta, G., & Fuentes, O. (2010). The imbalanced problem in morphological galaxy classification. *CIARP* (pp. 533-540). Sao Paulo, Brasil: Springer Verlag.

De Quiroga, A. (2004). *El proceso educativo según Paulo Freire y Enrique Pichon-Riviére.* San Pablo, Brasil: Plaza y Váldes S.A. de C.V.

Díaz Narváez, P. (2009). *Medotología de la investigación y bioestadística.* Santiago de Chile: RIL Editores.

Dirección General de Desarrollo de la Gestión e Innovación Educativa-SEP. (2013 йил julio). *SEcretaría de Educación: Gobierno del Estado de Jalisco.* From Modelo de Gestión Educativa Estratégica: Programa Escuelas de Calidad: http://portalsej.jalisco.gob.mx/sites/portalsej.jalisco.gob.mx.programa-escuelas-calidad/files/pdf/mgee.pdf

Drucker, P. F., Hammond, J., Raiffa, H., & Argyris, C. (2001). *On Decision Making.* Boston: Harvard Bussines Review.

Engedy, I. (2009). Artificial neural network based mobile robot navigation. *Intelligent Signal Processing, 2009.WISP 2009* (pp. 241-246). Budapest, Hungria: IEEE Xplore Digital Library.

Er, O., Yumusak, N., & Temurtas, F. (2010). Chest disease diagnosis using ANNS. *Expert systems with application, 37*(12), 7648-7655.

Ferrell, O. C., Hirt, G. A., & Ferrel, L. (2010). *Introducción a los negocios en un mundo cambiante.*Colorado, CA: McGraw Hill.

Fujita, H., & Revetria, R. (2012). *New treds in software methodologies, tools and techniques. Proceedings of the Eleventh SoMeT_I2.* Amsterdam: IOS Press.

Heckerling, P. S., Canaris, G., Flach, S. D., Tape, T. G., Wigton, R. S., & Gerber, B. S. (2007). Predictors of urinary tract infection based on ANNS & genetic algorithms. *International journal of Medical Informatics, 76*(4), 289-296.

Hellriegel, & Slocum. (2009). *Comportamiento Organizacional.* México, D.F.: Cengage Learning Editores. From http://books.google.com.mx/books?hl=es&lr=&id=__g324XjZNwC&oi=fnd&pg=PR25&dq=toma+de+decisiones+gerenciales&ots=7k7_vdYHXk&sig=yPHU5MSjY57MjLy7h5FNw83bgp4#v=onepage&q=toma%20de%20decisiones%20gerenciales&f=false

Isik, L., Leibo, J. Z., & Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. *Front. Comput. Neurosci, 6*(37).

Kadhim A-Shayea, Q. (2011). Artificial neuronal networking in medical diagnosis. *International Journal of Computer Science Issues, 8*(2).

Kelly, K. P. (1999). *Las técnicas para la toma de decisiones en equipo.* Buenos Aires, Argentina: Ediciones Granica.

Mitchell, T. M. (1997). *Machine Learning.* Singapore: McGraw-Hill.

Pérez, P., De La Calleja, J., Medina, M. A., & Benitez, A. (2012). Aplication of machine learning to classify dialetic retinopathy. *SPPRA*, 146-153.

Prieto Herrera, J. E. (2013). *Investigación de mercados.* Bogotá-Colombia: Ecoe Editores.

Render, B., Stair, R. J., & Hanna, M. E. (2006). *Métodos cuantitativos para los negocios.* Prentice Hall.

Rico García, M. G., & Sacristán Navarro, M. (2012). *Fundamentos empresariales.* Madrid: ESIC, Editorial.

Rodríguez, A. (s.f.). *La función directiva.* México D.F.: UNAM.

Roman, J. D., & Ferrández, M. (2008). *Liderazgo y coaching.* Libros en Red.

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3era. ed.). Prentice Hall.

Rzaev, R., Aliev, E., Akbarov, R., & Askerov, N. (2013). Estimation of the Efectiveness of Regional Investment Projects by Fuzzy Conclusion Method. In A. M. Gil-Lafuente, L. Barcellos-Paula, J. M. Merigó-Lindahl, F. A. Silva-Marins, & A. C. De Azevedo-Ritto, *Decision Making Systems in Business Administration* (pp. 27-37). Singapur: World Scientific Publishing.

Salles J., A. A. (2011). Lean production and bussiness efficiency: An artificial neural network analysis in auto parts companies. *Techonology management conference (ITMC)* (pp. 855-863). Sao Paulo: IEEE Xplore Digital Library.

Sinnexus. (2012). *Business Intelligence*. Obtenido de ¿Qué es Business Intelligence?: http://www.sinnexus.com/business_intelligence/

Sprenger, M. (2010). *Brain-Based Teaching in the Digital Age.* Aurora, USA: Assn for Supervision & Curricu.

Tirenni, G., Kaiser, C., & Herrmann, A. (2007). Applying decision trees for value-based customer relations management: Predicting airline customers' future values. *Database Marketing & Customer Strategy Management, 14*(2), 130–142.

Torres Pérez, V., Sánchez García, J., & Ramírez Gutiérrez, D. (2014). *Los 6 pecados capitales en la inteligencia de negocios*. Obtenido de IPADE: http://www.ipade.mx/editorial/Pages/articulo-los-6-pecados-capitales-en-la-inteligencia-de-negocios.aspx

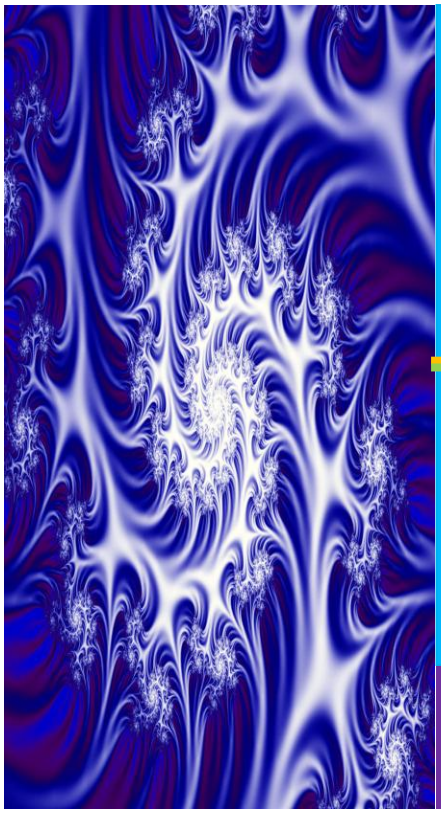
Universidad Politécnica de Puebla. (2011). *Certificaciones/ ISO 9000*. From Universidad Politécnica de Puebla: http://serpaguppue.uppuebla.edu.mx/ISO9000.php

Universidad Politécnica de Puebla. (2011). *Oferta Educativa*. From Universidad Politécnica de Puebla: http://serpaguppue.uppuebla.edu.mx/

Urbina Nájera, A. B., de la Calleja, J., Vega Lebrún, C. A., López Maldonado, N., & Pico González, B. (2014). Desarrollo y validación de un instrumento para identificar perfiles de tutorados y tutores de la modalidad virtual. *CAFVIR-2014* (págs. 227-234). Antigua Guatemala: CAFVIR.

Witten , I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques.* San Francisco, CA.: Morgan Kaufmann Publishers (Elsevier).

Yokono, J. J., & Poggio, T. (2009). Object Recognition Using Boosted Oriented Filter Based Local Descriptors. *IEEJ Transactions on Electronics, Information and Systems, 129*(5), 806-811.



Capítulo II
Marco Teórico

INTRODUCCIÓN

En este capítulo se hace una descripción de los temas fundamentales que sustentan este trabajo de investigación. Se presentan los conceptos principales de la toma de decisiones, su proceso y las relaciones con la minería de datos, se detalla a la inteligencia de negocios, sus factores, contribuciones y beneficios. Se particularizan las áreas de interés sustancial como es el caso de la inteligencia computacional, la minería de datos y el aprendizaje automático. Al mismo tiempo se detalla la herramienta Weka que se emplea para la experimentación. Finalmente, se expone la revisión de la literatura relacionada a la aplicación de algoritmos de aprendizaje automático en el contexto de estudio.

2.1 Planeación estratégica

La planeación estratégica formal con sus características modernas fue introducida por primera vez en algunas empresas comerciales a mediados de 1950. En aquel tiempo, las empresas más importantes fueron principalmente las que desarrollaron sistemas de planeación estratégica formal, denominados sistemas de planeación a largo plazo. Desde entonces, la planeación estratégica formal se ha ido perfeccionando al grado que en la actualidad todas las compañías importantes en el mundo cuentan algún tipo de este sistema, y un número cada vez mayor de empresas pequeñas la esta implementado para mejorar sus procesos (David, 1997).

La planeación estratégica está entrelazada de modo inseparable con el proceso completo de la dirección; por tanto, todo directivo debe comprender su naturaleza y realización. Además, a excepción de algunas empresas, cualquier compañía que no cuenta con algún tipo de formalidad en su sistema de planeación estratégica, se expone a un desastre inevitable. Algunos directores tienen conceptos muy distorsionados de ésta y rechazan la idea de intentar aplicarla; otros están tan confundidos acerca de este tema que lo consideran

sin ningún beneficio, y otros más ignoran las potencialidades del proceso tanto para ellos como para sus empresas. Existen quienes tienen cierto conocimiento, aunque no lo suficiente para convencerse que debería utilizarla (Hill & Jones, 2009).

Este tipo de planeación se utiliza para definir y alcanzar las metas de la organización al ocuparse de cuestiones fundamentales como dar respuesta a las preguntas ¿En qué negocio estamos?, ¿En qué negocio deberíamos estar?, ¿Quiénes son nuestros clientes?; al mismo tiempo, ofrece un marco de referencia para una planeación más detallada que coadyuva a la toma de decisiones ordinarias y como la alta gerencia participa activamente, es importante también, una toma de decisiones asertiva que incluya todos los aspectos de la organización. En este sentido, la planeación estratégica favorece la toma de decisiones asertivamente.

2.1.1 Fundamentos de planeación estratégica

Peter Drucker propone que el desempeño de un gerente sea juzgado mediante el doble criterio de la eficacia: 1) la habilidad para hacer las cosas "correctas" y 2) la eficiencia – la habilidad para hacerlas "correctamente". De estos dos criterios, Drucker sugiere que la efectividad es más importante, ya que ni el más alto grado de eficiencia posible podrá compensar una selección errónea de metas. Estos dos criterios se encuentran en paralelo con dos aspectos de la planeación: establecer las metas “correctas" y después elegir los medios "correctos" para alcanzar dichas metas. Para lograr esto, se requiere de una planeación estratégica. La figura 2.1 muestra la definición de estrategia considerando dos enfoques.

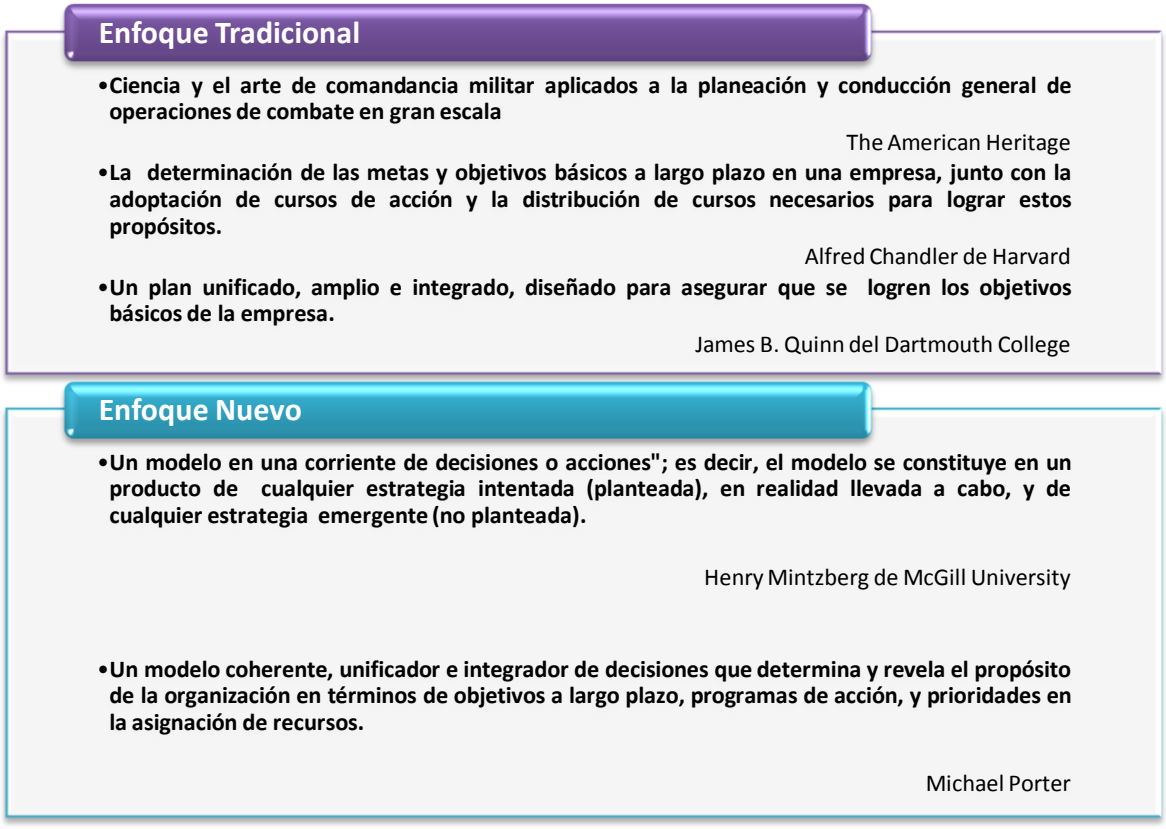


Figura 2.1. Definiciones sobre estrategia.

A diferencia de la planeación tradicional, la planeación estratégica contempla no solo el corto y el largo plazo, sino que hace énfasis en respuestas lógicas a necesidades de un futuro incierto, complejo y cambiante; busca prever los eventos futuros, y con ello, la posibilidad de describir el futuro de las decisiones actuales.

La planeación estratégica es una herramienta que permite a las organizaciones prepararse para enfrentar las situaciones que se presentan en el futuro, ayudando con ello a orientar sus esfuerzos hacia metas realistas de desempeño, por lo cual es necesario conocer y aplicar los elementos que intervienen en el proceso de planeación (David, 1997).

La figura 2.2 muestra el proceso de planeación estratégica propuesto por David Garvin, 1994.

Figura 2.2 Proceso de la Planeación Estratégica

De acuerdo a David Garvin, 1994, cada una de las etapas se describen a continuación.

- **Análisis del entorno.** Está enfocada en el diagnóstico, se analiza la industria para ver si es atractiva, también es necesario conocer el grado de competencia. Se trata de identificar, lo más objetivamente posible, las oportunidades y las amenazas. Asimismo es necesario un buen autodiagnóstico, hecho con objetividad hacia la empresa y hacia uno mismo, esto permitirá saber cuáles son las fuerzas (capacidades, competencias o habilidades) que lograrán aprovechar las oportunidades y también ayudará a identificar las debilidades (o limitaciones) que pueden evitar que una empresa compita eficazmente. Una vez analizada la industria y realizado el autodiagnóstico, valdría la pena crear dos o tres escenarios posibles, de situaciones que puedan presentarse, esto con la intención de "probarse", es decir, saber si se está preparado para enfrentarlos y para identificar las posibles señales de alarma que indiquen si estos escenarios pueden hacerse realidad.

- **Formulación.** Dentro del marco de referencia de la empresa que está definido por la misión (mi razón de ser), la visión (cómo me veo en el futuro cercano, en 10 o 20 años) y la filosofía y valores (las creencias y la cultura de la empresa), se debe tener un contexto desde el cual se podrá formular la estrategia con sus tres componentes: objetivos, plan de acción para lograrlos y capacidades y recursos que permitan llevar a cabo dicho plan de acción.
- **Programación.** Es la etapa de puente entre la formulación y la ejecución en donde se especificarán claramente las metas a alcanzar y se definirán, con cierta precisión, las actividades para alcanzar dichos objetivos.
- **Ejecución.** Se trata de llevar a cabo los programas, implementando las tareas. Coordinando las iniciativas, comunicando claramente las prioridades y dando un buen seguimiento.

En general, el ejercicio de la planeación estratégica constituye uno de los elementos centrales del proceso de toma de decisiones para la elección de la mejor alternativa y la óptima asignación de los recursos. En concreto, la planeación estratégica debe responder a tres preguntas: ¿Hacia dónde va?, ¿Cuál es el entorno? y ¿Cómo se logrará?. En este sentido, se aborda en el siguiente tema la toma de decisiones con el fin de buscar una respuesta a estas preguntas.

2.2 Toma de decisiones

Frecuentemente, las personas debemos elegir todos los días, entre varias, aquellas opciones que se consideran más convenientes. Es decir, hemos de tomar gran cantidad de decisiones en el actuar cotidiano, en mayor o menor grado de importancia, a la vez fáciles o difíciles de adoptar en función de las consecuencias o resultados derivados de cada una de ellas.

Es posible trasladar este planteamiento general al ámbito de la empresa. Cualquier gerente, directivo, empresario o mando intermedio de una empresa o persona que esté al frente de un negocio o de un grupo humano, se ve ineludiblemente obligado a tomar decisiones inagotablemente. Aunque, con frecuencia, sea un acto inconsciente. Es necesario elegircaminos, seleccionar personas, dar órdenes**,** elaborar programas, planificar acciones,

entre otras. Todas estas decisiones no serán del mismo nivel ni tendrán la misma importancia, y sin embargo; requieren tomar decisiones.

Schackle (1986) define la decisión como un corte entre el pasado y el futuro. Mientras que para Freemont E. Kast (1979) decidir significa adoptar una posición. Implica dos o más alternativas bajo consideración y la persona que decide tendrá que elegir entre ellas. Para Moody (1991), es una acción que debe tomarse cuando ya no hay más tiempo para recoger información. Leon Blank Buris (1990) considera que una decisión es la elección que se hace entre varias alternativas. Finalmente para P. Robbins (1996), decidir es la forma de como el hombre se comporta y actúa conforme a maximizar u optimizar cierto resultado, es decir, las decisiones se toman como reacción ante un problema. Es de gran utilidad conocer los procesos que se deben aplicar y abarcar para tomar decisiones efectivas.

Como se mencionó en el capítulo I, la toma de decisiones directivas en la UPPuebla, particularmente en el subproceso de tutoría, están relacionadas con el sesgo y la subjetividad; en este sentido, se busca proponer una forma diferente para mejorar la toma de decisiones apoyándonos de herramientas tecnológicas (Aprendizaje Automático) que conduzcan a su efectividad y asertividad, es por ello que en los siguientes apartados se describe el proceso de la toma de decisiones y las herramientas que contribuyen a la toma de decisiones asertivas.

2.2.1 Proceso de la toma de decisiones

En la toma de decisiones está inmersa la incertidumbre ya que no hay nada que garantice que las condiciones en las que se tomó la decisión sigan siendo las mismas. Tomar decisiones rápidamente en un mundo tan cambiante, complejo y en continua transformación, sugiere un proceso mental que conduce a seguir pasos para adoptar la decisión adecuada (Winograd, Fernández Lamarra, & Farrow, 1998).

La figura 2.3, muestra el proceso para la toma de decisiones propuesto por (Harvard business essentials, 2006).

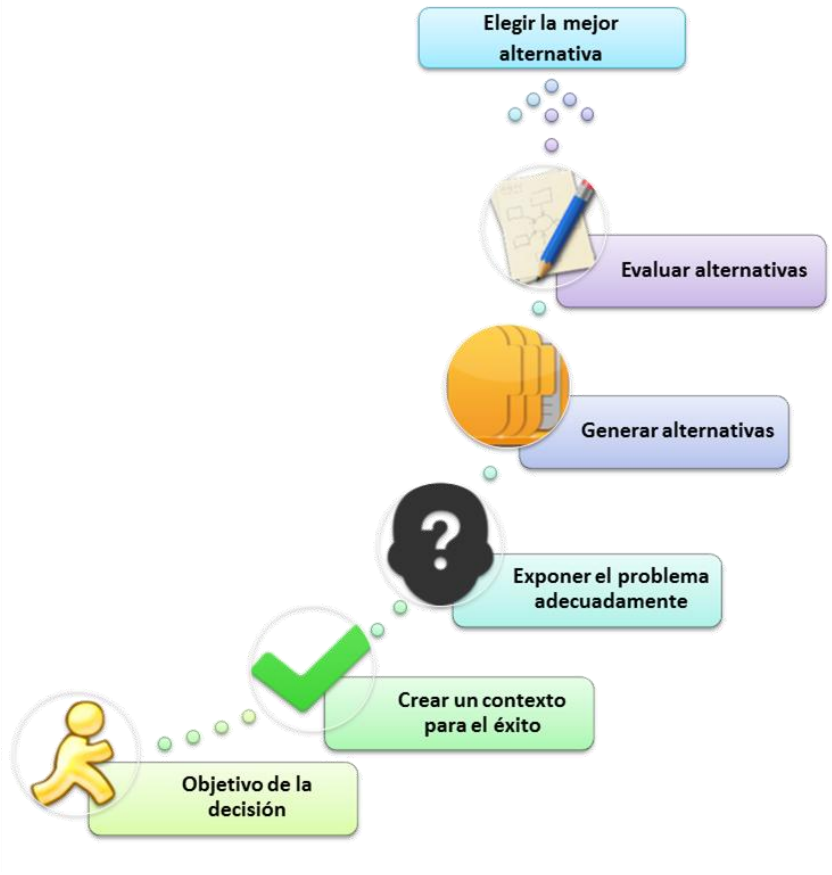


Figura 2.3 Proceso para la toma de decisiones

Objetivo de la decisión. Para esto es necesario contar con datos básicos e indicadores de tipo descriptivo, que sintetizan un conjunto de situaciones y medidas individuales para diferentes tipos de problemas. Dichos indicadores pueden ayudar al monitoreo de los problemas sociales, económicos, institucionales, ambientales, entre otros; a manera de proveer información para decidir mejor las acciones que se deben tomar en temas específicos. En otras palabras, estos indicadores se basan en información básica y estadística.

Establecer el contexto para el éxito. Uno de los mejores contextos para tomar decisiones efectivas es aquel que fomente el pensamiento creativo y la deliberación detallada; asegura, inherentemente que la gente correcta sea quien participe en el proceso. Además de, poseer reglas básicas que determinan cómo se tomará una decisión.

Exponer el problema adecuadamente. Toda decisión acertada depende de conocer claramente los problemas y de qué manera afecta cada uno de ellos a los objetivos de la empresa. Para ello, es fundamental determinar la naturaleza del problema.

Generar alternativas. Este paso consiste en la obtención de todas las alternativas viables que puedan tener éxito para la resolución del problema.

Evaluar alternativas. Una vez determinado el conjunto realista de alternativas, se tendrá que evaluar su viabilidad, además del riesgo e implicaciones de cada una de ellas.

Elegir la mejor alternativa. Cuando todos los pasos anteriores hayan sido elaborados cuidadosamente y el equipo de decisión haya acordado su objetivo, se podrá empezar a evaluar racionalmente cada una de las alternativas. También, este paso intenta que la decisión se lleve a cabo, e incluye dar a conocer la decisión a las personas afectadas y lograr que se comprometan con la misma. Si las personas que tienen que ejecutar una decisión participan en el proceso, es más fácil que apoyen con entusiasmo la misma. Estas decisiones se llevan a cabo por medio de una planificación, organización y dirección efectivas.

Es importante analizar reiteradamente el proceso de la toma de decisiones para saber si se ha corregido el problema. Si como resultado de esta evaluación se encuentra que todavía existe el problema tendrá que hacer el estudio de lo que se hizo mal. Las respuestas a estas preguntas nos pueden llevar de regreso a uno de los primeros pasos e inclusive al primer paso del proceso en la toma de decisiones.

2.2.2 Minería de datos y la toma de decisiones

La aplicación de la minería de datos (tema que se describe más adelante) en numerosos campos ofrece beneficios al emplear reglas o modelos para la toma de decisiones, la figura 2.4 muestra algunas aplicaciones de las técnicas de minería de datos para resolver diferentes problemas en la toma de decisiones en diversos sectores (Tufféry, 2011).

Figura 2.4 Aplicaciones de minería de datos en la toma de decisiones en diferentes sectores

También, la minería de datos ha tenido diferentes aplicaciones para resolver problemas como (Shawe-Taylor, 2006) y (Hamilton L. , 2014):

- La previsión del tráfico por carretera, día a día, o por intervalos de tiempopor hora
- La previsión de consumo de agua y electricidad
- Determinar si una persona posee una casa o larenta
- Cuando se planea ofrecer aislamiento o la instalación de un sistema de calefacción
- La mejora de la calidad de una red telefónica
- El uso de análisis de supervivencia en la industria con el objetivo de predecirla vida de un componente fabricado

- Conocer los perfiles de los solicitantes de empleo, con el fin de detectar a las personas desempleadas con mayor riesgo de desempleo a largo plazo y proporcionar asistencia rápida con base a sus circunstancias personales
- Recientemente, en el Reino Unido se ha empleado en el área de riesgo judicial. El proyecto tiene como objetivo estimar el riesgo de reincidencia en los casos de libertad anticipada, utilizando información sobre la liberación, antecedentes familiares, lugar de residencia, nivel educativo, los asociados, sus antecedentes, los informes de los trabajadores sociales y el comportamiento de la persona en cuestión en custodia y en prisión. Se pretende con ello, estandarizar la decisión sobre la reincidencia temprana, que actualmente varían ampliamente de una región a otra, especialmente bajo la presión de la opinión pública (Tufféry, 2011).

2.3 Inteligencia de negocios (IN)

Es un hecho incuestionable que la información es la clave de las organizaciones para generar una ventaja competitiva. En este mundo tan cambiante y competitivo se ha atenuado la necesidad de tener óptimos, más rápidos y más eficientes métodos para extraer, analizar y transformar los datos de una organización en información y distribuirla a lo largo de la cadena de valor[1] (Curto Díaz & Conera i Carat, 2010).

La inteligencia de negocios (*Business Intelligence, BI*) responde a esta necesidad al considerarse como una evolución de los sistemas de soporte de decisiones (DSS, *DecisionSupportSystems*) (Méndez del Río, 2006). Sin embargo, el concepto aún sigue siendo tema crítico en la mayoría de las empresas. La figura 2.5 muestra algunas definiciones de este concepto a lo largo de la historia(Williams & Williams, 2007), (Vercellis, 2009)(Sabherwal & Becerra-Fernández, 2011).

[1] De acuerdo a Michael E. Porter, la cadena de valor es un modelo teórico que permite describir las actividades que generan valor en una organización.
[2] Centroide es la medida de las puntuaciones de la discriminación de un grupo particular. Existen tantos centroides como grupos y un centroide por grupo. Las medias de un grupo en todas las funciones son los

Particularmente, el BI significa el aprovechamiento de los activos de información dentro de los procesos de negocio clave para lograr un mejor rendimiento del negocio (Williams &Williams, 2007) y (Vercellis, 2009). Se emplea esencialmente para: contextualizar los procesos de negocio clave, para la toma de decisiones y acciones de apoyo; y dar lugar a una mejora en los resultados empresariales.

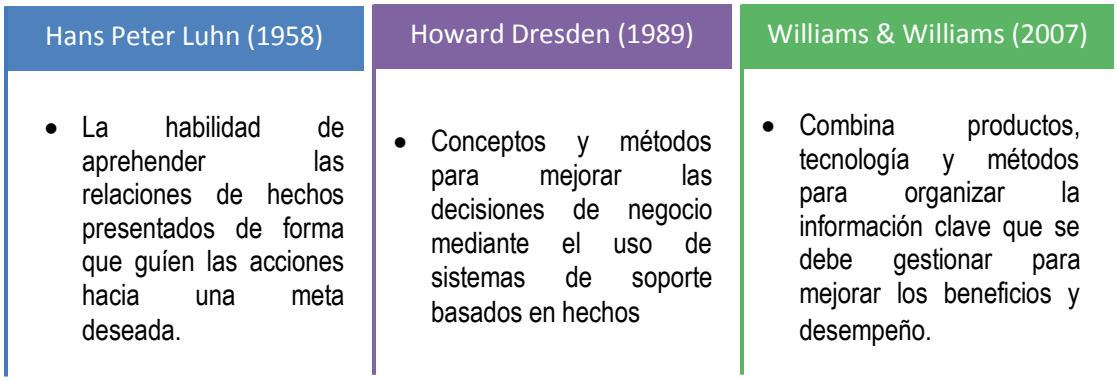


Figura 2.5 Definiciones de la inteligencia de negocios.

Actualmente, Sabherwal & Becerra-Fernández (2011) consideran que la IN tiene varias acepciones, unos lo consideran como el producto de un proceso, o la información y el conocimiento que son útiles para las organizaciones que ayudan a las actividades empresariales y la toma de decisiones. Suponen también que la IN se utiliza para referirse al proceso mediante el cual una organización obtiene, analiza y distribuye la información y el conocimiento. En este sentido, Sabherwal & Becerra-Fernandez, distinguen a la IN entre las herramientas desarrolladas por proveedores de IN y soluciones de IN desplegadas dentro de las organizaciones. Las soluciones de IN utilizan las herramientas de IN adquiridas por la organización y se basan en la gran cantidad de datos de los almacenes de datos existentes y los sistemas de procesamiento de transacciones, así como la información estructurada y no estructurada a partir de estas y otras fuentes para proporcionar información y conocimientos que facilitan la toma de decisiones.

Estos datos se información pueden referirse a aspectos tan diversos como la comprensión de las preferencias del cliente, hacer frente ala competencia, la identificación de

oportunidades de crecimiento y la mejora de la eficiencia interna. Las herramientas de IN, de este modo, se utilizan en soluciones de IN y estas soluciones IN apoyan al proceso a través del cual se proporcionan información y conocimientos valiosos. Las herramientas de la IN también pueden ayudar directamente en la obtención de dato se información. Algunas de estas herramientas de la IN (Curto Díaz & Conera i Carat, 2010)son:

- *Data warehouse*, *reporting*, análisis OLAP (*On-Line AnalyticalProcessing*), análisis visual, análisis predictivo, cuadro de mando, cuadro de mando integral, minería de datos, gestión del rendimiento, reglas de negocio, previsiones, *dashboards*, integración de datos (*ETL, Extract, Transform and Load*).

2.3.1 Factores que impulsan la IN

El incremento prominente de la IN está definido por factores que pueden ser clasificados en cuatro puntos (Sabherwal & Becerra-Fernández, 2011) descritos en la figura 2.6.

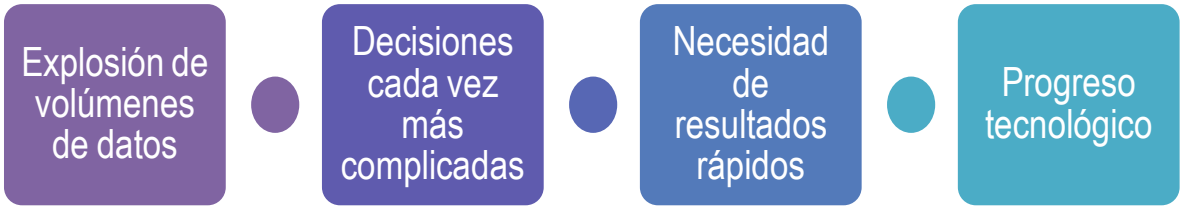


Figura 2.6 Factores que impulsan la inteligencia de negocios.

Explosión de volúmenes de datos. La necesidad de contar con almacenamiento de datos más barato, la existencia de más conexiones electrónicas (Internet, intranet, VPN), cambios regulatorios en el manejo y compartición de la información. Ante esto, la IN provee recursos con la capacidad de utilizar de forma eficaz estos volúmenes de datos.

Las decisiones cada vez resultan ser más complicadas debido a la competencia global y a la multi-industria; la toma de decisiones basadas en datos estructurados y no estructurados. En este sentido, la IN provee recursos con la capacidad de tomar decisiones que incorporan todos los factores importantes basados en la integración de este cúmulo de información.

Ante inmensa volatilidad, las ventanas de oportunidad, la conversión e integración de datos de diversas fuentes, poner a disposición, oportuna y rápida, los resultados ante quien toma las decisiones hacen de la IN una solución que ayuda a abordar cada uno de estas demoras.

El progreso tecnológico provee de sistemas de soporte a las decisiones (DSS), de sistemas empresariales de planificación de recursos (ERP), de almacenamientos de datos y de minería de datos. Los proveedores de BI tienen insumos necesarios para el desarrollo de herramientas eficaces; ante esto las organizaciones adoptan este tipo de plataformas para que se tengan soluciones de BI más eficaces.

2.3.2 Contribuciones y beneficios de la IN

El objetivo principal de los sistemas de inteligencia de negocio es proporcionar a los trabajadores conocimiento a través de herramientas y metodologías que les permitan tomar decisiones eficaces y oportunas. La figura 2.7 muestra de forma gráfica estos beneficios.

Decisiones eficaces. La aplicación de métodos analíticos rigurosos permite a quienes toman las decisionesse basenen la informacióny el conocimiento, que son más confiables. Como resultado, no son capaces de tomar mejores decisiones y elaborar planes de acción que permitan llegar a susobjetivos de manera efectiva (Vercellis, 2009).De hecho, quienes toman las decisiones emplean métodos analíticos como fuerzas formales para describir explícitamente tanto los criterios para la evaluación de opciones alternativas y los mecanismos que regulan el problema bajo investigación. Además, de que este proceso requiere de un examen en profundidad y pensamiento que conducen a una conciencia más profunda y a la comprensión de la lógica subyacente en el proceso de toma de decisiones (Sabherwal & Becerra-Fernández, 2011).

Decisiones oportunas. Las empresas operan en entornos económicos caracterizados por crecientes niveles de competencia y de alto dinamismo. Como consecuencia, la capacidad de reaccionar con rapidez a las acciones de los competidores y las nuevas condiciones del mercado es un factor crítico en el éxito o incluso en la supervivencia de una compañía. Si las personas que toman de decisiones pueden depender de un sistema de inteligencia de

negocios para facilitar la actividad, se espera que la calidad general del proceso de toma de decisiones se vea mejorada.

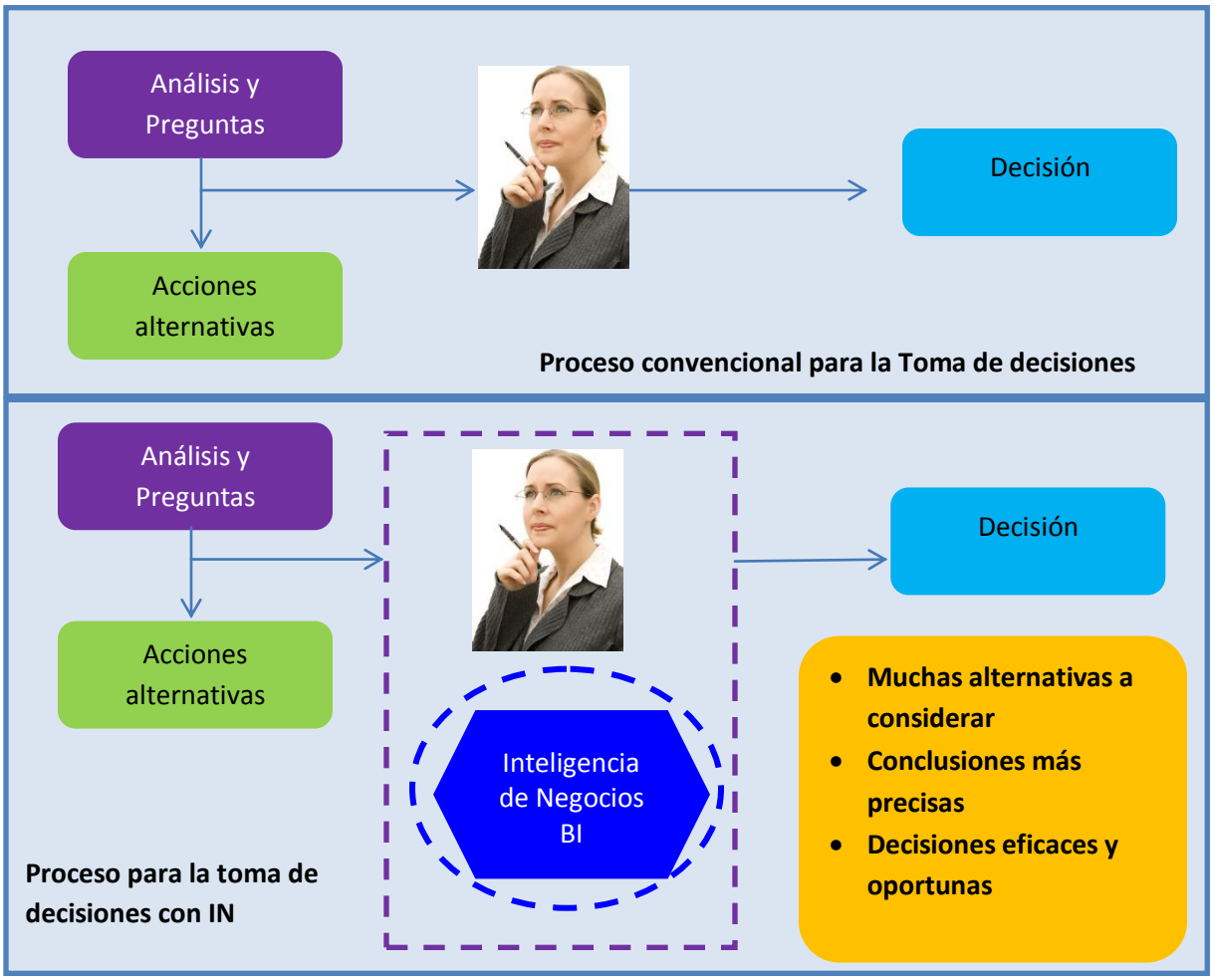


Figura 2.7 Proceso de la toma de decisiones con apoyo de la IN (Vercellis, 2009)

Con la ayuda de modelos matemáticos y algoritmos, en realidad es posible analizar un mayor número de acciones alternativas, lograr conclusiones más precisas, tomar decisiones eficaces, oportunas y sin sesgo. Por lo tanto, (Williams & Williams, 2007), (Vercellis, 2009) y (Sabherwal & Becerra-Fernández, 2011) concluyen que la principal ventaja derivada de la adopción de sistema de inteligencia de negocios se encuentra en el aumento de la eficacia del proceso de toma de decisiones.

2.4 Minería de datos

La inteligencia artificial (IA) fue consolidada en 1950 por Alan Turing proponiendo una prueba concreta para determinar si una máquina aprende o no, hasta ese entonces, Turing fue considerado el padre de la IA.

La IA es un área de las ciencias computacionales; considerada como una de las disciplinas que despiertan más la imaginación en el ser humano, aduciendo su potencial en la creación de robots que asemejen a un humano; esto gracias a que la ciencia ficción se ha encargado de crear sueños con máquinas capaces de razonar, de aprender, incluso de imitar las acciones y emociones de un ser humano, tal como se muestra en las películas escritas por Asimov, Silverberg, & Kazan (1999), Aldiss, Watson, & Spielberg (2001) yVintar, Goldsman, & Asimov(2004), entre otros.

La IA tiene como objetivo principal la creación de sistemas con cierto grado de "inteligencia" o autonomía, imitando la propia del ser humano (Russell & Norvig, 2009). De acuerdo a Muñoz Pérez (2010) la IA imita la inteligencia humana representando el conocimiento disponible del experto humano y utilizando mecanismos de razonamiento simbólico para resolver problemas de un dominio específico.

La IA se aplica en los sistemas reales en una gran variedad de ramas y problemas, tales como: gestión y control, procesos de fabricación, educación, cartografía, representación del conocimiento, estudio de las arquitecturas de agentes, coordinación y colaboración de multi-agentes, desarrollo de ontologías, procesamiento de voz y lenguaje, procesamiento de imágenes, etc. (García Fernández, 2004). Con esta motivación, ha surgido como disciplina la inteligencia computacional.

La IC se ocupa de la teoría, diseño, desarrollo y aplicación de modelos computacionales producidos lingüística y biológicamente, poniendo énfasis en redes neuronales, algoritmos genéticos, programación evolutiva, sistemas difusos y sistemas inteligentes híbridos. Algunos métodos que se consideran dentro de la IC incluyen aprendizaje supervisado y no supervisado mediante sistemas adaptativos, agrupando enfoques neuronales, difusos, evolutivos, probabilísticos y estadísticos tales como las redes Bayesianas y los métodos

basados en núcleos (Muñoz Pérez, 2010). La figura 2.6 muestra algunos paradigmas que conforman la IC.

Figura 2.8 Algunos paradigmas de la Inteligencia Computacional.

Como se observa en la figura 2.8, el paradigma o área objeto de aplicación en este trabajo de tesis es el aprendizaje automático y minería de datos. La minería de datos es el proceso de extraer conocimiento oculto a partir de grandes volúmenes de datos en bruto. Técnicamente, la minería de datos es el proceso de encontrar correlaciones o patrones entre los miles de campos en grandes bases de datos. En los sistemas de bases de datos, los registros se devuelven de acuerdo a una consulta, mientras que en el descubrimiento de conocimiento, lo que se recupera no está explícita en la base de datos, es decir, patrones implícitos. La minería de datos encuentra estos patrones y relaciones utilizando las herramientas y técnicas de análisis de datos para construir modelos, es decir, unsistema de aprendizaje (Witten & Frank, 2005).

Los datos adoptan la forma de un conjunto de ejemplos, mientras que la salida toma la forma de predicciones sobre nuevos ejemplos. Estos ejemplos son las "cosas" que se desean clasificar, asociar o agrupar, y comúnmente se denominan **instancias**. Cada instancia es un ejemplo individual, independiente del concepto que hay que aprender. Además, cada uno se caracteriza por los valores de un conjunto de atributos predeterminados, que será la entrada a los algoritmos de aprendizaje automático.

El valor de un atributo es una medida de la cantidad a la que se refiere el atributo (Han, Kamber , & Pei, 2011), por ejemplo un atributo “edad” puede tener valores numéricos. A continuación se describe el proceso de la minería de datos, como el descubrimiento de conocimiento en bases de datos (KDD, *Knowledge Discovery in Databases*), que se concentra principalmente en las siguientes tareas: colección de datos, pre-procesamiento de los datos, selección de atributos y aplicación de algoritmos de aprendizaje automático (Figura 2.9).

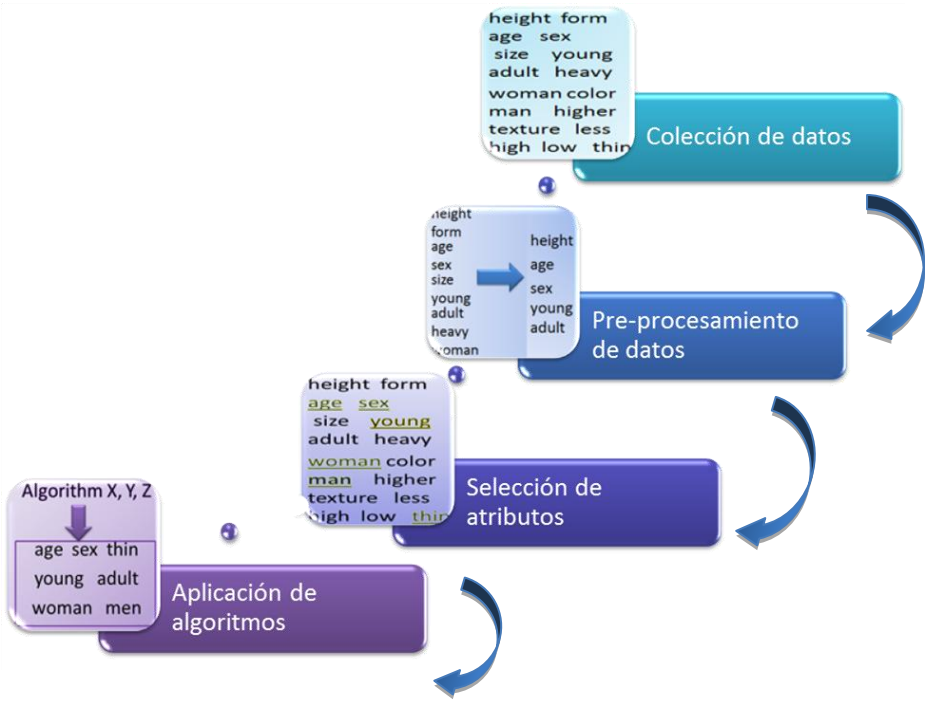


Figura 2.9 Proceso de descubrimiento de conocimiento útil en datos.

Colección de datos

Los datos es toda aquella información con la que una computadora trabaja y procesa para obtener un resultado (ejemplo: texto, imágenes, documentos,etc.). Hoy en día, existe una gran cantidad de datos en diferentes formatos y diferentes bases de datos, tales como ventas, costos, inventarios, datos de previsión, entre otros.

El proceso de recolección de datos incluye todos los pasos necesarios para obtener los datos deseados en un formato digital. Los métodos de recolección de datos incluyen la adquisición y almacenamiento de las nuevas observaciones, consultar bases de datos existentes de acuerdo con el problema, y si es necesario realizar cualquier combinación de datos(Han, Kamber , & Pei, 2011).

Pre-procesamiento de los datos

El pre-procesamiento consiste en manipular, enriquecer, reducir o transformar los datos originales para ser accesibles, posteriormente, con mayor facilidad (Han, Kamber , & Pei, 2011).

Los algoritmos requieren atributos, es decir, los valores de los campos de datos que describen las propiedades de cada instancia, pudiendo ser numéricos o nominales. Los atributos numéricos, a veces llamados continuos, toman números reales o enteros. Por otro lado, los atributos nominales pueden tomar valores de un conjunto finito establecido de antemano. Es posible transformar atributos numéricos a nominales y viceversa.

En general, los datos pueden tener valores incorrectos o inclusive la ausencia de algunos. Para el primer caso, estos valores pueden simplemente eliminarse, mientras que para los valores faltantes se puede hacer interpolación usando los datos existentes (Witten & Frank, 2005).

Selección de Atributos

Es frecuente que se tenga un gran número de atributos para cada instancia en un conjunto de datos, sin embargo no todos pueden ser relevantes para caracterizar al objeto. De hecho,

si se utilizan todos los atributos pueden, en muchos casos, causar un problema (Witten & Frank, 2005). Esto se puede describir de la siguiente manera, el gran número de atributos representa un espacio de alta dimensión, por lo que es necesario llevar a cabo una reducción de la dimensionalidad, seleccionando sólo unos pocos atributos. Este pequeño conjunto de atributos debe conservar la mayor cantidad de información posible y que describan a los ejemplos (Bishop, 2007). Existen diversas técnicas para la reducción de datos que incluyen el análisis de componentes principales, análisis de componentes independientes, matriz de Fisher, entre otros.

2.5 Aprendizaje automático

El aprendizaje automático (AA) es la rama de la Inteligencia Artificial que se dedica al estudio de los agentes/programas que aprenden o evolucionan basados en su experiencia, para realizar una tarea determinada cada vez mejor. El objetivo principal de todo proceso de aprendizaje es utilizar la evidencia conocida para poder crear una hipótesis y poder dar una respuesta a nuevas situaciones no conocidas (Mitchell, 1997).

La figura 2.10 muestra las disciplinas y ejemplos en los cuales el aprendizaje automático ha tenido influencia.

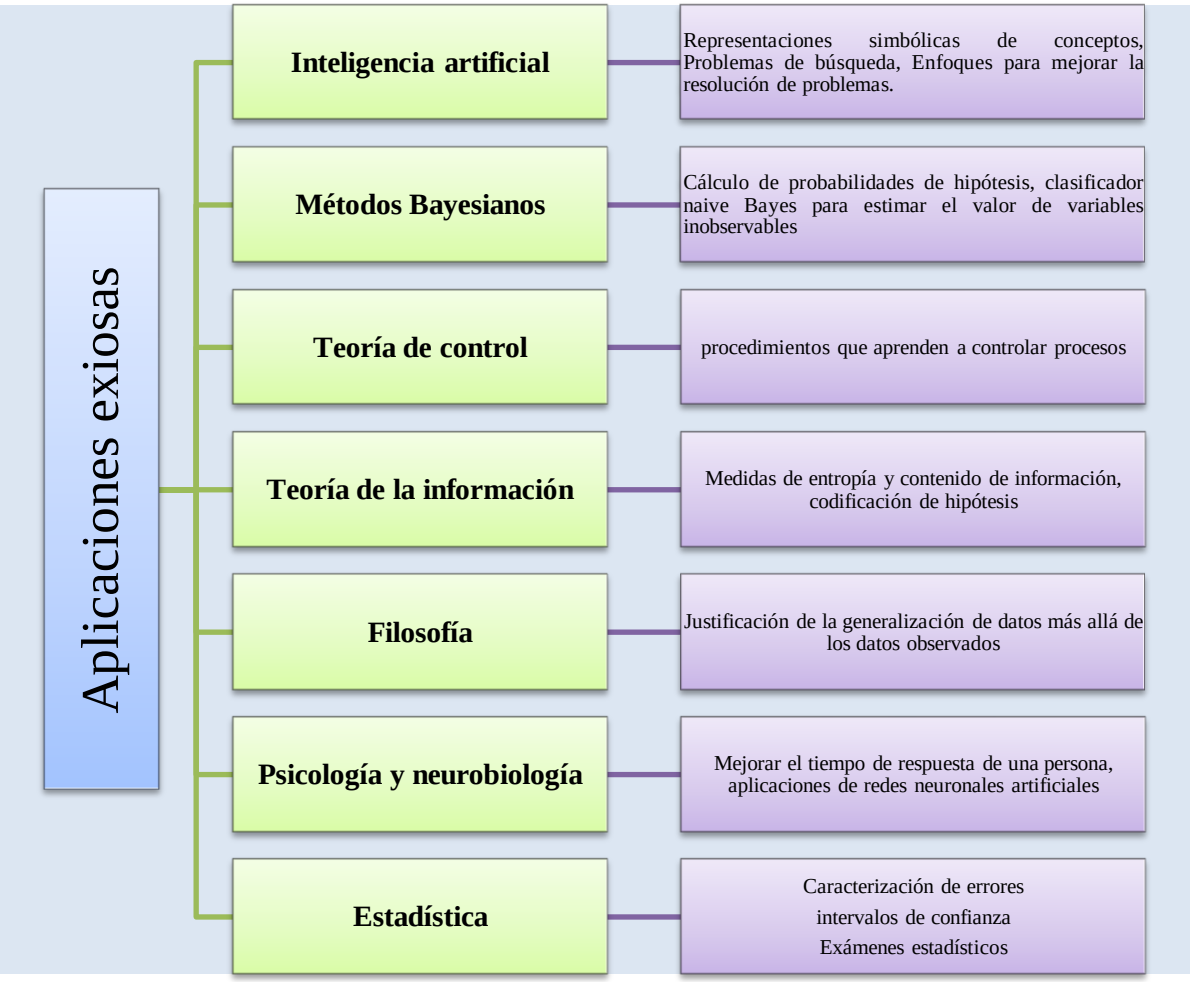


Figura 2.10 Aplicaciones exitosas del aprendizaje automático. (Mitchel, 1997)

El AA, también ha tenido impacto en otras áreas del conocimiento según (Álvarez & Zwir, 2001), (Shawe-Taylor, 2006), (Rodríguez & Mislej, 2013), (Hamilton, 2014), mismas que se mencionan a continuación:

- **Biología molecular**: se ha empleado para realizar un análisis inductivo de la estructura super-secundaria de proteínas, regulación molecular de alfa-virus, codificación basada en conocimiento de la topología de proteínas, producción de la topología protéica a través de satisfacción de estricciones, predicción de ángulos de torsión, sistemas expertos para la purificación de proteínas, entre otros.

- **Economía, mercadotecnía y finanzas:** se ha empleado para conocer las tendencias geográficas de ventas, análisis de la canasta de compras, realizar perfiles demográficos, perfiles de comportamientos de cliente, manejo de portafolios, fraudes e intromisiones no deseadas, etc.
- **Ingeniería, control y robótica:** detección y predicción de fallas, monitoreo de procesos de manufactura, modelado de sistemas altamente no lineales, localización y sistemas de almacenamiento, para regular el desplazamiento de robots, planificación de tareas, etc.
- **Medicina y salud:** conocer las tendencias de salud, descubrimiento de patrones geográficos, clasificación y caracterización de pacientes, sistemas expertos de apoyo en diagnósticos, análisis causal, predecir el tiempo de espera de los pacientes en la sala de urgencias, extracción de criterios de diagnóstico de insuficiencia cardiaca, predecir reingresos hositalarios, entre otros.
- **Ciencias biológicas:** clasificación de especies, reconocimiento de tumores o arritmias, patrones en las cadenas de ADN, predicción de patrones de movilidad, etc.
- **Recursos Humanos:** creación de modelos para la automatización del control de acceso de empleados que buscan minimizarlaintervención humanarequerida paraotorgar o revocarel acceso de empleados
- **Veterinaria:** identificación de ballenas en el océano basados en grabaciones de audio para que los barcos eviten golpearlas, así como el reconocimiento de aves através de una grabación de audio obtenida en condiciones de campo.

2.5.1 Notación

De acuerdo a Mitchell (1997) la notación para el AA está definido por elementos denominados conjunto de *instancias*, denotada por X. Para comprender mejor la notación, Mitchell (1997) propone un ejemplo que se describe en la tabla 2.1.

Al momento de aprender el concepto objetivo, al algoritmo se le presenta un conjunto de ejemplos de entrenamiento, cada uno compuesto de una instancia $x \epsilon X$, junto con su valor

concepto objetivo *c(x)* (Véase tabla 2.1). Las instancias para cada $c(x) = 1$ se llaman ***ejemplos positivos***, o miembros del concepto objetivo. Las instancias $c(x) = 0$ se denominan ***ejemplos negativos***, o no miembros del concepto objetivo. Frecuentemente, se escribe un par ordenado $\{x, c(x)\}$ para describir el ejemplo de entrenamiento que consiste en la instancia x y su valor concepto objetivo $c(x)$. Se utiliza el símbolo D para indicar el conjunto de ejemplos de entrenamiento disponibles.

Dado:

Instancias X

Posibles días, cada uno descrito con sus atributos
- *Cielo* (con posibles valores soleado, nublado, soleado)
- *Temperatura del Aire* (con valores caluroso, frio)
- *Humedad* (con valores normal y alto)
- *Viento* (con valores fuerte y débil)
- *Agua* (con valores caliente y frío)
- *Pronóstico* (con valores iguales y cambiantes)

Hipótesis H

Cada hipótesis es descrita por una conjunción de restricciones en los atributos *cielo, temperatura del aire, humedad, viento, agua y pronóstico*. Las restricciones pueden ser "?" (cualquier valor es aceptable), "∅" (ningún valor es aceptable), o un valor específico

Concepto objetivo ***c***

Disfrutar un deporte: $X \rightarrow \{0,1\}$

Ejemplos de entrenamiento D

D: ejemplos Positivos o Negativos para la función objetivo, descrita a continuación:

Ejemplo	Cielo	Temperatura del aire	Humedad	Viento	Agua	Pronóstico	Disfrutar un deporte
1	Soleado	Caluroso	Normal	Fuerte	Caliente	Igual	Si
2	Soleado	Caluroso	Alto	Fuerte	Caliente	Igual	Si
3	Lluvioso	Frío	Alto	Fuerte	Caliente	Cambiante	No
4	Soleado	Caluroso	Alto	Fuerte	Fría	Cambiante	Si

Determinar:

- Una hipótesis $h$en Hde tal forma que $h(x) = c(x) para\ toda\ x\ en\ X$

Tabla 2.1 Concepto de la tarea de aprendizaje

Dado un conjunto de ejemplos de entrenamiento del concepto objetivo c, los problemas que enfrenta el algoritmo es crear la hipótesis o estimar el valor para c.

Se utiliza el símbolo H para denotar el conjunto de todas las ***hipótesis positivas*** que el algoritmo puede considerar en relación con la identidad del concepto objetivo. Normalmente H está determinado por el diseñador humano quien elige la representación de la hipótesis. En general, cada hipótesis $h\ en\ H$representa una función booleana defina sobre c, es decir $h: X \rightarrow \{0,1\}$. El objetivo del algoritmo es encontrar una hipótesis H tal que que $h(x) = c(x) para\ toda\ x\ en\ X$.

2.5.2 Tipos de aprendizaje

La figura 2.11 muestra los tipos de AA. Para efectos e intereses de este trabajo de tesis, solamente nos centraremos en la descripción del aprendizaje supervisado y aprendizaje no supervisado.

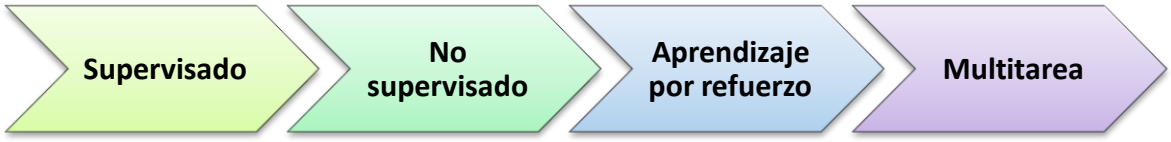


Figura 2.11 Tipos de aprendizaje

Aprendizaje supervisado

Este tipo de aprendizaje tiene como objetivo inferir una función a partir de datos de entrenamiento etiquetados, es decir, se tiene una clasificación de los datos. Los datos de entrenamiento son un conjunto de ejemplos de entrenamiento. En este tipo de aprendizaje, cada ejemplo es un par que consta de un objeto de entrada y un valor de salida deseado. Un algoritmo de aprendizaje supervisado analiza los datos de entrenamiento y produce una función deducida que puede utilizarse para el mapeo de nuevos ejemplos (Bishop, 2007).

La notación formal para este tipo de aprendizaje es la siguiente:

- N = ejemplos de entrenamiento
- $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ donde x_i es el vector de características de $i$ejemplos y y_i es la etiqueta.
- $g: X \rightarrow Y$ donde X es el espacio de entrada y Y es el espacio de salida
- La función g es un elemento de algún espacio posible en la función Gusualmente llamada espacio de hipótesis, en este sentido, conviene representar a $g$usando la función $f: X\ x\ Y\ \rightarrow \mathbb{R}$ tal que g se define como el valor de salida que da la máxima puntuación: $g(x) = argmax\ f(x, y)$.F denota el espacio de puntuación.

Con el fin de resolver un problema determinado de aprendizaje supervisado, es necesario realizar los siguientes pasos (Bishop, 2007) y (Wikipedia, 2014):

2. ***Determinar el tipo de ejemplos de entrenamiento.***el usuario debe decidir qué tipo de datos se va a utilizar como un conjunto de entrenamiento.

3. ***Recolectar el conjunto de entrenamiento***.El conjunto de entrenamiento debe ser representativo del mundo real de la función.Por lo tanto, un conjunto de objetos de entrada y salidas correspondientes deben ser recolectados, ya sea por los expertos humanos o basados en mediciones.

4. ***Determinar la representación de entidad de entrada de la función aprendida***.La precisión de la función aprendida depende en gran medida de cómo se representa el objeto de entrada.Típicamente, el objeto de entrada se transforma en un vector de características, que contiene una serie de tipologías que son descriptivas del objeto.El número de características no debe ser demasiado grande; pero debe contener suficiente información para predecir con precisión la salida.

5. ***Determinar la estructura de la función aprendida y algoritmo de aprendizaje correspondiente.*** Se pueden emplear múltiples algoritmos como k-medias, árboles de decisión, selección de atributos, k-vecinos más cercanos, entre otros.

6. ***Completar el diseño***.Ejecutar el algoritmo de aprendizaje en el conjunto de entrenamiento que fue recolectado.Algunos algoritmos de aprendizaje supervisado requieren del usuario para determinar ciertos parámetros de control.Estos parámetros pueden ajustarse mediante la optimización del rendimiento en un subconjunto (llamado un conjunto *de validación)* del conjunto de entrenamiento, o por medio de validación cruzada.

7. ***Evaluar la exactitud de la función aprendida***.Después de que los parámetros se ajustaron y aprendieron, el desempeño de la función resultante se debe medir en un conjunto de pruebas que está separado del conjunto de entrenamiento.

Se encuentra disponible una amplia gama de algoritmos de aprendizaje supervisado como las redes neuronales artificiales, estadística bayesiana, estimaciones Kernel, k-vecinos más cercanos, clasificador naiveBayes, etc., cada uno con sus fortalezas y debilidades.No existe un algoritmo de aprendizaje único que funcione mejor en todos los problemas de aprendizaje supervisado (Wikipedia, 2014).

Aprendizaje no supervisado

El objetivo de este tipo de aprendizaje es el de tratar de encontrar la estructura oculta en los datos no etiquetados, es decir, ahora en este tipo de aprendizaje no se tiene una clasificación de instancias. Está constituido por un conjunto de reglas que dan al algoritmo la habilidad de aprender asociaciones entre los patrones que ocurren en conjunto. Una vez que los patrones se han aprendido como asociación le permite a los algoritmos realizar tareas útiles de reconocimiento de patrones y la habilidad de recordar.

Este tipo de aprendizaje está estrechamente relacionado con la estimación de la densidad en las estadísticas; aunque también abarca otras técnicas que tratan de resumir y explicar las principales características de los datos. Algunos algoritmos de aprendizaje no supervisado

abarcan la agrupación (k-medias, agrupación jerarquica, etc), modelos de Markov, entre otros (Bishop, 2007).

2.6 Algoritmos para "aprender"

Como se mencionó anteriormente, el aprendizaje automático es una rama de la inteligencia computacional cuyo objetivo es crear técnicas que permitan a las computadoras "*aprender"*. Es decir, se trata de desarrollar programas capaces de generalizar comportamientos a partir de una información no estructurada suministrada en forma de ejemplos, a esto se le conoce como un proceso de inducción del conocimiento.

En esta sección se describen los algoritmos que se emplearán para la experimentación en este trabajo de tesis. A saber, árboles de decisión, *k*-medias y selección de atributos.

2.6.1 Árboles de decisión

Los árboles de decisión (*decisión tree*, DT por sus siglas en inglés) se ubican dentro de una rama del aprendizaje automático denominada aprendizaje simbólico, en la que también se encuentran los modelos de reglas de decisión, estrechamente relacionados con los árboles.

El aprendizaje mediante árboles de decisión es una técnica que permite analizar decisiones secuenciales basadas en el uso de resultados y probabilidades asociadas. Mitchel (1997) lo define como "*Un método de aproximación de una función objetivo de valores discretos en el cual la función objetivo es representada mediante un árbol de decisión. Los árboles aprendidos también pueden representarse como un conjunto de reglas Si–entonces*…" son uno de los métodos de aprendizaje inductivo más usado en los algoritmos de inferencia inducción y han sido aplicados exitosamenteen diversos ámbitos. La figura 2.12 muestra las ventajas y desventajas de este algoritmo.

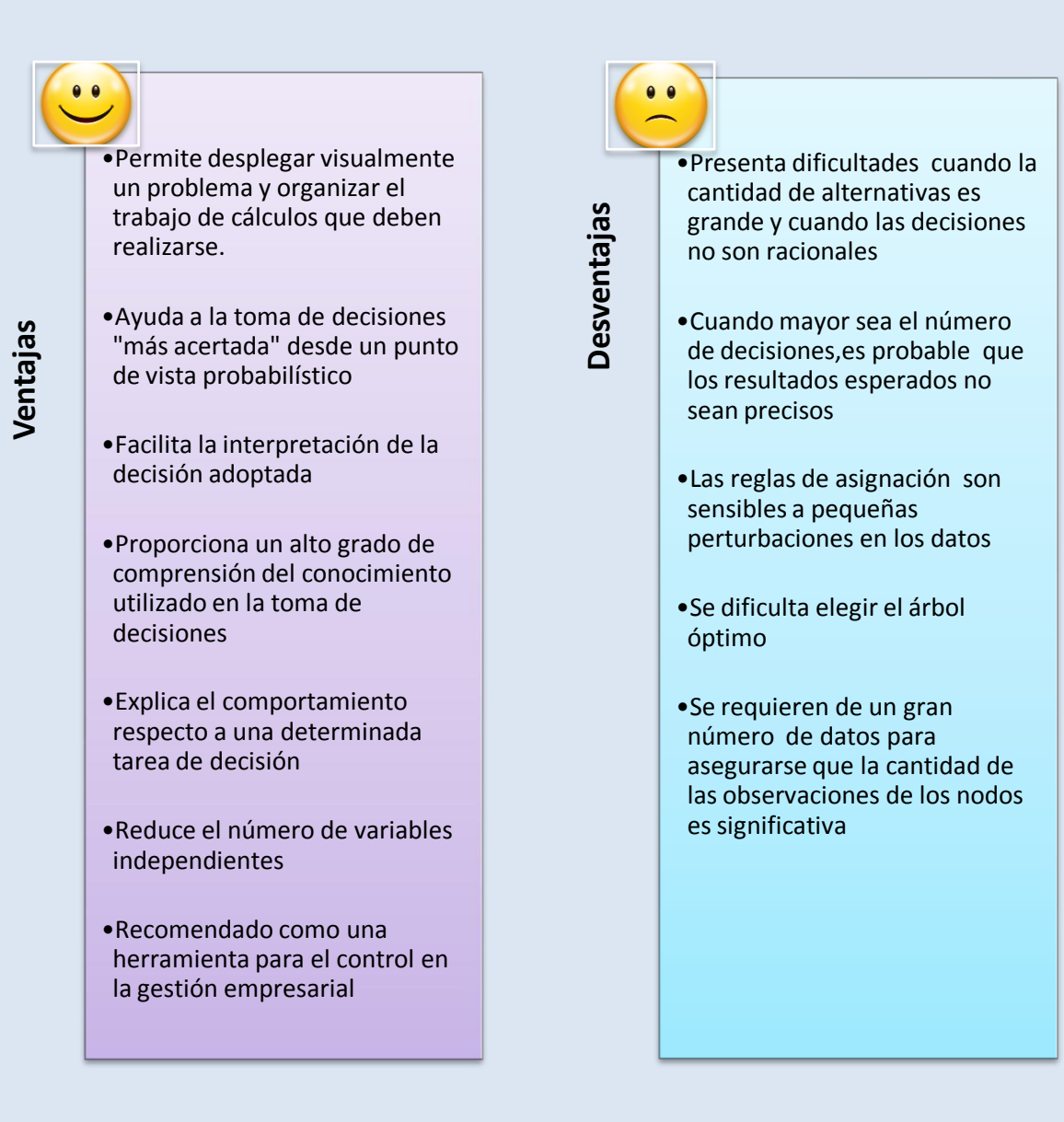


Figura 2.12. Características de los árboles de decisión

Un ejemplo clásico, es el planteado por (Mitchell, 1997) en donde el árbol de decisión clasifica las mañanas de domingo de acuerdo al clima apropiado para jugar tenis. Esto de acuerdo a la instancia dada por {Cielo=soleado, Temperatura=calor, Humedad= Alta, Aire=fuerte}, en la figura 2.13 se muestra la representación de esta instancia.

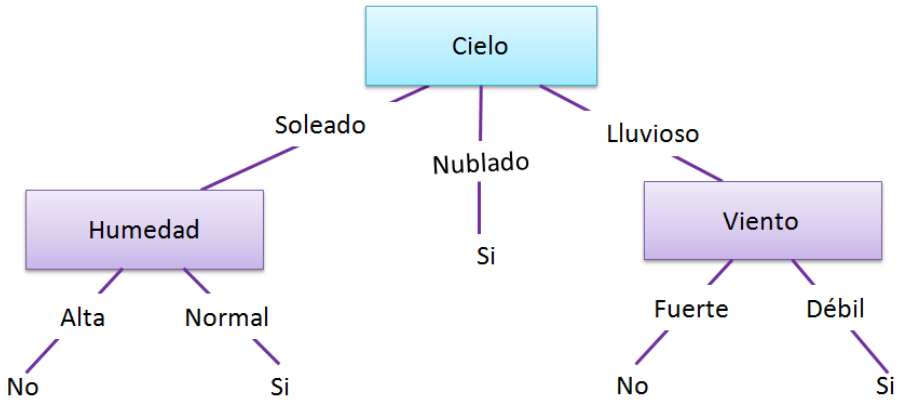


Figura 2.13 Un árbol de decisión para decidir su se puede realizar el partido de tenis. (Mitchel, 1997)

El árbol de la figura 2.13 corresponde a la hipótesis:

(*Cielo = Soleado* ^ *Humedad = Normal*)

(*Cielo = Nublado*)

(*Cielo = Lluvioso* ^ *Viento = Débil*)

Representación:

- Cada nodo que no sea una hoja representa un atributo.
- Las aristas que partan del nodo etiquetado con el atributo *A* están etiquetadas con cada uno de los posibles valores del atributo *A*.
- Cada hoja corresponde a un valor de la clasificación

La mayoría de losalgoritmosque han sido desarrolladospara el aprendizaje delos árboles de decisiónson variaciones deun algoritmoque emplea unnúcleode arriba hacia abajo (*top-down*). Este enfoque, particularmente, incluye el algoritmo ID3 y su sucesor C4.5, ambos desarrollados por Quinlan en 1986 y 1993, respectivamente.

Algoritmo ID3 (*Induction Decision Trees*)

Es un sistema de aprendizaje supervisado que construye árboles de decisión a partir de un conjunto de ejemplos. Estos ejemplos o tuplas están constituidos por un conjunto de atributos y un clasificador o clase. Los dominios de los atributos y de las clases deben ser discretos. Además, las clases deben ser disjuntas. En general, es un algoritmo que genera descripciones que clasifican a cada uno de los ejemplos del conjunto de entrenamiento (Mitchell, 2000).

Algoritmo C4.5

Es una extensión de ID3, permite trabajar con valores continuos para los atributos, separando los posibles resultados en dos ramas. Los árboles que genera son menos frondosos porque cada hoja no cubre una clase en particular sino una distribución de clases. C4.5 genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente, según la estrategia de profundidad-primero *(depth-first)*. Antes de cada partición de datos, el algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información o en la mayor proporción de ganancia de información. Para cada atributo discreto, se considera una prueba con *n* resultados, siendo *n* el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una *prueba binaria* sobre cada uno de los valores que toma el atributo en los datos (Mitchell, 2000).

Entropía y ganancia de información

Para construir un DT es necesario determinar qué atributos son los mejores, particularmente, cuál es el atributo que debe colocarse en el nodo raíz. Así, la entropía y la ganancia de información son utilizadaspara dar respuesta a estas incógnitas.

De acuerdo a (Mitchel, 1997) la entropía es una medida que permite calcular el grado de incertidumbre de una muestra. Si la muestra es completamente homogénea, su $entropía = 0$, al contrario de una muestra igualmente distribuida, cuya $entropía = 1$.

Una colección S, contiene ejemplos negativos y positivos de algún concepto objetivo, la entropía de S es relativa a una clasificación booleana expresada como:

$$Entropía(S) \equiv -p_{\oplus} log_2 p_{\oplus} - p_{\ominus} log_2 p_{\ominus} \qquad \text{(E1)}$$

Donde $p_{\oplus}$ es la proporción de los ejemplos positivos en S y $p_{\ominus}$ es la proporción de ejemplos negativos en S.

La ganancia de información, $Ganancia\ (S, A)$de un atributo A, relativo a una colección de ejemplos S es definido como:

$$Ganancia\ (S, A) \equiv Entropía(S) - \sum \frac{|\ S_v|}{|\ S|} Entropía(S_v) \qquad \text{(E2)}$$

Donde los valores para A son un conjunto de posibles valores para el atributo A, y S_v es el subconjunto de S para cualquier atributo A. El primer término de la ecuación (E2) es precisamente la entropía de la colección original S y el segundo término es el valor esperado de la entropía después de que S es particionada usando atributos v. La (E2) permite determinar el mejor atributo cuando la ganancia de información es la más alta.

Para ejemplificar, supóngase que Ses un conjunto de entrenamiento con 14 ejemplos. 9 ejemplos positivos y 5 negativos ([9+,5-]). Unos de los atributos, *Viento*, puede tomar los valores *Débil* y *Fuerte*. La distribución de ejemplos positivos y negativos según los valores de $viento = positivos\ \{débil = 6, fuerte = 3\},\ negativos\{débil = 2, fuerte = 3\}$.

La ganancia de informaciónque se tiene si se clasifican los 14 ejemplos según el atributo *Viento* (de acuerdo a E2) es:

$$Ganancia\ (S, A) \equiv Entropía(S) - \sum \frac{|\ S_v|}{|\ S|} Entropía(S_v)$$

$$= Entropía(S) - \sum \frac{|\ 8|}{|\ 14|} Entropía(S_{Débil}) - Entropía(S) - \sum \frac{|\ 6|}{|\ 14|} Entropía(S_{fuerte}\ )$$

$$= 0.940 - \frac{|\ 8|}{|\ 14|} 0.811 - \frac{|\ 6|}{|\ 14|} 1.00$$

$$= 0.048$$

Entonces, ¿cuál es el atributo que está mejor clasificado?. La figura 2.14 muestra la representación gráfica que da respuesta a esta pregunta.

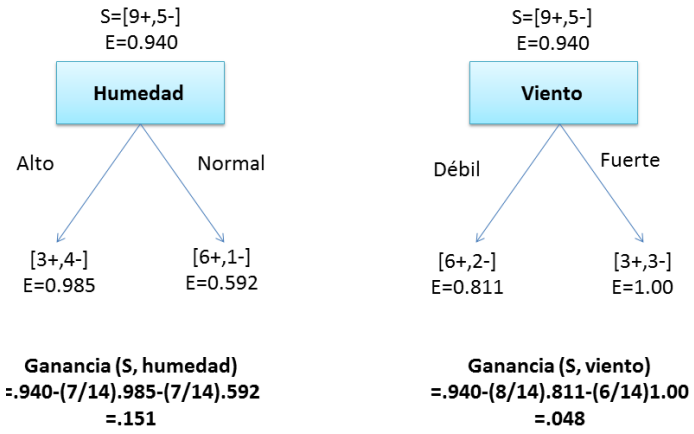


Figura 2.14 Representación gráfica del atributo mejor clasificado. (Mitchel, 1997)

En la figura 2.14, *E* representa la entropía y *S* representa la colección original de ejemplos. Se proporciona un valor inicial a la colección *S* de 9 ejemplos positivos y 5 ejemplos negativos. [9+,5-], clasificando estos valores para el atributo *humedad* que produce una colección de [3+,4-] con (humedad=alto) y [6+,1-] con (humedad=normal). La ganancia de información para esta partición es de 0.151, comparado con la ganancia de solo 0.048 para el atributo *viento*. Es por ello, que el atributo *humedad* proporciona la mayor ganancia de información que *viento*, relativo a la clasificación objetivo.

En general, los árboles de decisión son una herramienta para elegir entre varias alternativas, útiles para la toma de decisiones. Las decisiones pueden estar afectadas por incertidumbre, costos asociados y utilidad. Contienen nodos que representan decisiones, nodos que representan situaciones aleatorias y, finalmente, aparecen las consecuencias de las decisiones. Estas decisiones finales pueden estar asociadas a costos (económicos) o utilidades (otros factores además de los económicos, emocionales, prácticos, etc.). Una manera de mejorar el entendimiento del proceso de toma de decisiones consiste en realizar

un análisis de sensibilidad, es decir, realizar cambios en los parámetros hasta que las conclusiones sean afectadas.

2.6.2 *K*-medias

El algoritmo *K*-medias, creado por MacQueen en 1967 es el algoritmo de agrupamiento o *clustering* más conocido y utilizado ya que su aplicación es simple y eficaz; realiza un aprendizaje no supervisado construyendo grupos o *clusters* donde se clasifican a determinadas instancias que tienen características en común.

El nombre de *K*-medias proviene de la representación de cada uno de los clusters por la media (o media ponderada) de sus puntos, es decir, por su centroide[2]. La representación mediante centroides posee la ventaja de que tiene un significado gráfico y estadístico inmediato. Cada *cluster* por tanto es caracterizado por su centro o centroide que se encuentra en el centro o el medio de los elementos que componen el *cluster*. La figura 2.15, muestra las características de este algoritmo obtenidas de (Tan, Steinbach, & Kumar, 2005).

K-medias tiene por objetivo separar el conjunto de datos en *k clusters* de manera que cada dato pertenezca a un grupo y sólo a uno; busca mediante un método iterativo los centroides (medias, medianas) de los *k clusters* y asigna cada individuo a un *cluster*.

[2] Centroide es la medida de las puntuaciones de la discriminación de un grupo particular. Existen tantos centroides como grupos y un centroide por grupo. Las medias de un grupo en todas las funciones son los centroides del grupo.

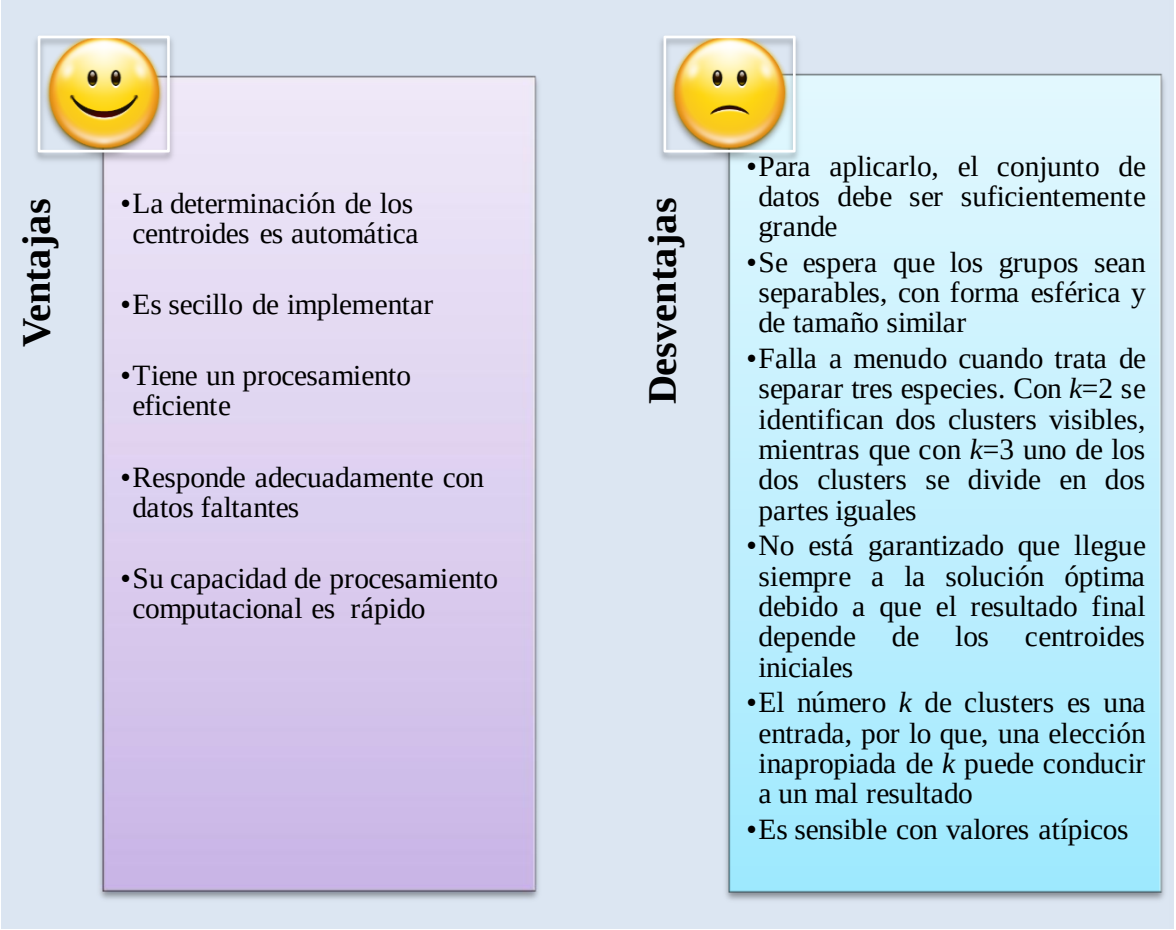


Figura 2.15 Ventajas y desventajas del algoritmo k-medias

Pasos del algoritmo (Tan, Steinbach, & Kumar, 2005)

Paso 1: Se debe seleccionar el valor de K, es decir, el número de grupos o *clusters* en los que se desea dividir las instancias del conjunto de datos. Una vez elegido dicho número, el algoritmo selecciona aleatoriamente las K instancias para continuar con el paso 2

Paso 2: Las instancias seleccionadas en el paso1 son consideradas los centroides de los *clusters* y el resto de instancias forma parte del conjunto de datos y se clasifican como pertenecientes a la clase del centroide más cercano

Paso 3. Una vez creada la primera aproximación de lo que será los *clusters* definitivos, el centroide es modificado eligiendo a la instancia que ocupa el lugar central como nuevo centroide.

Paso 4. En este paso se hace una repetición del paso 2 y paso 3 hasta que se alcanza un equilibrio y no existan más modificaciones en la estructura y composición de los grupos.

El algoritmo se considera que ha alcanzado la convergencia cuando en una iteración no se produce ningún cambio, o se cumple un *criterio de parada*. La figura 2.16, muestra esta secuencia de iteraciones[3].

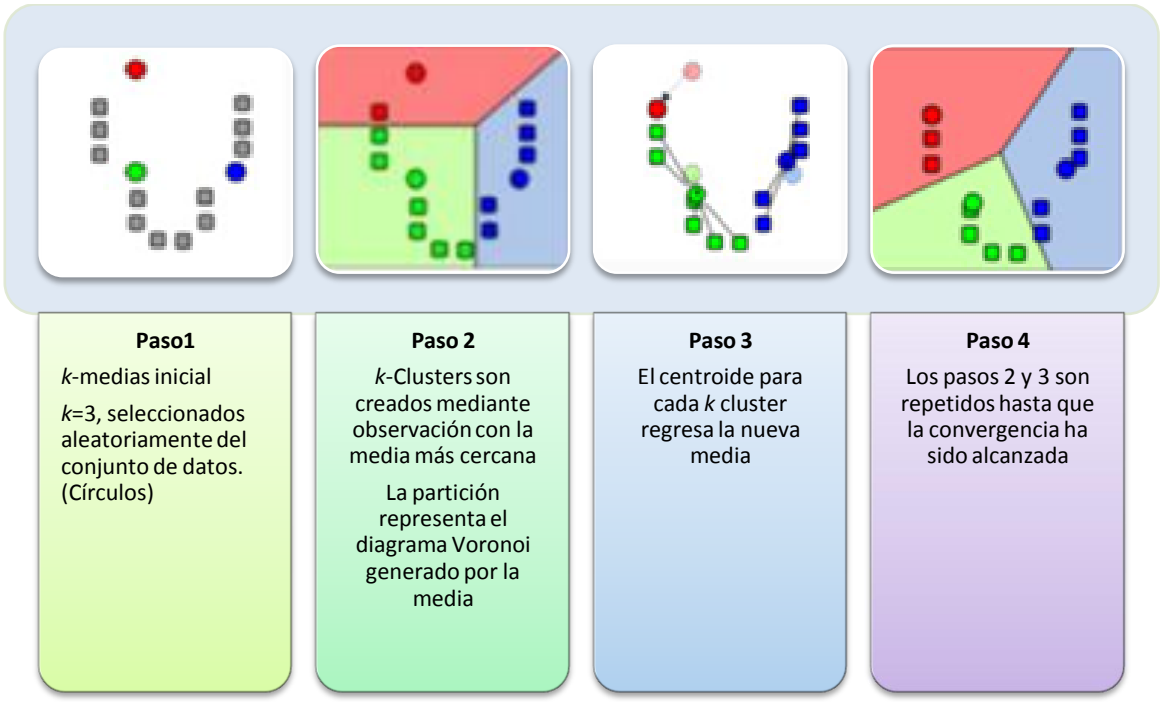


Figura 2.16 Funcionamiento estándar del algoritmo *k*-medias.

[3] Diagrama de Voronoi. Son estructuras fundamentales dentro de la geometría computacional; pues en ellos se almacena toda la información referente a la proximidad entre puntos. Básicamente, es una construcción de teoría de grafos en la que se crean las regiones con menor distancia a ciertos puntos; se genera a partir de las mediatrices de los puntos.

2.7 Evaluación del desempeño de algoritmos

Un algoritmo debe ser analizado para determinar el uso de los recursos que emplea y sobre todo su desempeño para realizar alguna tarea (clasificar, reconocer, identificar, agrupar, categorizar, entre otros). Existen varias medidas para estimar el desempeño de un algoritmo y este desempeño dependerá de qué medida se esté empleando como prioridad. Por ejemplo, la prioridad podría ser la salida del algoritmo, la minimización de la memoria, entre otras.

El criterio más obvio para estimar el rendimiento de un clasificador es su exactitud predictiva en instancias que no se ven. El número decasosque no se ven, algunas veces es potencialmentegrande(si no es que infinita), por lo tanto,una estimacióndebeser calculadoen unconjunto de pruebas. A esto se le conocecomúnmente comola validación cruzada (M.P. van der Aalst, 2011).

La validación cruzada (*cross-validation*) es una técnica empleada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición de datos de entrenamiento y prueba. Razón por la cual es el tema principal de este apartado, mismo que se describe a detalle a continuación.

2.7.1 Matriz de confusión

Unamatriz de confusión, también llamada matriz de predicción o de clasificación, es una herramienta de visualización que se emplea en el aprendizaje supervisado; para obtener información sobre las clasificaciones reales y predicciones realizadas por un sistema de clasificación (Bird, Klein, & Loper, 2009).

La matriz de confusión es una tabla donde cada celda [*i,j*] indica cómo se clasificó algú dato con respecto a su clase pre-establecida (véase tabla 2.2).

Los casos bien clasificados se encuentran en las entradas de la diagonal (ejemplo, las celdas [*i,j*] de la tabla 2.2) pues los grupos pronosticados y reales son los mismos; los elementos

fuera de la diagonal se encuentran mal clasificados. La suma de los elementos de la diagonal dividida entre el total de casos representa la proporción de aciertos (Witten & Frank, 2005), (Montero Lorenzo, 2007), (Bird, Klein, & Loper, 2009) y (Hamilton H. , 2009).

		PREDICCIÓN	
		Negativo	Positivo
ACTUAL	Negativo	***a***	***b***
	Positivo	***c***	***d***

Tabla 2.2 Matriz de confusión cuando se tienen dos posibles resultados de clasificación: Negatigo y Positivo.

En la tabla 2.2, la columna de la matriz representa el número de predicciones de cada clase, mientras que la fila representa las instancias en la clase real. Donde:

- *a* es el número de predicciones correctas cuando una instancia es negativa (VP)
- *b* es el número de predicciones incorrectas cuando una instancia es positiva (FP)
- *c* es el número de predicciones incorrectas cuando una instancia es negativa y (FN)
- *d* es el número de predicciones correctas cuando una instancia es positiva (NV)

En otras palabras, la matriz de confusión presenta cuatro posibles resultados obtenidos de una sola predicción para un casode dos clasesconclases "1" ("sí")y "0" ("no"), que son: verdaderospositivos (VP), negativos verdaderos (NV), falsos positivos (FP) yfalsos negativos (FN). Un falso positivoescuando el resultadose clasificaincorrectamente como"sí" (o "positivo"), cuandoen realidad es"no" (o "negativo"). Un falsonegativo escuando el resultadose clasificaincorrectamente comonegativo cuandoen realidad espositiva.Los verdaderospositivos y verdaderos negativosson, evidentemente,las clasificacionescorrectas (Freitas, 2002) y (Bird, Klein, & Loper, 2009).

Las métricas frecuentemente usadas, y que se obtienen de la matriz de confusión son: Exactitud, Precisión, Recuerdo, Medida F, cuya descripción fueron obtenidas de (Freitas, 2002)(Witten & Frank, 2005), (Bird, Klein, & Loper, 2009) y (M.P. van der Aalst, 2011) y

se describen a continuación. Considere la tabla 2.2, para cada una de las ecuaciones presentadas.

- **Exactitud (*Accuracy*):** Es la proporción del número total de predicciones que son correctas. Se determina aplicando la ecuación (AC).

$$AC = \frac{a + d}{a + b + c + d}$$ Ecuación AC

En otras palabras, la exactitudmidela fracciónde casos en la diagonal de la matriz de confusión. La ecuación AC puede no ser una medida de rendimiento adecuado cuando el número de casos negativos es mucho mayor que el número de casos positivos.

- **Precisión (*Precision*):** Algunas veces llamada consistencia o confidencia. Es la proporción de la predicción de los casos positivos correctos. Se calcula aplicando la ecuación *P*.

$$P = \frac{d}{b + d}$$ Ecuación *P*

- **Recuperación (*Recall*):** También llamado Verdadero positivo, completitud o sensibilidad. Se calcula empleando la ecuación *R*.

$$R = \frac{a}{p}$$ Ecuación *R*

Donde *p* se puede interpretar como el número de casos que deberían haber sido recuperados en función de algunos criterios de búsqueda. Es posible tener una alta precisión y baja recuperación, cuando algunos de los casos que han buscado son devueltos por la consulta, pero los que se devuelven son relevantes. También es posible tener una recuperación alta y baja precisión, cuando se devuelven muchos casos (incluidos los más relevantes), pero también se devuelven muchos casos irrelevantes.

- **Medida F (*F-Measure*):** Es una medida que combina la precisión con la recuperación para dar una puntuación única. Se define con la ecuación *MF*.

$$MF = \frac{2\ x\ Precisión\ X\ Recuperación}{Precisión + Recuperación}$$ Ecuación *MF*

2.7.2 Validación cruzada con k iteraciones

Como ya se mencionó, la validación cruzada es una técnica que se emplea para evaluar modelos. Se tiene un conjunto de datos que se divide enun conjunto de entrenamientoy un conjuntode pruebas.Elconjunto de entrenamientose utiliza paraobtenerun modelo, mientras que el conjunto de pruebasse utiliza para evaluareste modelobasado enejemplosque no se consideraron; de tal forma que la función de aproximación sólo se ajusta al conjunto de datos de entrenamiento y a partir de allí se calculan los valores de salida para el conjunto de datos de prueba.

La figura 2.17 muestra el procedimiento para aplicación de la validación cruzada, mismo que se describe a continuación.

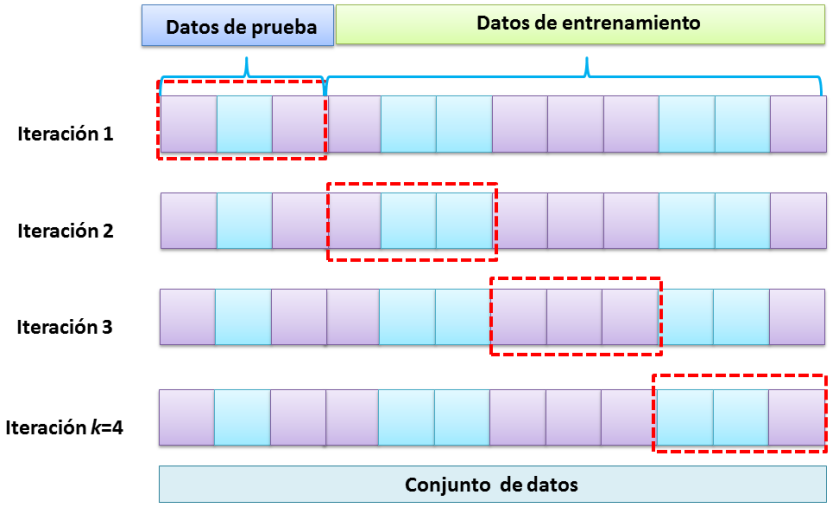


Figura 2.17 Validación cruzada de K iteraciones con *K*=4.

El procedimiento que se muestra en la figura 2.17 es el siguiente:

1. Dividir la muestra en *k* particiones (manualmente o aleatoriamente)
2. Formar el conjunto de datos de entrenamiento con k-1 particiones, el conjunto de pruebas se forma con los valores restantes
3. Repetir el proceso hasta *k* iteraciones, dejando como control cada una de las particiones
4. Realizar la media aritmética de los resultados de cada iteración para obtener un único resultado

Este método es muy preciso puesto que se evalúa a partir de *k* combinaciones de datos de entrenamiento y de prueba. Su principal desventaja, es que es lento desde el punto de vista computacional. En la práctica, la elección del número de iteraciones depende de la medida del conjunto de datos. Lo más común es manipular la validación cruzada de 10 iteraciones (10-*fold cross-validation*), si la muestra es muy grande ($k>10$) entonces $k=3$, y si la muestra es muy pequeña, entonces se toma el valor máximo de *k*.

2.8 Experimentación con Weka

Weka es un ave no voladora con una naturaleza inquisitiva que solamente habita en las islas de Nueva Zelanda. Esta ave da nombre a una herramienta de software que "*es una colección de algoritmos de aprendizaje automático para realizar tareas de minería de datos. Los algoritmos se pueden aplicar directamente a un conjunto de datos o llamadas desde su propio código JAVA. Contiene herramientas para los datos de pre-procesamiento, clasificación, regresión, conglomerados, reglas de asociación y visualización. También es adecuado para el desarrollo de nuevos sistemas de aprendizaje de máquina"* (The University of Waikato, 2013).

Sus características principales son:

- Tiene licencia GNU (*PublicLicense*) con libre distribución y difusión.

- Está programado en JAVA lo que la hace independiente de la arquitectura, pues funciona en cualquier plataforma sobre la que exista una máquina virtual JAVA disponible

La figura 2.18 muestra el selector de interfaces de Weka, las cuales son: Explorador (Explorer), Interfaz Simple de línea de comandos (*Simple Cli)*, Experimentador (*Experimenter)* y Flujo de conocimiento (*Knowledgeflow*), mismos que se describen a continuación.

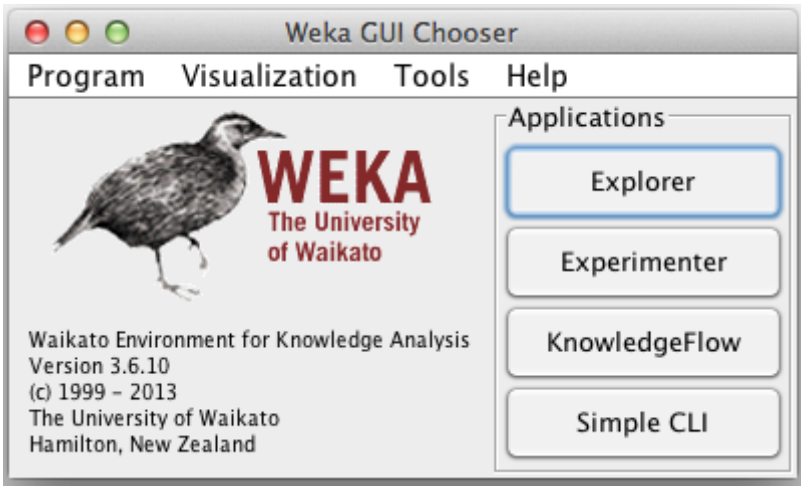


Figura 2.18 Selector de interfaces Weka. (The University of Waikato, 2013)

1. **Explorador**
    a. La interfaz en modo Explorador es el comúnmente usado; permite realizar operaciones sobre un solo archivo de datos y tareas como: pre-procesamiento de datos y aplicación de filtros, clasificación, agrupamiento, búsqueda de asociaciones, selección de atributos y visualización de datos.
2. **Interfaz Simple de Línea de Comandos**
    a. Proporciona una consola para poder introducir instrucciones. Considerada actualmente como solamente una herramienta de pruebas.

3. **Experimentador**
    a. Es útil para aplicar uno o varios métodos de clasificación sobre un gran conjunto de datos, con el fin de realizar comparativos estadísticos entre ellos y obtener otros índices estadísticos.
4. **Flujo de conocimiento**
    a. Muestra de una forma explícita el funcionamiento interno del programa. Su funcionamiento es gráfico, sitúa en el panel de trabajo elementos base de permita crear un "circuito" que defina el experimento desarrollado

Es importante comentar que para la experimentación de este trabajo de investigación, se empleará la interfaz **"Explorador"** del cual se emplearán las tareas de pre-procesamiento, clasificación y agrupamiento.

Weka trabaja con archivos en un formato denominado arff, acrónimo de *Attribute-Relation File Format*. Este formato está compuesto por una estructura de tres partes:

1. **Cabecera.** Se define el nombre de la relación. Su formato es el siguiente:

```
@relation<nombre-de-la-relación>
```

Donde <nombre-de-la-relación> es de tipo *String*. Si dicho nombre contiene algún espacio será necesario expresarlo entrecomillarlo.

2. **Declaraciones de atributos.** En esta sección se declaran los atributos que describen el conjunto de datos aunado a su tipo. La sintaxis es la siguiente:

```
@attribute<nombre-de-la-relación><tipo>
```

Donde <nombre-del-atributo> es de tipo *String* teniendo las mismas restricciones que el punto anterior. Weka acepta diversos tipos de datos como: *numeric*, *integer*, *date*, *string*, enumerado.

3. **Sección de datos.** Se declaran los datos que componen la relación separando entre comas los atributos y con saltos de línea las relaciones, tal y como sigue.

```
@data
4, 3.2
```

Aunque éste es el modo “completo" es posible definir los datos de una forma abreviada (*sparse data*). Si se tiene una muestra en la que hay muchos datos que sean 0; estos datos se pueden expresar prescindiendo de los elementos que son nulos, rodeando cada una de las filas entre llaves y situando delante de cada uno de los datos el número de atributo.

Un ejemplo de esto es el siguiente

```
@data
{1 4, 3 3}
```

En este caso se ha prescindido de los atributos 0 y 2 (como mínimo) y se asigna al atributo 1 el valor 4 y al atributo 3 el valor 3.

En el caso de que algún dato sea desconocido se expresará con un símbolo de cerrar interrogación (“?"). Es posible añadir comentarios con el símbolo “%”, que indicará que desde ese símbolo hasta el final de la línea es todo un comentario. Los comentarios pueden situarse en cualquier lugar del archivo.

2.9 Revisión de la literatura

Como se ha visto en capítulos anteriores, la minería de datos y el aprendizaje automático ha tenido diversas aplicaciones en diferentes áreas del conocimiento como: Medicina, robótica, astronomía, visión por computadora, etc.

No obstante, el objetivo de esta sección es presentar las aplicaciones del aprendizaje automático en el ámbito educativo, exclusivamente. En el anexo 3: Producción científica se encuentra la información completa de lo presentado en esta sección.

Autores: Minaei-Bidgoli, Kashy, Kortemeyer, & Punch (2003)
Algoritmo empleado: Clasificador bayesiano, 1-NN, k-vecinos más cercanos, red neuronal artificial, *parzen-window* y árboles de decisión
Aplicación o algoritmo a resolver: Clasificar a los alumnos con el fin de predecir su nota final correspondiente en las características extraídas de los datos registrados en un sistema académico basado en la web
Hallazgos: K-vecinos más cercanos obtuvo la mejor precisión cuando se experimentó con dos clases. CART (árboles de clasificación y regresión) obtuvo mejor precisión para experimentos de 3 a 9 clases

Autores y año: Thomas &Galambos (2004)
Algoritmo empleado: CHAID (**CH**i-squared**A**utomatic**I**nteraction**D**etection) Chi cuadrada de Detección Automática de Interacción.
Aplicación o problema a resolver: Conocer la satisfacción de los estudiantes en tres ámbitos: experiencias académicas, la integración social y el campus de servicios-instalaciones.
Hallazgos: Se obtuvieron resultados importantes que ayudaron a la toma decisiones en esos ámbitos.

Autores y año: Kotsiantis, Pierrakeas, &Pintelas (2004)
Algoritmo empleado: C4.5, backpropagación, NaiveBayes, 3-NN, regresión logística y máquina de vectores de soporte
Aplicación o problema a resolver: Predecir el comportamiento de los nuevos estudiantes de la ingeniería informática.
Hallazgos: El algoritmo naiveBayes es el más adecuado para construir una herramienta de apoyo basada en software para predecir el comportamiento de los nuevos estudiantes. Al

mismo tiempo, fue el más satisfactorio para obtener una precisión del 72,48% y una sensibilidad global del 78%, además de ser el más fácil de implementar.

Autores: García, Amandi, Schiaffino, & Campo (2005), (2007) y (2008)
Aplicación o algoritmo a resolver: Identificar los estilos de aprendizaje basado en un sistema web.
Hallazgos: Se obtuvo un modelo que permite agregar más dimensiones de aprendizaje
Infiere el estilo de aprendizaje basado en el modelado de su comportamiento
Ayudar de forma proactiva al sugerir cursos personalizados a los estudiantes.

Autores y año: Acevedo Orduña, Caicedo Bravo, & Loaiza Correa (2007)
Algoritmo empleado: De Levenberg Marquardt de red de tipo perceptrón multicapa
Aplicación o problema a resolver: Optimizar los procesos de administración de personal de la Armada de la República de Colombia con el fin de disminuir los niveles de la subjetividad en la selección de ingreso
Hallazgos: La exactitud de predicción aumenta a medida que la red neuronal es entrenada con más datos para abordar los problemas psicosociales de alta complejidad.
También destacan la eficiencia de las redes neuronales de base radial para resolver problemas de clasificación de patrones.

Autores: Márquez, Ortega, González-Abril, & Velasco (2008)
Algoritmo empleado: Colonia de hormigas y red bayesiana
Aplicación o algoritmo a resolver: Encontrar la ruta para adquirir una competencia profesional mediante un LMS
Hallazgos: Se obtuvo una ruta que se adapta a las preferencias y necesidades de los estudiantes, considerando el aprendizaje de la unidad pedagógica y el comportamiento social de los estudiantes en el LMS

Autores y año: Oladokun, Adebanjo, & Charles-Owaba (2008)
Algoritmo empleado: Red neuronal artificial

Aplicación o problema a resolver: Predecir el rendimiento de un posible candidato para la admisión a la Universidad.

Hallazgos: Demostraron el potencial de las redes neuronales artificiales para mejorar la eficiencia del sistema de la Universidad para poder acceder.

No todas las características de rendimiento de la escuela se pueden obtener a partir del registro de pre-admisión y por esta razón, que sugieren que la aplicación de una entrevista oral puede mejorar el modelo real.

Autores y año: Karamouzis, Stamos T.; Vrettos, Andreas (2008)

Algoritmo empleado: Red neuronal artificial

Aplicación o problema a resolver: Predecir el número de estudiantes a graduarse en la universidad

Hallazgos: Emplear una red para predecir la tasa de graduados. Ayuda a los administradores a conocer las razones por las cuales los estudiantes no egresan.

Autores y año: Baylari&Montazer (2008)

Algoritmo empleado: Red neuronal artificial

Aplicación o problema a resolver: Construir un sistema de *e-learning*multi-agente basado en la Teoría de Respuesta al Ítem personalizada

Hallazgos: El agente propuesto podría recomendar los materiales personalizados según el curso con una precisión del 83,3%.

La capacidad de aprendizaje adaptativo basado en la evidencia puede acelerar la eficiencia y la eficacia del aprendizaje.

Autores y año: Ranjan&Khalil (2008)

Algoritmo empleado: Árboles de decisión, clasificación y regresión logística

Aplicación o problema a resolver: Búsqueda de patrones de ¿cómo los estudiantes interactúan entre sí? ¿Cómo se realiza el proceso de admisión?, ¿cuáles son los mecanismos de asesoramiento y cómo elegir los cursos?

Hallazgos: Se identificó que la minería de datos es útil para que los maestros organicen sus clases y cursos, para entender sus estilos de aprendizaje y promover la proactividad. Concluyeron que la minería de datos es útil para predecir el éxito de los programas educativos y útiles para el aprendizaje académico.

Autores y año: Vialardi, Bravo, Shafti, &Ortigosa (2009)

Algoritmo empleado: C 4.5

Aplicación o problema a resolver: Ayudar a los estudiantes a tomar mejores decisiones acerca de cuántos y cuáles cursos deben inscribirse

Hallazgos: La recomendación que se hace está relacionada con su rendimiento académico global o de ciertos cursos y por lo tanto, el estudiante pude decidir libremente basado en ello. La información analizada se podría utilizar como entrada para una eventual modificación al currículo.

Autores y año: KumarBaradwaj& Pal (2011)

Algoritmo empleado: ID3

Aplicación o problema a resolver: Extraer un conjunto de datos académicos para evaluar el rendimiento estudiantil

Hallazgos: Predecir el rendimiento de los estudiantes la final del semestre. Así como, identificar a los estudiantes que requieren atención especial para reducir el índice de fracaso escolar y tomar medidas adecuadas para el próximo examen semestral.

Autores y año: Ayinde, Adetunji, Bello, &Odeniyi (2013)

Algoritmo empleado: NaiveBayes y DecisionStump

Aplicación o problema a resolver: predecir y clasificardatos académicos relacionados con la evaluación del desempeño o para estudiar el comportamiento de los estudiantes.

Hallazgos: Se produjo una lista de predicción para el desempeño de los nuevos estudiantes Identificar a estudiantes que requieren atención especial para un buen desempeño en su disciplina a los largo de sus estudios.

Autores: Yukselturk, Ozekes, &Türel (2014)

Algoritmo empleado: K-vecinos más cercanos, árboles de decisión, naiveBayes y redes neuronales artificiales.

Aplicación o algoritmo a resolver: Para clasificar a los alumnos que abandonan la escuela

Hallazgos: Encontraron que 3-nn y el árbol de decisión son más sensibles.

Predecir la deserción de estudiantes y Predecir con éxito las razones por las cuales los estudiantes abandonan sus estudios.

Referencias del capítulo

Aldiss, B., Watson, I., & Spielberg, S. (2001). Artificial Intelligence A.I. Los Ángeles, California, Estados Unidos.

Álvarez, J., & Zwir, I. (2001). *Aprendizaje Automático (AA).* Obtenido de brains and machines: http://www-2.dc.uba.ar/materias/aa/aa.html#Qu%C3%A9%20significa

Amaya Amaya, J. (2010). *Toma de decisiones gerenciales: Métodos cuantitativos para la administración.* Bogotá, Colombia: ECOE EDICIONES.

ANUIES. (2011). *Programas institucionales de tutoría una propuesta de la ANUIES 3a Edición.* México, D.F.: ANUIES.

Arciniega, S., Del Rosario, M., Calderón, B., & M. L. (2006). *Validez y confiabilidad del estudio socieconómico.* México, DF.: UNAM.

Asimov, I., Silverberg, R., & Kazan, N. (1999). Bicentennial Man. Los Ángeles, California, Estados Unidos.

Berberena González, V. H. (2011). *El Patrón de Lealtad de Clientes: una ventaja competitiva sostenible.* Obtenido de Negociación Comercial: negociacioncomercial .com.mx/archivos/archivo_107.pdf

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python.* USA: O´Really Media, Inc.

Bishop, C. M. (2007). *Pattern recognition and Machine Learning.*Singapore: Springer.

Curto Díaz, J., & Conera i Carat, J. (2010). *Introducción al Business Intelligence.* Barcelona: Editorial UOC.

David, F. R. (1997). *Concepto de Administración Estratégica.*México, D.F.: Pearson Education.

Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms.*The Netherlands: Springer-Verlag.

García Fernández, L. A. (2004). Usos y aplicaciones de la inteligencia artificial. *La ciencia y el hombre, XVII*(3), http://www.uv.mx/cienciahombre/revistae/vol17num3/articulos/inteligencia/.

García Morate, D. (2006). *Weka en castellano.* Obtenido de Diego García Morate: http://www.metaemotion.com/diego.garcia.morate/

Hamilton, H. (2009). *Site Map of Course Notes*. Obtenido de Computer Science 831: Knowledge Discovery in Databases: http://www2.cs.uregina.ca/~dbd/cs831/index.html

Hamilton, L. (2014). *Six Novel Machine Learning Applications.*Obtenido de Forbes: http://www.forbes.com/sites/85broads/2014/01/06/six-novel-machine-learning-applications/

Han, J., Kamber , M., & Pei, J. (2011). *Data mining: concepts and techniques.*Amsterdam: Morgan Kaufmann.

Harvard business essentials. (2006). *Toma de decisiones para conseguir mejores resultados.* EE.UU.: Deusto.

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la investigación.* México, DF.: Mc Graw Hill/Interamericana Editores.

Hill, C. W., & Jones, G. R. (2009). *Administración Estratégica.* México, D.F.: Mc Graw Hill.

Hitt, M. A., Black, J. S., & Porter, L. W. (2014). *Management.* Edinburg Gate, Harlow: Pearson Education Limited.

Kelly, P. K., & Gorín, J. (1999). *Las Tecnicas para la Toma de Decisiones en Equipo: Guia Practica para Obtener Buenos Resultados.*Ediciones Granica SA.

Krogerus, M., & Tschäppeler, R. (2011). *The decision book.* London, UK: Profile Books LTD.

M.P. van der Aalst, W. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes (Google eBook).*London, UK: Springer-Verlag.

Méndez del Río, L. (2006). *Más allá del Business Intelligence: 16 experiencias de éxito.*Barcelona: Gestión 2000.

Mitchel, T. M. (1997). *Machine Learning.* Singapore: Mc Graw Hill.

Mitchell, T. M. (2000). *Decision Tree Learning.* Obtenido de Washington State University: http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CD0QFjAA&url=http%3A%2F%2Fwww.cs.princeton.edu%2Fcourses%2Farchive%2Fspr07%2Fcos424%2Fpapers%2Fmitchell-dectrees.pdf&ei=QAY7U-ngA4Xo2QXw7oCwAQ&usg=AFQjCNGMko1QbJWpXS8K8lzSZ

Monahan , G. E. (2000). *Management Decision Making.* Cambridge, UK.: Cambridge University Press.

Montero Lorenzo, J. M. (2007). *Minería de Datos: Técnicas y herramientas.* Madrid, España: Thomson Ediciones Paraninfo S.A.

Muñoz Pérez, J. (2010). *Inteligencia computacional inspirada en la vida.* Málaga, España: Servicio Publicaciones UMA.

Nuevas Tecnologías. (2011). *Fernández Editores*. Obtenido de La elección de una carrera o un trabajo: http://www.tareasya.com.mx/index.php/tareas-ya/secundaria/formacion-civica-y-etica/el-individuo/1728-Habilidades,-aptitudes-e-intereses.html

Prawda, J. (2004). *Métodos y modelos de investigación de operaciones: modelos determinísticos.* México: Limusa.

Real Academia Española. (2013). *Diccionario de la Lengua Española*. Obtenido de afinidad: http://lema.rae.es/drae/?val=afinidad

Rodríguez, R., & Mislej, E. (2013). *Aprendizaje Automático - Machine Learning.* Obtenido de Departamento de Computación- Universidad de Buenos Aires: http://www.dc.uba.ar/materias/aa/2013/cuat1

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3era. ed.). Prentice Hall.

Sabherwal, R., & Becerra-Fernández, I. (2011). *Business Intelligence: Practices, Technologies and Management.* United States of America: Wiley.

Schmitz, C. (2010). *LimeSurvey*. From The project: https://www.limesurvey.org

Serra de la Figuera, D. (2004). *Métodos cuantitativos para la toma de decisiones.*España: Ediciones Gestión 2000.

Shawe-Taylor, J. (2006). *Machine learning research topics and application field.* Obtenido de Centre for Computational Statistics and Machine Learning: http://www.csml.ucl.ac.uk/courses/msc_ml/?q=node/37

Slade, S. (1994). *Goal-based decision making: an interpersonal model.* New Jersey: Lawrence Erlbaum Associates, Inc. Publishers.

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining.* EUA.: Addison-Wesley.

The University of Waikato. (2013). *Weka 3: Data Mining Software in Java*. Obtenido de Machine Learning Group at the University of Waikato: http://www.cs.waikato.ac.nz/ml/weka/

Tufféry, S. (2011). *Data Minign and Statistics for Decision Making.* United Kingdom: John Wiley & Sons.

Vercellis, C. (2009). *Business Intelligence: Data mining and optimization for decision making.* United Kingdom: Wiley.

Vintar, J., Goldsman, A., & Asimov, I. (2004). I Robot. Los Ángeles, California, Estados Unidos.

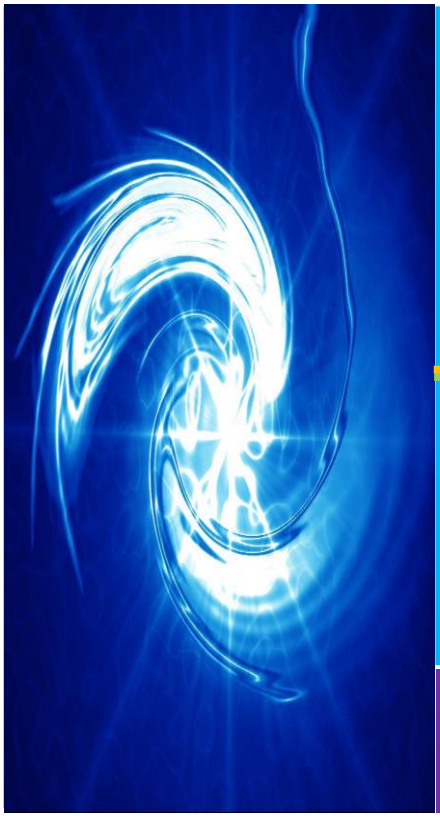
Wikipedia. (2014). *Supervised Learning*. Obtenido de Wikipedia: http://en.wikipedia.org/wiki/Supervised_learning

Wikipedia. (2014). *Unsupervised Learning*. Obtenido de Wikipedia: http://en.wikipedia.org/wiki/Unsupervised_learning

Williams, S., & Williams, N. (2007). *The profit impact of Business Intelligence.*San Francisco, CA.: Elsevier.

Winograd, M., Fernández Lamarra, N., & Farrow, A. (1998). *Herramientas Para la Toma de Decisiones en América Latina Y El Caribe: Indicadores Ambientales Y Sistemas de Información Geográfica.*CIAT.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques.* San Francisco, CA: ELSEVIER.



Capítulo III
Metodología

INTRODUCCIÓN

En este capítulo se define la técnica y tipo de investigación. Así mismo, se describe a la población a partir de la selección de la muestra, se detalla la recolección de los datos, el diseño del instrumento para recolectarlos, su validación y método de aplicación. Finalmente, se puntualiza el diseño del experimento para Weka y la forma en cómo se realizará el análisis de los resultados obtenidos.

3.1. Técnica de Investigación

El diseño de la investigación se basa fundamentalmente en la revisión de la literatura sobre el tema propuesto y se apoya en las áreas de aplicación comunes que tiene al AA. La recolección de esta información nos permitirá hacer un análisis cualitativo y cuantitativo, para proponer la aplicación del AA en la toma de decisiones que potencialicen la inteligencia de negocios como ventaja competitiva en las empresas e instituciones.

3.1.1. Tipo de investigación

El alcance de esta investigación es exploratorio[1](Díaz Narváez, 2009) pues como se observó en el capítulo II, la revisión de la literatura reveló que la aplicación del aprendizaje automático en la inteligencia de negocios es vaga o poco estudiada y con ello se establecen antecedentes para su aplicación en la inteligencia de negocios. Al mismo tiempo, tiene un alcance descriptivo[2](Prieto Herrera, 2013) ya que se busca especificar las competencias y

[1] Es aquella que se efectúa sobre un tema u objeto desconocido o poco estudiado, por lo que sus resultados constituyen una visión aproximada de dicho objeto, es decir, un nivel superficial de conocimiento.
[2] Consiste, fundamentalmente, en caracterizar un fenómeno o situación concreta indicando sus rasgos más peculiares o diferenciadores. Su meta no se limita a la recolección de datos, sino a la predicción e identificación de las relaciones que existen entre dos o más variables.

afinidades particulares de un grupo de personas que permiten enriquecer la productividad cuando se agrupan como equipo de trabajo.

3.2. Selección de la muestra

De acuerdo a Hernández Sampieri, Fernández Collado, & Baptista Lucio, (2010), la muestra **"***es un subgrupo de la población de interés (sobre el cual se recolectaran datos, y que tiene que definirse o delimitarse de antemano con precisión), este deberá ser representativo de la población".*

Para seleccionar una muestra, es prescindible definir la unidad de análisis, es decir, personas, organizaciones, periódicos, comunidades, etc. que se estudiarán. Para definir esta unidad, el sobre quién o sobre qué se van a recolectar datos, se debe considerar el planteamiento del problema a investigar y los alcances del estudio. Para nuestro caso, la unidad de análisis son los tutores y tutorados de la Universidad Politécnica de Puebla.

3.2.1. Población

La población comprende a todos aquellos estudiantes de ingeniería de la UPPuebla inscritos en el cuatrimestre enero-abril 2014 y tutores de todas las carreras activos en el mismo cuatrimestre.

Para seleccionar la muestra de estudiantes se emplea el muestreo probabilístico, puesto que todos los elementos de la población tienen la misma probabilidad de ser elegidos, además de que se emplea una "encuesta" en donde se pretende hacer estimaciones de variables en la población, dichas variables serán medidas por el instrumento de medición H-A (véase figura 3.5).

Para determinar el tamaño de la muestra se emplea la ecuación (E3.1).

$$n = \frac{n\prime}{1+\frac{n\prime}{N}} \qquad \text{(E3.1)}$$

En donde *N*= tamaño de la población, *n'*= tamaño de la muestra sin ajustar. Cabe mencionar que para obtener *n'* se requiere de la ecuación (E3.2).

$$n' = \frac{S^2}{V^2} \qquad \text{(E3.2)}$$

En donde S^2= varianza de la muestra, V^2=varianza de la población al cuadrado. Para calcular ambas variables es necesario aplicar las ecuaciones E3.3 y E3.4.

$$s^2 = p(1-p) \ (E3.3) \qquad V^2 = se^2 \quad (E3.4)$$

En donde *p*=porcentaje estimado de la muestra, es decir, la probabilidad de ocurrencia del fenómeno; *se*= error estándar determinado por el investigador.

Para obtener el tamaño de la muestra de estudiantes se realiza el procedimiento mostrado en la tabla 3.1.

Es preciso comentar que los datos fueron obtenidos de la Unviersidad Politécnica de Puebla (2013).

***N*= 1199** ***se*= 5% (suele ser lo convencional)** **Nivel de confianza= 95%** **A aplicar todas las ecuaciones anteriores se obtiene:**		
$(\boldsymbol{E5.3})$	$s^2 = p(1-p)$	$s^2 = 0.95(1-0.95) = 0.0475$
$(\boldsymbol{E5.4})$	$V^2 = se^2$	$V^2 = (0.05)^2 = \ 0.0025 \ \ (E3.4)$
(E3.2)	$n' = \frac{s^2}{V^2}$	$n' = \frac{0.0475^2}{0.0025^2} = \frac{0.00225}{0.00000625} = 360$
(E.3.1)	$n = \frac{n'}{1 + \frac{n'}{N}}$	$n = \frac{360}{1+\frac{360}{1199}} = \frac{360}{1.30025} = 276.8 \ \ (redondeando\ 277\ casos)$

Tabla 3.1 Obtención de la muestra para estudiantes.

Entonces, se tendrán que encuestar a 277 estudiantes; y dado que la muestra de tutores es pequeña (35 tutores) se encuestarán a un porcentaje superior al 50%.

3.3. Recolección de datos

La recolección de datos implica elaborar un plan detallado de procedimientos que conduzcan a reunir datos con un propósito específico. Este plan se muestra en la figura 3.1.

Figura 3.1. Plan para la recolección de datos

Para recolectar datos existe un sinnúmero de técnicas y estrategias como: entrevistas, encuestas, observación, sesiones de grupo, diagramas de flujo, cuestionarios, etc.; incluso existe un gran número de cuestionarios elaborados previamente para aplicarlos sin requerimiento de validarlos. No obstante, dadas las características de este estudio y después de realizar una extensa búsqueda infructuosa sobre instrumentos similares al objeto de estudio, se desarrolló un instrumento propio, mismo que se describe en el siguiente tema.

3.3.1. Diseño del instrumento

La estrategia empleada para el diseño del instrumento, se ilustran en la figura 3.2 y que se describen a continuación.

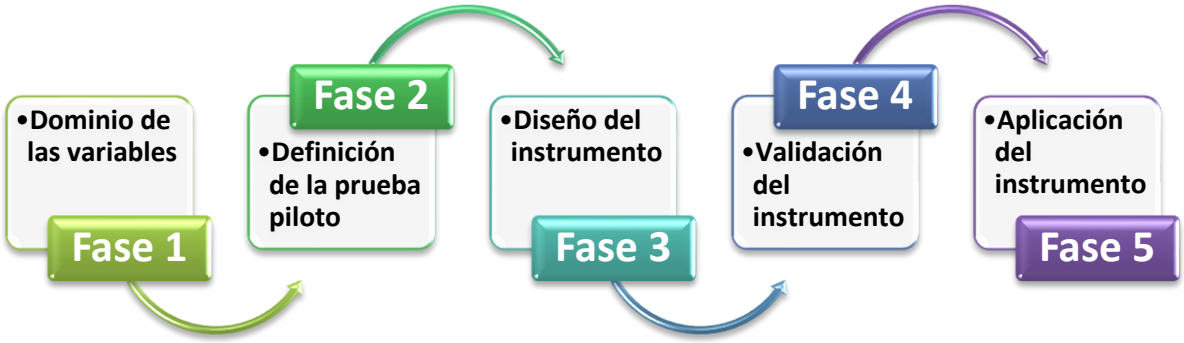


Figura 3.2. Metodología empleada para la creación y validación del instrumento propuesto. Fuente: Propia

FASE 1: Esta fase consiste en la identificación del dominio de las variables a medir y sus indicadores a partir de lo descrito por Hernández Sampieri, Fernández Collado, & Baptista Lucio (2010).

FASE 2: Previo a la validación del instrumento, se efectuará una prueba piloto que consiste en aplicar el instrumento a un grupo de individuos. Para elegir dicho grupo, los criterios fueron: licenciaturas de la Universidad Virtual de la UNACH que tuviesen más de 3 tutorados y al menos 2 tutores.

FASE 3: En esta fase se construye el instrumento completo que implica la generación de todos los *ítems*, así como la definición del procedimiento de aplicación. Este último consiste de los siguientes pasos. 1) Se elige una estrategia que permita obtener datos de manera rápida y de fácil tratamiento de los resultados. Para ello, se sugiere el empleo de alguna herramienta de software vía web que brinde a los participantes tiempo suficiente para su contestación, así como la libre elección del espacio físico para ello. Para efectos prácticos de esta investigación se emplea *LimeSurvey*[3] (Schmitz, 2010). 2) Previo a la aplicación se les envía a tutorados y profesores una invitación vía correo electrónico

[3]Es una aplicación *open source* para la aplicación de encuestas en línea, que posibilita a usuarios sin conocimientos de programación, desarrollar, publicar y recolectar fácilmente las respuestas de dichas encuestas. https://www.limesurvey.org/en/

indicando el propósito del estudio, importancia de la participación, agradecimiento, tiempo aproximado de respuesta, identificación de quienes lo aplican y una cláusula de confidencialidad, de acuerdo con Hernández Sampieri, Fernández Collado, & Baptista Lucio, (2010).

FASE 4: Esta fase valida el instrumento. El método de consistencia interna basado en el Alfa de *Cronbach* permite estimar la confiabilidad de un instrumento de medición a través de un conjunto de *ítems* que se espera midan el mismo instrumento. Ésta métrica asume que los *ítems*, medidos en escala de Likert, miden un mismo instrumento y están altamente correlacionados (Arciniega, Del Rosario, Calderón, & Ma., Validez y confiabilidad del estudio socieconómico, 2006). Cuanto más cerca se encuentra el valor del Alfa a 1, mayor es la consistencia interna de los *ítems* analizados. Esta medida tiene la ventaja de evaluar cuánto mejoraría la confiabilidad de la prueba si se excluyera un determinado *ítem*.

3.3.2. Validación del instrumento

Hernández Sampieri, Fernández Collado, & Baptista Lucio (2010) afirman que un instrumento de medición es aquel que registra datos observables que representan conceptos o variables que se tienen en mente. Toda medición o instrumento de recolección de datos debe reunir tres requisitos esenciales: confiabildiad, validez y objetividad.

La ***confiabilidad*** de un instrumento de medición se refiere al grado en que su aplicación repetida al mismo sujeto u objeto produce resultados iguales. La ***validez***[4] se refere al grado en que un instrumento realmente mide la variable que se pretende medir y finalmente, la ***objetividad*** se refiere al grado en que el instrumento es permeable a la influencia de sesgos y tendencias del investigador o investigadores que lo administran, califican e interpretan (Hernández Sampieri, Fernández Collado, & Baptista Lucio, Metodología de la investigación, 2010).

[4] Validez total=validez de contenido + validez de criterio + validez de constructo

Es importante comentar que para obtener una validez total es necesario tener validez de contenido, validez de criterio, y validez de constructo. En donde la validez de contenido se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide. La validez de criterio se establece al validar un instrumento comparado con algún criterio externo que pretende medir lo mismo; en tanto que la validez de constructo debe explicar el modelo teórico empírico que subyace a la variable de interés. Para el caso del instrumento diseñado solamente se realiza una validez genérica, misma que se describe más adelante.

En consecuencia, tras la búsqueda de información relacionada a los métodos de validación de instrumentos de medición y de acuerdo con Cervantes H. (2005) y Ledesma, Molina Ibañez, & Valero Mora (2002), el método de consistencia interna Alfa de *Cronbach* es el método más empleado para validar un instrumento, además que sistemas como SPSS, Statistica o SAS lo incluyen dentro de sus opciones de análisis. Razón por la cual, dicho método también se emplea para validar el intrumento de medición propuesto para encontrar las habilidades y afinidades en tutores y tutorados de la UPPuebla.

El cálculo del valor de Alfa de *Cronbach* se realiza mediante la siguiente ecuación:

$$\alpha = \frac{K}{K-1}\left[1 - \frac{\sum V_i}{V_t}\right] \qquad \ldots (1)$$

dondeα= Alfa de *Cronbach*, k= Número de ítems, V_i= Varianza de cada ítem, V_t= Varianza total.

Los resultados obtenidos de la validación del instrumento se detallan en Urbina Nájera, de la Calleja, Vega Lebrún, López Maldonado, & Pico González (2014).

3.3.3. Descripción del instrumento

Con base en los objetivos que se persiguen, se describe a continuación cada una de las variables consideradas para identificar las habilidades y afinidades para construir el instrumento deseado (véase tabla 3.2).

Variable a medir: Demográfica

Definición conceptual: De acuerdo con Hernández Sampieri, Fernández Collado, & Baptista Lucio (2010) este tipo de variable, también llamada de ubicación del sujeto encuestado, se considera obligatoria en todos los tipos de instrumentos a diseñar. No obstante, se debe analizar las que resulten pertinentes y útiles para efectos de la investigación a realizar. En este caso, las preguntas asociadas a esta variable son: sexo, edad, estado civil, si trabaja; dado que ayudará a determinar el rango de edad tanto de estudiantes como de profesores y una posible afinidad dado su sexo, estado civil y trabajo.

Variable	Categoría	Indicador	Opción de respuesta
Demográfica	Datos generales	Identificación de las características demográficas de la población	Diversa
Competencias	Comunicativa	Grado en el que el estudiante/profesor percibe que posee estas competencias	Escala Likert • Totalmente de acuerdo • De acuerdo • Ni en acuerdo ni en desacuerdo • En desacuerdo • Totalmente en desacuerdo
	Intrapersonal		
	Interpersonal		
	Autodirección		
	Digital		
Autonomía en el proceso de aprendizaje	Capacidades y responsabilidades	Grado en el que el estudiante / profesor percibe que posee autonomía en su aprendizaje	
Afinidades e intereses	Emoción o gusto por algo	Grado en el que el estudiante/profesor percibe que posee cierta emoción o gusto por algo	

Tabla 3.2 Clasificación de variables

Variable a medir: Competencias

Definición conceptual: Este tipo de variable pretende estimar el nivel de habilidades que el encuestado tiene en función de las habilidades personales según las categorías que mencionan en común Cruz Bejarán ,2010;Sanz de Acedo Lizarraga, 2010; Repetto & Beltrán, 2009; Vaello Orts, 2009;Cabrera Dokú & González F. ,2006; y Bautista, Borgues, & Forés, 2006; a saber: competencias interpersonales, competencias intrapersonales, competencias comunicativas, competencias de autodirección y competencias digitales.

Variable a medir: Autonomía en el proceso de aprendizaje

Definición conceptual: De acuerdo aRaz, 1986; citado porBenson, 2007; la autonomía es la creación de un mundo propio sin estar sujetos a la voluntad de otros. Es por ello que la autonomía en el proceso de aprendizaje se puede concebir según Rué (2009) como el acto que el estudiante refleja para dar una respuesta a las demandas específicas de conocimiento, escogiendo por sí mismo sólo aquellas condiciones que estime necesarias para obtener dicha respuesta. Es decir, la autonomía se atribuye a las condiciones y no al resultado de aprendizaje.

Variable a medir: Afinidades e intereses

Definición conceptual: Afinidad es la atracción o adecuación de caracteres, opiniones, gustos, etc. que existe entre dos o más personas (Real Academia Española, 2013), mientras que los intereses personales consisten en los gustos o inclinaciones por actividades, personas u objetos; en función de factores sociales, culturales, académicos e incluso propios de la edad (Nuevas Tecnologías, 2011). En la tabla 3.3 se muestra el instrumento H-A completo.

Categoría	Ítem
Demográfica	1. Edad
	2. Sexo
	3. Estado civil
	4. Trabaja
Competencia comunicativa	5. Tengo buena construcción gramatical para redactar reportes y ensayos 6. Tengo habilidad de presentación, discusión y argumentación 7. Represento fácilmente mis ideas con diagramas, presentaciones, entre otros.
Competencia interpersonal	8. Explico e interpreto fácilmente la realidad 9. Comparto mis ideas con otros de manera sencilla 10. Produzco ideas originales que permiten crear e innovar 11. Aplico fácilmente conceptos, valores y herramientas en la realidad natural o social 12. Propongo alternativas para solucionar problemas y selecciono fácilmente las opciones viables 13. Enfrento problemas y los supero con facilidad
Competencia de autodirección	14. Considero que tengo autonomía intelectual y moral 15. Realizo actos con responsabilidad, ética, social y ambiental
Competencia intrapersonal	16. Demuestro de manera oral, escrita o física las cualidades propias 17. Defino necesidades de aprendizaje y busco satisfacerlas con el máximo provecho y mínimo esfuerzo 18. Desarrollo un plan con objetivos y metas realistas para alcanzarlos
Competencia digital	19. Utilizo recursos tecnológicos e informáticos para facilitar mi aprendizaje 20. Aprovecho recursos digitales de forma responsable, pertinente y ágil en mi actividad académica 21. Encuentro información de interés de forma rápida y sencilla
Autonomía en el proceso de aprendizaje	22. Organizo mi tiempo adecuadamente para realizar actividades académicas 23. Tengo la capacidad de automotivarme a pesar de la dificultad de ciertas actividades 24. Tengo responsabilidad y compromiso para realizar cada una de las actividades planeadas 25. Tengo una mente abierta para las diferentes opiniones que se generan en el grupo 26. Realizo una autoevaluación constante de mis progresos 27. Aplico constantemente estrategias para evitar mis deficiencias
Afinidades e intereses	28. Disfruto de ir al cine / teatro / museo 29. Practico o veo algún deporte 30. Asisto frecuentemente a espectáculos 31. Me distraigo viendo televisión o escuchando música 32. Disfruto de participar frecuentemente en redes sociales

	33. Opto por salir a cenar a un restaurante o bar 34. Disfruto pasear por el parque / jardín / avenida 35. Me distraigo cocinando 36. Me gusta ir al spa / gimnasio 37. Visito frecuentemente a un familiar o amigo

Tabla 3.3 Instrumento de medición H-A

3.3.4. Aplicación del instrumento

La aplicación del instrumento H-A se realizó en dos etapas:

1) **Prueba piloto.** Se administró el instrumento H-A a una muestra pequeña, misma que se describió en la fase 2 del tema 3.3.1. con el fin de determinar la confiabilidad del inicial y de ser posible, la validez del instrumento.

    - La primera versión del instrumento conformada por 31 reactivos se aplicó a este grupo de individuos y se realizaron ajustes sobre los resultados; posteriormente, se generó una segunda versión conformada por 39 reactivos administrada a este mismo grupo; se realizaron cambios, hasta obtener una versión final de 37 reactivos (compuesto por 33 *ítems* divididos en dos grupos: 18 ítems para identificar habilidades y 15 para identificar afinidades; más 4 *ítems* que identifican las características demográficas de los encuestados). Para obtener la participación completa en la prueba piloto; se envió una invitación vía correo electrónico para tutores y tutorados. El instrumento fue contestado por 16 participantes; 13 tutorados y 3 tutores.

2) **Versión final.** El instrumento H-A preliminar se modifica, ajusta y mejora (se quitan o agregan *ítems*, se cambian palabras, se otorga más tiempo para responder, etc.). Se obtendrá la versión final para administrar a la muestra completa de sujetos de estudio.

Cabe mencionar que por cuestiones de prontitud, la prueba piloto se aplicará en línea y la versión final será aplicada en papel.

Como resultado de la prueba piloto, se obtuvieron resultados favorables al identificar que la pregunta 28 y 32 significan lo mismo, por lo que se mejoró la redacción para hacer una sola pregunta. Así mismo, se redefinieron las preguntas 3, 14, 16, 19 y 31. Con estas mejoras al instrumento se obtuvo un cuestionario de 33 afirmaciones. Mejora que se observa en la figura 3.3.

En seguida, se vaciarán los datos (tanto de tutores como de tutorados) en una hoja de cálculo para ser procesados en Weka.

FOLIO: _____

CUESTIONARIO PARA IDENTIFICAR AFINIDADES Y HABILIDADES

Introducción: La Universidad Autónoma de Chiapas y la Universidad Politécnica de Puebla trabajan en un estudio que servirá para identificar el perfil de los estudiantes con el objetivo de asignarles un tutor acorde a las características del grupo. Solicitamos de su ayuda para contestar este cuestionario.

Indicaciones: Le pedimos que conteste el cuestionario con la mayor sinceridad posible. No hay respuestas correctas ni incorrectas. Todas las preguntas tienen cinco opciones de respuesta, elije solamente la que mejor describa lo que consideres apropiado. Por último, le pedimos evitar dejar preguntas sin contestar.

Información demográfica

Edad:_____ **Sexo:** M F **Estado civil:** _____ **Trabaja:** SI NO

CUESTIONARIO

5: Totalmente de acuerdo, **4:** Parcialmente de acuerdo, **3:** Ni en acuerdo ni en desacuerdo, **2:** Parcialmente en desacuerdo, **1:** Totalmente en desacuerdo

1. Tengo buena construcción gramatical para redactar reportes y ensayos					
2. Tengo habilidad de presentación, discusión y argumentación					
3. Represento fácilmente mis ideas con mnemotecnia (diagramas, esquemas, presentaciones, etc.)					
4. Explico e interpreto fácilmente la realidad					
5. Comparto mis ideas con otros de manera sencilla					
6. Produzco ideas originales que permiten crear e innovar					
7. Aplico fácilmente conceptos, valores y herramientas en la realidad natural o social					
8. Propongo alternativas para solucionar problemas y selecciono fácilmente las opciones viables					
9. Enfrento problemas y los supero con facilidad					
10. Considero que tengo autonomía intelectual y moral					
11. Realizo actos con responsabilidad ética, social y ambiental					
12. Demuestro de manera oral, escrita o física las cualidades propias					
13. Defino necesidades de aprendizaje y busco satisfacerlas con el máximo provecho y mínimo esfuerzo					
14. Suelo realizar planes para alcanzar metas realistas					
15. Utilizo recursos tecnológicos e informáticos para mi aprendizaje					
16. Utilizo recursos tecnológicos e informáticos para facilitar mi aprendizaje o actividad laboral					
17. Aprovecho recursos digitales de forma responsable, pertinente y ágil en mi actividad académica					
18. Encuentro información de interés de forma rápida y sencilla					
19. Organizo mi tiempo adecuadamente para realizar actividades académicas o personales					
20. Tengo la capacidad de automotivarme a pesar de la dificultad de ciertas actividades					
21. Tengo responsabilidad y compromiso para realizar cada una de las actividades planeadas					
22. Tengo una mente abierta para las diferentes opiniones que se generan en el grupo					
23. Realizo una autoevaluación constante de mis progresos					
24. Aplico constantemente estrategias para evitar mis deficiencias					
25. Disfruto de ir al cine/teatro/museo					
26. Practico o veo algún deporte					
27. Asisto frecuentemente a espectáculos					
28. Me distraigo realizando alguna actividad dentro de casa					
29. Disfruto de participar frecuentemente en redes sociales					
30. Frecuentemente, opto por cenar fuera de casa					
31. Disfruto de los paseos al aire libre					
32. Me gusta ir al spa/gimnasio					
33. Visito frecuentemente a un familiar o amigo					

Gracias por su tiempo y participación.

Figura 3.3. Versión final del instrumento H-A

3.4. Experimentación en Weka

Un experimento se lleva a cabo para analizar si una o más variables independientes afectan a una o más variables dependientes y por qué lo hacen. Se entiende como experimento a la "*situación de control en la cual se manipulan, de manera intencional, una o más variables independientes (causas) para analizar las consecuencias de tal manipulación sobre una o más variables dependientes (efectos)*" (Hernández Sampieri, Fernández Collado, & Baptista Lucio, Metodología de la investigación, 2010).

La figura 3.4, muestra los pasos para realizar un experimento de acuerdo a Hernández Sampieri, Fernández Collado, & Baptista Lucio (2010).

Paso 1	
Decidir cuantas variables independientes y dependientes deberán incluirse en el experimento.	• 33 variables para identificar habilidades • 4 variables tipo demográficas
Paso 2	
Elegir los niveles o modalidades de manipulación de las variables y traducirlos en tratamientos experimentales	La variables fueron clasificadas en: Demográficas, competencias, autonomía en el proceso de aprendizaje y afinidades e intereses
Paso 3	
Desarrollar el instrumento o instrumentos para medir las variables	Se desarrolló un instrumento de 33 reactivos con 5 opciones de respuesta
Paso 4	
Seleccionar una muestra de personas para el experimento	Se seleccionaron a 300 tutorados y 35 tutores de acuerdo al tema 5.2
Paso 5	
Reclutar a los sujetos participantes del experimento	El instrumento se aplicará in situ para ambos sujetos de estudio
Paso 6	
Seleccionar el diseño experimental o cuasi experimental apropiado para el estudio	Experimento de laboratorio
Paso 7	
Planear como vamos a manejar a los sujetos que participen en el experimento.	La participación de los sujetos de estudio solamente consistió en responder el cuestionario

Figura 3.4. Diseño del experimento

Hernández Sampieri, Fernández Collado, & Baptista Lucio (2010) proponen los dos últimos pasos (8 y 9) para completar el experimento que consiste en llevar el control de los grupos y en la aplicación de pre-pruebas o pos-pruebas, respectivamente. No obstante, para el caso del experimento que se muestra en la figura 3.4, y de acuerdo a las características de este estudio; dichos pasos no se requieren ejecutar.

Por otro lado, es importante mencionar que dadas las características del estudio en cuestión, se realizarán experimentos de laboratorio, pues de acuerdo a (Hernández Sampieri, Fernández Collado, & Baptista Lucio) generalmente logran un control más riguroso que los experimentos de campo y que a pesar de las múltiples críticas sobre este tipo de experimentos, en este caso, se le dará validez externa[5] al obtener una muestra significativa de la población de estudio; por lo que los datos si se pueden generalizar a toda la población de la UPPuebla y a poblaciones con características similares.

Como se mencionó en el capítulo III, se empleará la herramienta Weka para realizar los experimentos, el proceso para cada experimento se generaliza en la figura 3.5.

[5]**Validez externa** Posibilidad de generalizar los resultados de un experimento a situaciones no experimentales, así como a otras personas y poblaciones. (Hernández Sampieri, Fernández Collado, & Baptista Lucio, Metodología de la investigación, 2010)

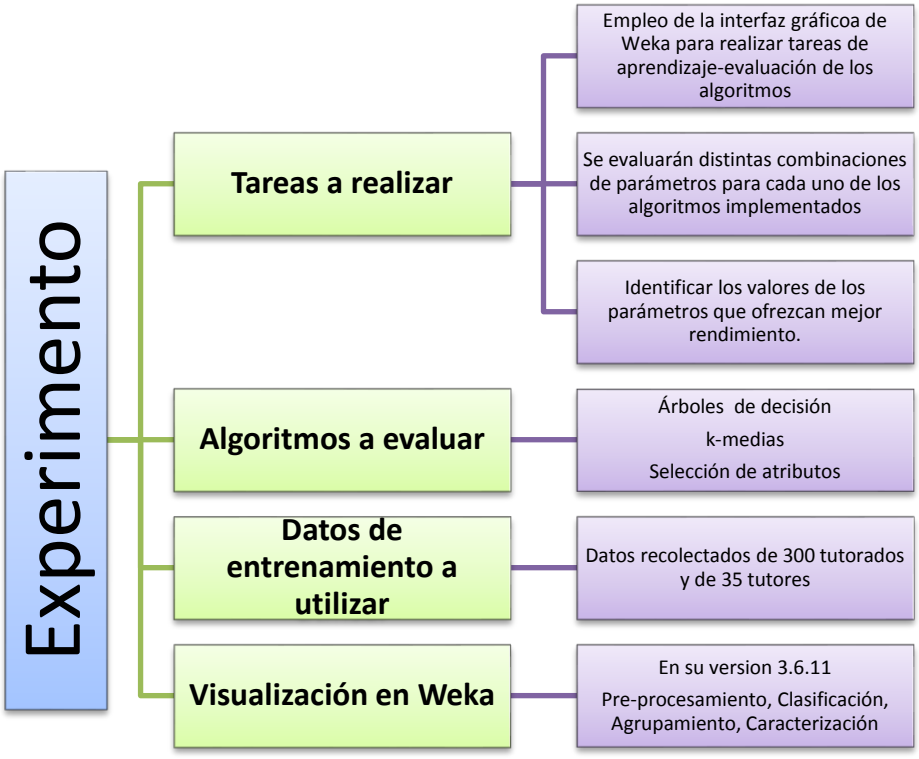


Figura 3.5. Diseño de un experimento para Weka

3.4.1. Aplicación de algoritmos y análisis de datos

Los algoritmos que se aplicarán serán: árboles de decisión, k-medias y selección de atributos.

El algoritmo árboles de decisión, permitirá generar un árbol con las características principales del capital humano, a fin de caracterizarlos como líderes o colaboradores.

- Como se mencionó en el capítulo II se empleará la matriz de confusión para obtener información sobre las clasificaciones reales y predicciones realizadas por un sistema de clasificación; puesto que en esta matriz se definen diferentes métricas que proporcionan información sobre el desempeño de los algoritmos

de aprendizaje automático. Al mismo tiempo, se describió la validación cruzada que se empleará para evaluar modelos obtenidos.

- En general, para estimar el desempeño del algoritmo árboles de decisión se considerarán las métricas de exactitud, precisión, medida F y *recall*.

El algoritmo k-medias, permitirá formar grupos de individuos basados en sus competencias y afinidades personales.

- Para crear grupos se realizará desde dos aristas: 1) enfoque didáctico-pedagógico y 2) empresarial.
    - Para desarrollar el enfoque didáctico-pedagógico se realizará una búsqueda de la literatura que contribuya a determinar el número de individuos que debe tener un grupo de estudiantes.
    - Para cumplir con el enfoque empresarial, se realizará una búsqueda sobre las diez empresas más importantes del País, a fin de determinar el número promedio de líderes que hay en esas diez empresas.
- Para realizar la experimentación con este algoritmo, se emplearán dos distancias Manhattan y Euclídea; dado que Weka a pesar de tener más distancias, solamente implementa estas dos.

La selección de atributos, permitirá identificar aquellas competencias principales (comunicación, interpersonal, intrapersonal, autodirección, autonomía, digital) que se deben reconocer en un líder o un colaborador.

- En la experimentación con este algoritmo se emplearán diversos evaluadores que permitan obtener aquellos atributos más importantes, a saber: CfsSubsetEval, GainRatioAttributeEval, InfoGainAttributeEval, PrincipalComponents, SymericalUncerAttributeEval; con sus respectivos métodos de búsqueda: RandomSearch y ranker.

Referencias del capítulo

Díaz Narváez, P. (2009). *Medotología de la investigación y bioestadística.* Santiago de Chile: RIL Editores.

Hernández Sampieri, Fernández Collado & Baptista Lucio (2010). Metodología de la investigación. McGraw Hill.

Prieto Herrera, J. E. (2013). *Investigación de mercados.* Bogotá-Colombia: Ecoe Editores.

Urbina Nájera, A. B., de la Calleja, J., Vega Lebrún, C. A., López Maldonado, N., & Pico González, B. (2014). Desarrollo y validación de un instrumento para identificar perfiles de tutorados y tutores de la modalidad virtual. *CAFVIR-2014* (págs. 227-234). Antigua Guatemala: CAFVIR.

Capítulo IV

MÉTODOS PARA MEJORAR LA TOMA DE DECISIONES

INTRODUCCIÓN

En este capítulo se detallan las propuestas que dan origen a este trabajo de investigación, considerando dos aristas, por un lado un enfoque didáctico-pedagógico y por el otro, un enfoque empresarial. Los datos para la experimentación fueron obtenidos de la UPPuebla y de la lista de las diez empresas más importantes del País. Se aplicó el algoritmo *k-medias* para formar grupos de individuos basados en sus competencias y afinidades personales; se aplicó el algoritmo selección de atributos para identificar aquellas competencias sobresalientes que definen a un líder y a un colaborador; así como, determinar si las afinidades personales son importantes en esta caracterización. Por otro lado, se aplicó el algoritmo árboles de decisión para conocer la característica más importante entre las competencias y afinidades personales del capital humano y con ello determinar todas aquellas competencias que caracterizan a un líder y a un colaborador. Finalmente, se particularizan las propuestas para crear procesos que permitan tomar decisiones al caracterizar al capital humano en líder o colaborador.

4.1. Procesamiento de datos

En esta sección se describe el procedimiento empleado para extraer y procesar el conjunto de datos

4.1.1 Extracción del conjunto de datos

Debido a que resulta incierto conocer el número de individuos que debe tener un grupo de trabajo y que no hay forma de determinarlo; la experimentación se realizará desde dos aristas: 1) Creación de grupos en términos didácticos-pedagógicos, 2) Creación de grupos en función del aprendizaje automático orientado a la empresa. A continuación se describe la extracción de los datos para ambos casos.

Conjunto de datos para la experimentación con un enfoque didáctico-pedagógico. Los experimentos se realizarán con la información de 19 PTC (Profesores de Tiempo Completo) a quienes a partir de este momento se les llamará Líderes, que equivalen al 59.4% del total y con la información de 277 estudiantes, a quienes a partir de ahora se les llamará colaboradores, que equivalen al 27.7% de la población total. Este conjunto de datos fue extraído a través de la aplicación de un muestreo probabilístico aleatorio simple y una muestra representativa de la población basada en (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010), descrita en el capítulo anterior.

Conjunto de datos para la experimentación desde un enfoque empresarial. Para realizar esta experimentación se considerarán a las 10 empresas más importantes del país, deacuerdo a lo publicado en CNNexpansion, (2014); esto con el fin de conocer el número promedio de líderes que tiene una empresa.

4.1.2 Pre-procesamiento de datos

En este apartado se mencionan los pasos requeridos para crear los archivos de los conjuntos de datos desde un archivo *.xls (Excel) a un archivo *.arff que utiliza la herramienta Weka, para ser procesados y posteriormente, analizados.

Paso 1

Obtener los datos origen en archivo *.xls (véase figura 4.1)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL
1	E	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	R23	R24	R25	R26	R27	R28	R29	R30	R31	R32	R33	edac	sexo	edo civil	Trabaja
2	1	2	3	2	2	4	2	2	2	3	3	4	3	4	4	4	4	4	4	3	3	4	4	1	2	5	5	3	5	2	4	5	1	4	22	M	Soltero	S
3	2	5	5	4	5	5	5	5	4	5	5	5	4	4	4	5	5	5	5	5	5	5	5	4	4	5	4	4	5	4	2	5	2	5	25	M	Soltero	N
4	3	3	4	3	4	4	5	3	5	4	3	3	4	3	2	4	4	4	4	2	5	5	5	3	4	5	2	2	2	2	1	5	2	5	22	F	Soltero	N
5	4	3	3	2	4	4	3	4	5	4	4	3	2	4	4	5	5	5	5	5	5	4	5	1	5	5	5	3	2	4	4	5	5	3	27	M	Soltero	N
6	5	4	3	3	4	4	4	3	3	4	4	5	4	3	4	5	5	4	4	3	4	4	5	2	3	5	5	3	4	5	5	5	5	4	19	M	Soltero	N
7	6	4	4	4	5	3	3	4	4	4	4	5	3	3	5	4	4	3	3	4	5	5	5	3	5	5	5	2	4	2	4	5	5	3	18	M	Soltero	N
8	7	4	3	4	4	4	4	4	4	3	3	4	3	2	4	3	2	5	4	3	3	3	3	4	4	4	4	3	4	4	3	5	3	2	22	M	Soltero	S
9	8	5	5	5	5	4	4	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	3	1	2	1	5	5	5	21	F	Soltero	N
10	9	3	4	4	3	4	3	5	5	4	3	4	5	3	4	5	5	4	4	5	4	4	5	2	4	4	1	3	2	4	1	5	3	4	19	M	Soltero	S
11	10	4	4	4	5	5	4	5	5	4	5	5	4	4	4	4	4	4	3	3	5	4	4	3	3	5	3	4	4	3	1	5	2	2	19	F	Soltero	N
12	11	4	3	4	3	4	5	4	3	4	5	5	5	3	3	5	1	4	4	3	4	4	5	3	3	5	5	2	4	3	1	5	5	4	19	F	Soltero	N
13	12	4	4	5	4	4	4	4	4	4	4	3	3	3	4	4	3	3	3	4	4	4	4	3	3	3	3	4	4	3	3	3	3	3	21	M	Soltero	S
14	13	3	3	4	5	4	3	5	4	5	5	5	4	4	4	5	5	5	4	5	5	5	3	2	3	5	3	2	5	1	1	5	4	3	19	F	Soltero	N
15	14	4	4	4	4	3	4	4	4	4	4	3	4	4	4	5	5	4	4	4	4	4	4	3	3	5	5	4	4	5	2	4	3	4	20	M	Soltero	N
16	15	4	3	5	4	3	4	5	5	3	5	5	4	5	5	4	4	5	5	4	5	4	4	4	4	5	5	4	5	5	4	5	3	5	20	F	Soltero	S
17	16	5	4	5	3	3	4	3	2	3	4	5	5	3	5	5	5	5	3	4	4	5	5	3	3	5	1	3	1	2	1	5	1	3	19	F	Soltero	N
18	17	3	4	4	5	5	4	3	4	4	4	3	4	4	3	4	4	4	5	3	5	4	5	4	3	4	4	2	3	2	1	5	3	5	23	M	Soltero	N
19	18	2	4	5	4	3	4	4	4	3	4	5	5	2	2	3	3	4	5	5	5	4	5	3	4	5	5	5	3	4	1	5	4	5	23	F	Soltero	S
20	19	2	3	2	2	1	2	2	3	2	2	1	2	2	1	1	2	1	3	3	2	3	1	1	2	3	2	5	1	1	3	1	5	4	20	M	Soltero	N
21	20	4	4	3	4	3	3	4	4	4	4	5	3	3	4	5	5	5	5	2	5	3	5	1	3	5	5	3	4	3	1	5	3	4	20	M	Soltero	N
22	21	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4	3	5	4	5	4	5	4	3	4	4	22	F	Soltero	N
23	22	5	4	4	3	3	4	4	4	4	5	5	3	4	4	4	4	5	5	3	5	3	4	3	4	5	5	3	5	2	3	5	1	5	20	F	Soltero	S

Figura 4.1 Datos origen desde Excel.

Paso 2

Seleccionar la ruta: Menú archivo→Guardarcomo→tipo CVS (MS-DOS)

- Escribir nombre del archivo→ Guardar
- Aparecerán dos ventanas emergentes, presionar Aceptar→ sí.

Paso 3

Abrir el archivo en la aplicación“Bloc de Notas”, mismo que aparecerá como se observa en la figura 4.2

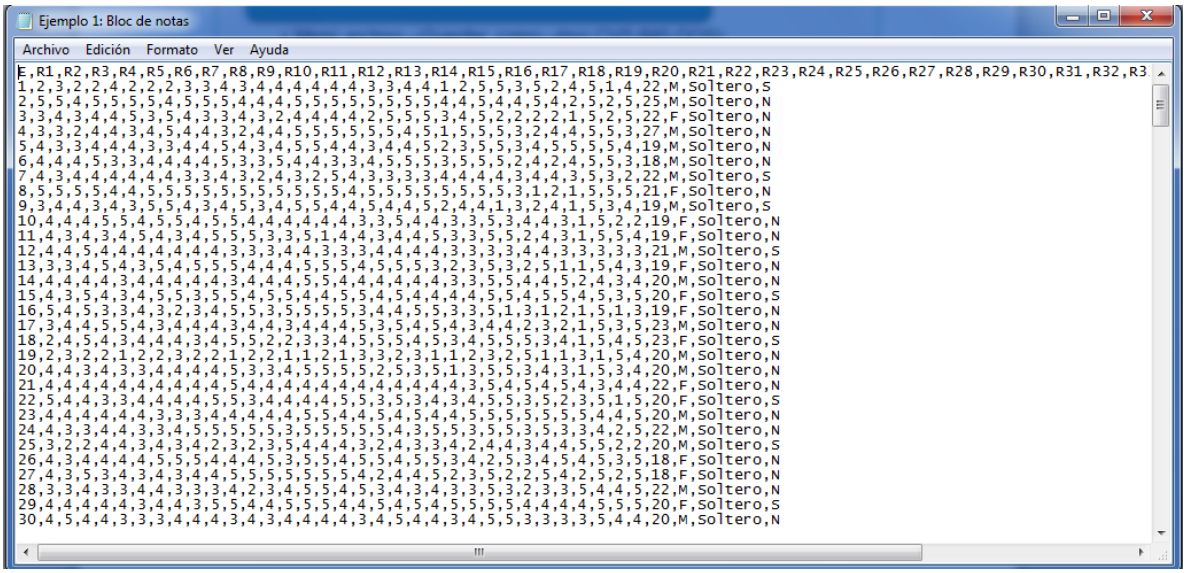


Figura 4.2 Datos originales en formato CVS (MS-DOS) abiertos en el Bloc de notas

Paso 4

Escribir el encabezado que se encuentra en la figura 4.3., en el cual se define el nombre del archivo, sus atributos y el conjunto de datos. Los atributos comunicación, interpersonal, intrapersonal, digital, autonomía y afinidades tienen valores predeterminados de 1 a 5, donde 1 equivale a Totalmente en desacuerdo y 5 Totalmente de acuerdo. El atributo edad está definido para cualquier valor numérico, el atributo sexo solamente acepta una F equivalente a femenino y una M equivalente a masculino, para el atributo edo_civil y de acuerdo a los resultados obtenidos, solamente se definen dos valores: casado y soltero. Finalmente, para el atributo trabaja tiene dos valores S/N, que significa sí y no.

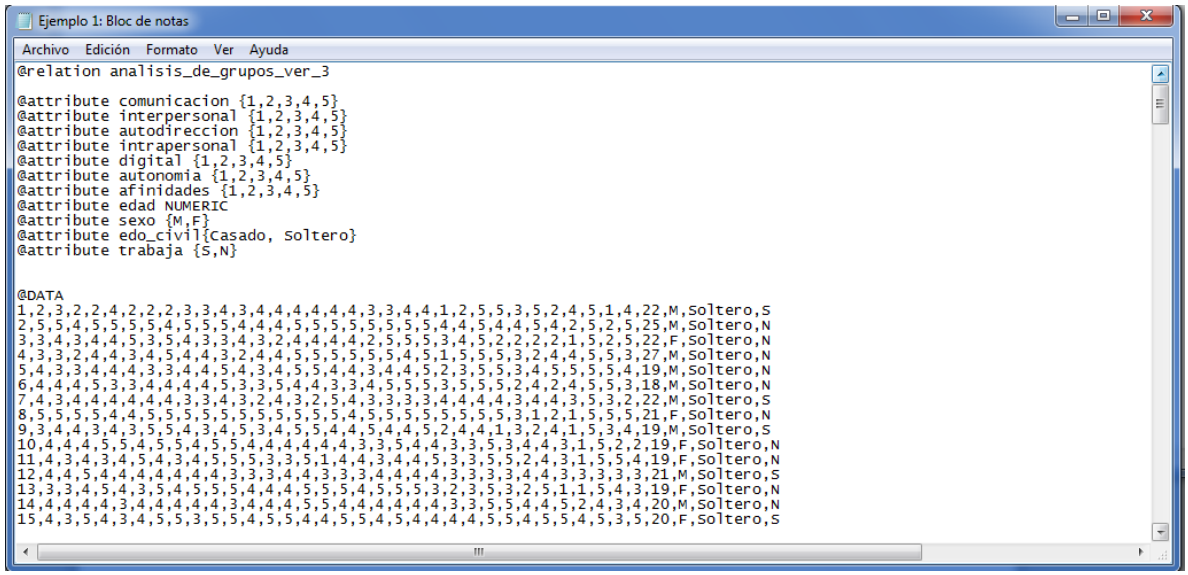


Figura 4.3 Encabezado para archivos *.arff

Paso 5

Posterior a la captura del encabezado correspondiente y de escribir los comentarios necesarios (líneas que empiezan con %), se guarda el archivo bajo la siguiente ruta (véase figura 4.4):

- Archivo→Guardar como
- Escribir en el campo nombre el deseado y agregar la extensión *.arff
- Seleccionar en Tipo→Todos los archivos
- Seleccionar en codificación→ANSI

Paso 6

El último paso consiste en abrir el archivo en Weka (en este trabajo se empleó la versión 3.6.11), tal y como se describió en el capítulo anterior.

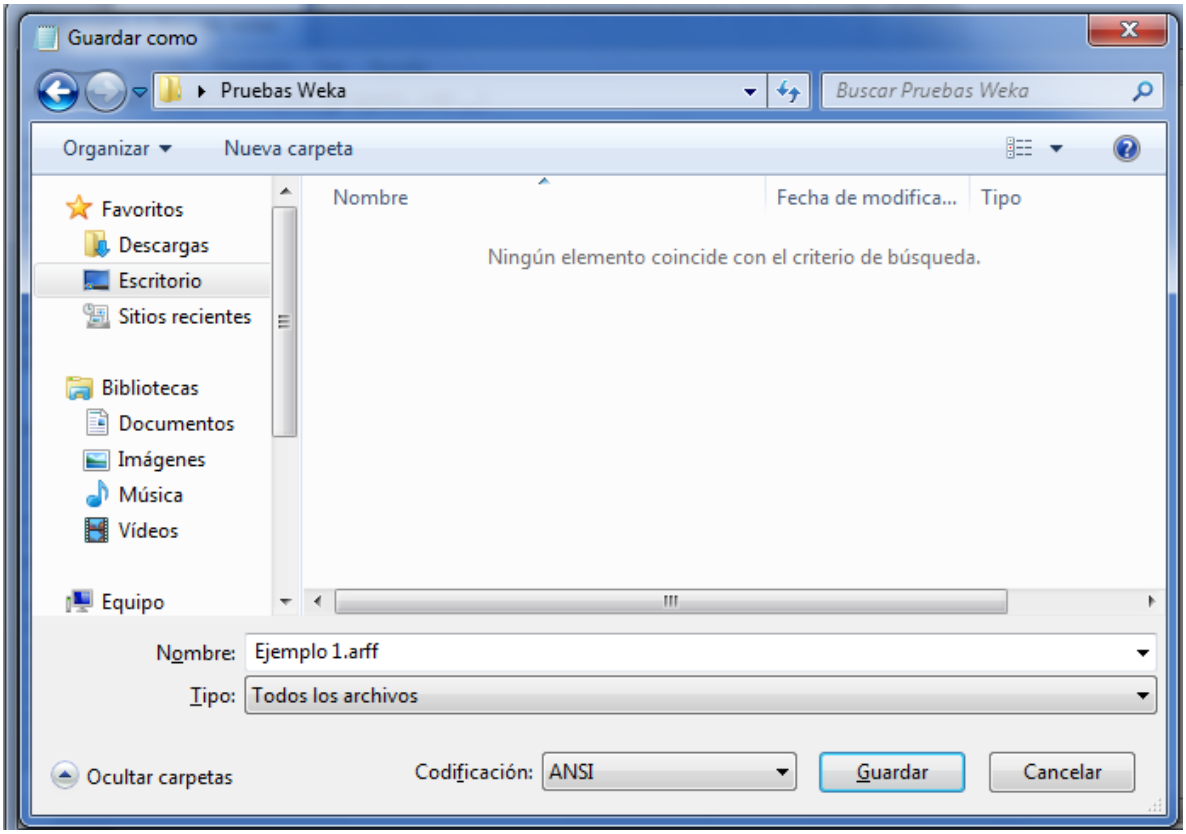


Figura 4.4 Archivo con extensión *.arff

Como se ha mencionado en el marco teórico, el objetivo del aprendizaje automático es construir un clasificador que permita agrupar, identificar, clasificar o reconocer objetos de forma automática; en particular para este trabajo de investigación se desea la agrupación de personas en términos de sus competencias de comunicación, interpersonal, intrapersonal, autodirección, digital, autonomía en el proceso de aprendizaje y sus afinidades (respecto a pasatiempos y formas de distracción).

Con los datos mostrados en la figura 4.5 se realizarán experimentos desde dos aristas: 1) Creación de grupos en términos didácticos-pedagógicos, 2) Creación de grupos en función del aprendizaje automático orientado a la empresa.

La aplicación del algoritmo *k-medias* fue realizada con dos archivos, por un lado el archivo que contiene el conjunto de datos de 277 colaboradores y por el otro, un archivo con el conjunto de datos de 19 líderes.

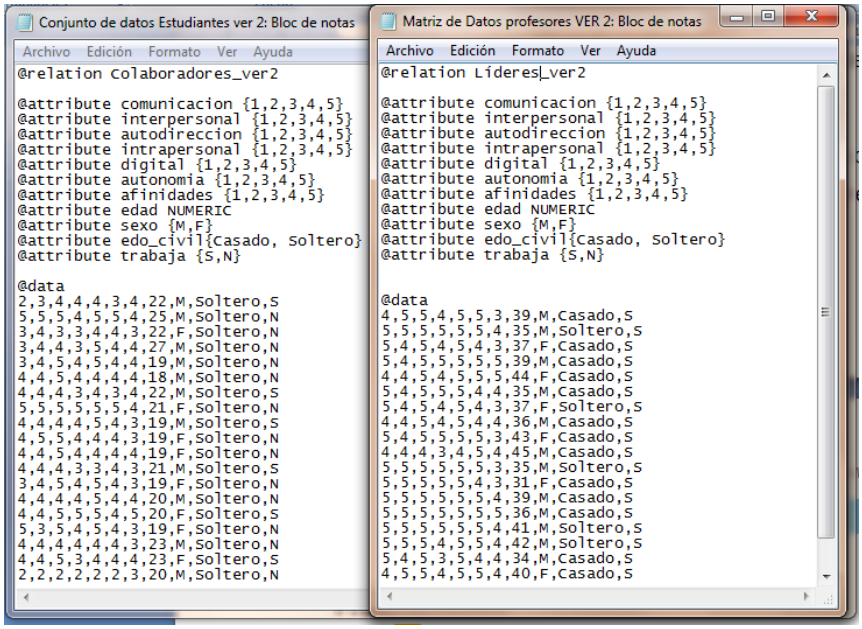


Figura 4.5 Conjunto de datos para aplicar *k-medias*

Procedimiento

En este apartado se describe el procedimiento realizado para explicar y sustentar cada una de las aristas: Enfoque didáctico-pedagógico y enfoque empresarial.

La figura 4.6 muestra la información que contiene archivo “análisis_” al ser abierto desde Weka. Para ello se ejecuta; y enseguida en la pestaña *Preprocess*, se selecciona la ruta: open file→ colaboradores.

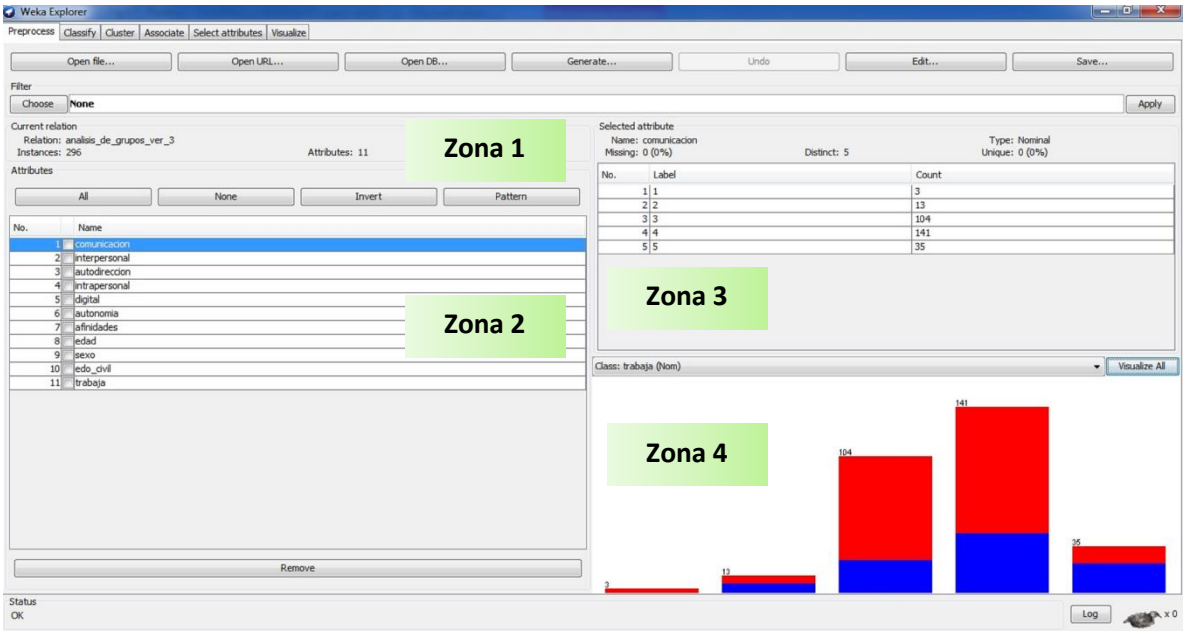
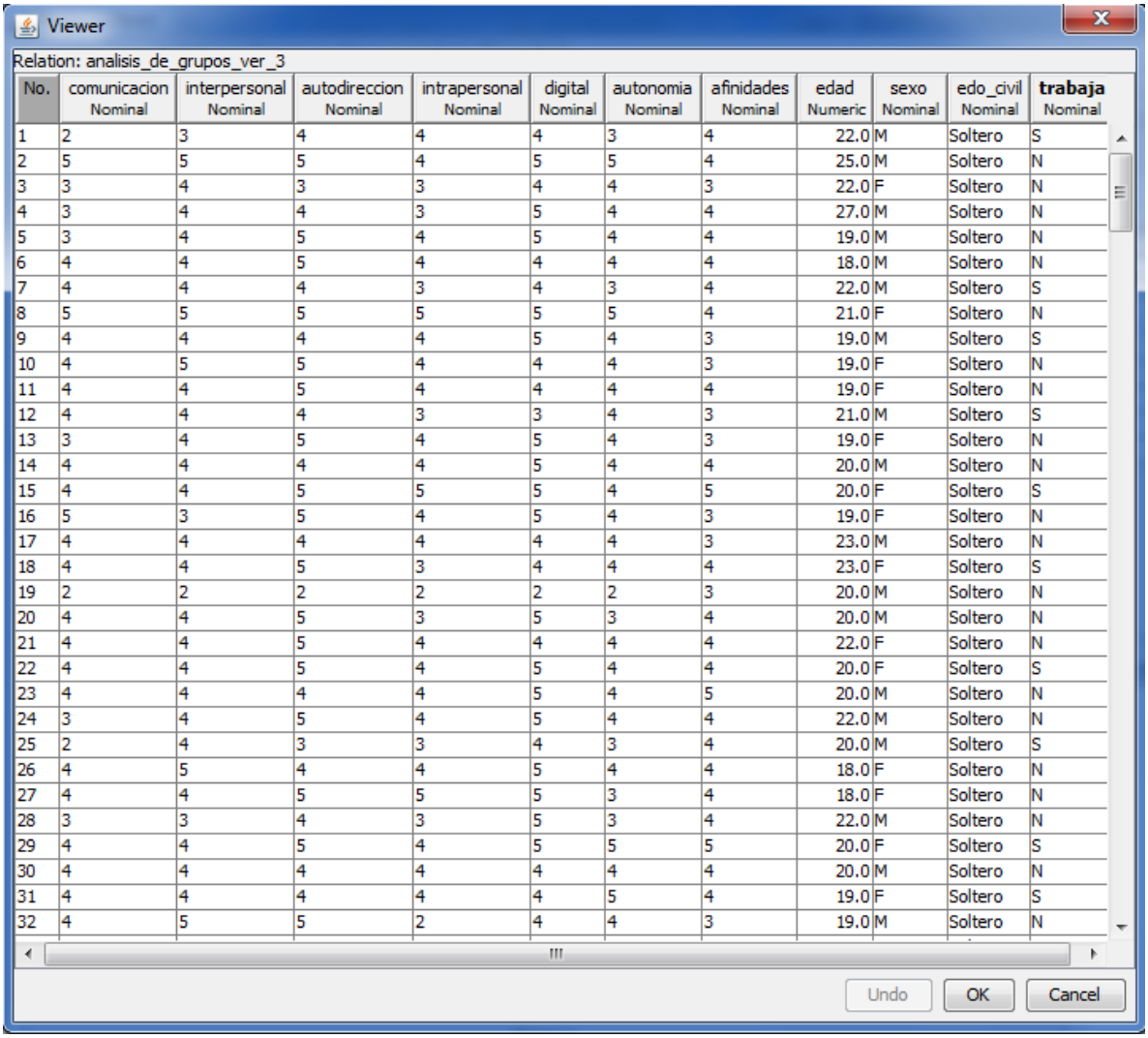


Figura 4.6 Información que muestra Weka al abrir un archivo para su procesamiento

Para una mejor descripción de cada interfaz presentada en Weka, se ha divido en zonas. En la figura 4.6, en la zona 1 se muestran las características actuales del archivo, como: nombre del archivo: análisis_de_grupos_ver_3, Instancias: 296 y Atributos: 11.

En la zona 2 se muestra el listado de los atributos (las siete competencias que se evaluaron, más cuatro criterios demográficos (edad, sexo, estado civil y si trabaja). En la zona 3 se listan de forma individual las características de cada atributo conforme se seleccionan en la zona 2; estas características son: nombre, tipo, valores faltantes, total de valores y porcentajes. Al mismo tiempo muestra una tabla con el número de valores, la etiqueta y el conteo total de ocurrencias por cada valor. Finalmente en la zona 4, se muestra al presionar *Visualice All*, el histograma (véase figura 4.8) de los valores presentados en la zona 3.

Si se desea una comprobación de los datos leídos desde el archivo seleccionad, se presiona *edit* en la ventana inicial (véase figura 4.7).



Viewer

Relation: analisis_de_grupos_ver_3

No.	comunicacion Nominal	interpersonal Nominal	autodireccion Nominal	intrapersonal Nominal	digital Nominal	autonomia Nominal	afinidades Nominal	edad Numeric	sexo Nominal	edo_civil Nominal	**trabaja** Nominal
1	2	3	4	4	4	3	4	22.0	M	Soltero	S
2	5	5	5	4	5	5	4	25.0	M	Soltero	N
3	3	4	3	3	4	4	3	22.0	F	Soltero	N
4	3	4	4	3	5	4	4	27.0	M	Soltero	N
5	3	4	5	4	5	4	4	19.0	M	Soltero	N
6	4	4	5	4	4	4	4	18.0	M	Soltero	N
7	4	4	4	3	4	3	4	22.0	M	Soltero	S
8	5	5	5	5	5	5	4	21.0	F	Soltero	N
9	4	4	4	4	5	4	3	19.0	M	Soltero	S
10	4	5	5	4	4	4	3	19.0	F	Soltero	N
11	4	4	5	4	4	4	4	19.0	F	Soltero	N
12	4	4	4	3	3	4	3	21.0	M	Soltero	S
13	3	4	5	4	5	4	3	19.0	F	Soltero	N
14	4	4	4	4	5	4	4	20.0	M	Soltero	N
15	4	4	5	5	5	4	5	20.0	F	Soltero	S
16	5	3	5	4	5	4	3	19.0	F	Soltero	N
17	4	4	4	4	4	4	3	23.0	M	Soltero	N
18	4	4	5	3	4	4	4	23.0	F	Soltero	S
19	2	2	2	2	2	2	3	20.0	M	Soltero	N
20	4	4	5	3	5	3	4	20.0	M	Soltero	N
21	4	4	5	4	4	4	4	22.0	F	Soltero	N
22	4	4	5	4	5	4	4	20.0	F	Soltero	S
23	4	4	4	4	5	4	5	20.0	M	Soltero	N
24	3	4	5	4	5	4	4	22.0	M	Soltero	N
25	2	4	3	3	4	3	4	20.0	M	Soltero	S
26	4	5	4	4	5	4	4	18.0	F	Soltero	N
27	4	4	5	5	5	3	4	18.0	F	Soltero	N
28	3	3	4	3	5	3	4	22.0	M	Soltero	N
29	4	4	5	4	5	5	5	20.0	F	Soltero	S
30	4	4	4	4	4	4	4	20.0	M	Soltero	N
31	4	4	4	4	4	5	4	19.0	F	Soltero	S
32	4	5	5	2	4	4	3	19.0	M	Soltero	N

Undo OK Cancel

Figura 4.7 Visualización del conjunto de datos para comprobar que son leídos desde el archivo seleccionado

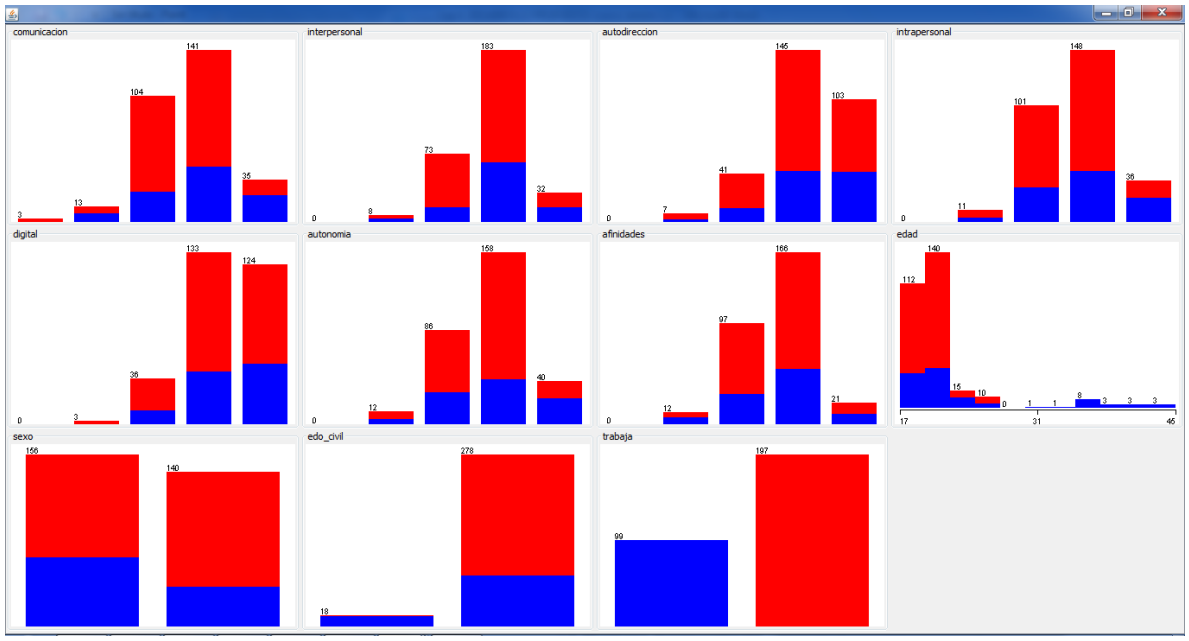


Figura 4.8 Histograma de los atributos que se evalúan dentro de Weka

La figura 4.8 muestra la visualización de todos los atributos. Esta es una manera útil para comprobar la efectividad de separación de cada uno de los atributos, considerados de forma individual.

En el color azul se grafican los valores perfectamente separados y en color rojo los que no se han separado perfectamente. En este sentido se espera que las clases (rojo y azul) estén perfectamente separados al presentar barras de un solo color como ocurre en el atributo *trabaja*. Es posible que esta separación se deba a que los valores que se consideraron para su evaluación (si y no) son dicotómicas, que a diferencia del resto de las variables fue evaluado bajo la escala Likert (5: Totalmente de acuerdo hasta 1: Totalmente en desacuerdo) y que por lo tanto, los valores son diversos.

4.2. Método para la creación de grupos de trabajo aplicandoel algoritmo *k-medias*

El primer método que se propone es crear de manera automática grupos de trabajo, basados en sus competencias interpersonales, intrapersonales, comunicación, digital, autonomía,

autodirección y afinidades personales. Para lograr esto, se aplica el algoritmo *k-medias* que permite agrupar individuos en función de un número determinado de ellos. El procedimiento de este algoritmo se describió en el capítulo anterior; por lo que en este apartado se describe su aplicación basada en los conjuntos de datos descritos previamente.

Para la ejecución del algoritmo *k-medias*es necesario seleccionar la pestaña *Cluster* de la ventana principal. En el botón *Choose* se elige el algoritmo en cuestión (véase figura 4.6). La figura 4.9 muestra los valores predeterminados que el algoritmo emplea.

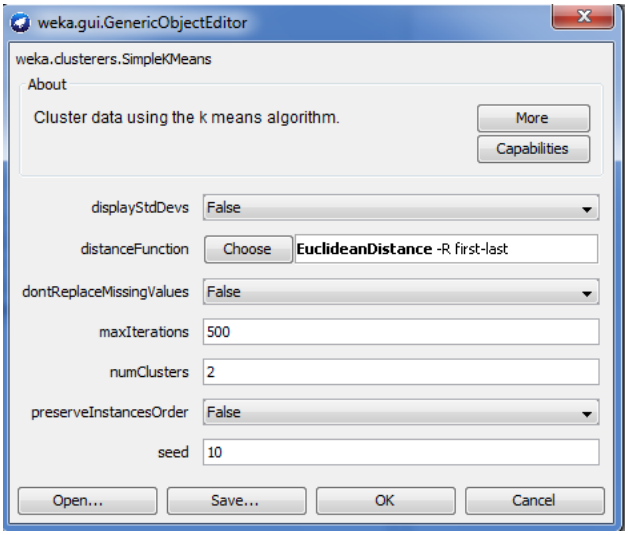


Figura 4.9 Valores predeterminados del algoritmo *k-medias*

A continuación se describe cada uno de los valores predeterminados del algoritmo *k-medias*.

- ***displayStdDevs:*** Muestra la desviación estándar de los atributos numéricos y nominales. Sus valores son: Falso/verdadero y el valor predeterminado es: Falso

- ***distanceFunction:*** Es la función de distancia a utilizar para la comparación de instancias. Sus valores son: *ChebyshevDistance* [1] , *EditDistance* [2] , *EuclideanDistance*[3], *ManhattanDistance*[4]. El valor predeterminado es: *Euclidean-Distance*
- ***dontReplaceMissingValues:*** Reemplazar los valores perdidos. Sus valores son: Falso/verdadero. El valor predeterminado es: Falso
- ***maxIterations:*** Establece el número máximo de iteraciones. Puede tomar cualquier valor mayor a cero. El valor predeterminado es 500
- ***numClusters***: Número de agrupaciones. El valor predeterminado es 2
- ***preserveInstancesOrder:*** Preserva el orden de las instancias. Sus valores son:Falso/verdadero. El valor predeterminado es Falso
- ***Seed***: El número aleatorio de semilla a ser utilizado. Puede tomar cualquier valor mayor a cero. El valor predeterminado es 10

Con el objetivo de mostrar los resultados que se obtienen al aplicar este algoritmo para algún conjunto de datos, se realizó una prueba con los valores predeterminados del algoritmo.

Los resultados de la ejecución se muestran en la figura 4.10. En color verde se resaltan los *clusters* formados, en el que se observa que para las 296 instancias el valor promedio de cada competencia es 4 (Parcialmente de acuerdo), que la edad promedio es de 21.28 años, el sexo promedio es masculino, estado civil promedio corresponde a soltero y en promedio no trabajan.

[1] También conocida como distancia de tablero de ajedrez (*chessboarddistance*): Número de movimientos que el rey ha de hacer para llegar de una casilla a otra en un tablero de ajedrez.
[2] Distancia de Edición. Implementa la distancia de Levenshtein. Es el número de operaciones necesario para transformar una cadena en otra.
[3] Es la distancia "ordinaria" (que se mediría con una regla) entre dos puntos de un espacio euclídeo, la cual se deduce a partir del teorema de Pitágoras.
[4] Llamada también métrica Taxicab. Es una forma de geometría en la cual la métrica usual de la geometría Euclídea es reemplazada por una nueva métrica en la cual la distancia entre dos puntos es la suma de las diferencias (absolutas) de sus coordenadas.

El primer *cluster* está formado por 181 personas cuyo promedio en la competencia comunicación es igual a 3 (Ni de acuerdo ni en desacuerdo) y para el resto de las competencias y afinidades es 4 (Parcialmente de acuerdo), el promedio de edad es de 20.18 años, masculinos, solteros y no trabajan. En tanto, que el *cluster* dos está formado por 115 personas, cuya media es igual en las competencias comunicación, interpersonal, intrapersonal, autonomía y afinidades al obtener un valor de 4, el resto de las competencias autodirección y digital coinciden al responder 5 (Totalmente de acuerdo), una media de 23 años, mujeres, solteras que trabajan. Se aprecia entonces, que el algoritmo hizo la separación simple en los atributos sexo y trabaja; se infiere que la razón sea debe a que son dicotómicas, es decir, sus valores son dos (para sexo: femenino/ masculino y para trabaja: si/no). En color azul se resalta el tiempo que tomó la ejecución, mismo que es muy bajo al procesar los datos en 0.01 segundos. En color rosa se destaca el modelo y evaluación de los datos de entrenamiento. El porcentaje de instancias en el *cluster*, a saber, para el *cluster* 0=61% y para el *cluster* 1=39%. Dejando entrever que el sexo masculino prevalece en los datos y que de igual forma son individuos que no trabajan.

Clusterer output

```
=== Run information ===

Scheme:weka.clusterers.SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10
Relation:     analisis_de_grupos_ver_3
Instances:    296
Attributes:   11
              comunicacion
              interpersonal
              autodireccion
              intrapersonal
              digital
              autonomia
              afinidades
              edad
              sexo
              edo_civil
              trabaja
Test mode:evaluate on training data

=== Model and evaluation on training set ===


kMeans
======
Number of iterations: 4
Within cluster sum of squared errors: 1074.0186282704299
Missing values globally replaced with mean/mode

Cluster centroids:
                            Cluster#
Attribute        Full Data          0          1
                     (296)      (181)      (115)
=================================================
comunicacion             4          3          4
interpersonal            4          4          4
autodireccion            4          4          5
intrapersonal            4          4          4
digital                  4          4          5
autonomia                4          4          4
afinidades               4          4          4
edad               21.2838    20.1878    23.0087
sexo                     M          M          F
edo_civil          Soltero    Soltero    Soltero
trabaja                  N          N          S


Time taken to build model (full training data) : 0.05 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      181 ( 61%)
1      115 ( 39%)
```

Figura 4.10 Resultados que obtiene *k-medias* formando dos *clusters* usando valores predeterminados

Recordando que el objetivo del algoritmo *k-medias* es minimizar la disimilaridad de los elementos dentro de cada *cluster* y maximizar la disimilaridad de los elementos que caen en diferentes *clusters*. Es decir, se tiene un conjunto de datos *S* y el número *k* de *cluster* a formar; para obtener una lista *L* de los *clusters* en que caen las características de *S*.

En la experimentación de este algoritmo se obtendrá el promedio de cinco ejecuciones para los valores de semilla =10, 2, 4, 7, 16; determinados aleatoriamente, tanto para la experimentación con distancia Euclídea como con distancia Manhattan. Se sabe que existen diferentes métodos para obtener la distancia entre dos puntos, sin embargo, Weka 3.6.11 (versión actual para *Windows*) solamente tiene implementadas estas dos distancias.

Antes de comenzar con la experimentación de este algoritmo, es pertinente aclarar cómo es que estas distancias actúan y cómo afectan los resultados obtenidos tras la experimentación.

La distancia Euclídea, es la comúnmente conocida y empleada para obtener la distancia entre dos puntos "la distancia más corta entre dos puntos es la línea recta". Si se tienen dos puntos en el plano cartesiano (*a*,*b*) y (*c*,*d*) respectivamente, y se desea calcular la distancia entre ellos, por el teorema de Pitágoras, se sabe que la distancia Euclídea se mide como:

$$dE = \sqrt{(c-a)^2 + (d-b)^2} \qquad (E4.1)$$

En tanto, la distancia Manhattan o distancia L_1 es la longitud del camino más corto formado por segmentos horizontales y verticales que une a los dos puntos. Comúnmente utilizada para el diseño y optimización de rutas de distribución. Entonces, la distancia Manhattan se mide con la fórmula:

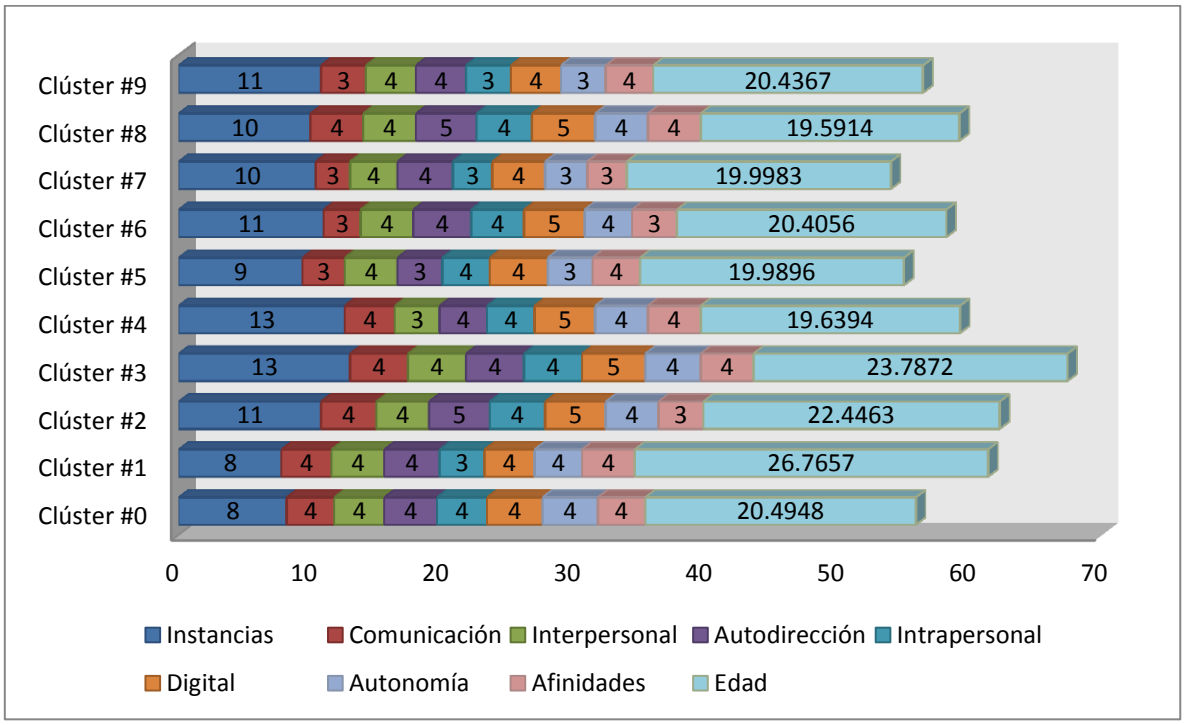
$$d = \sum_{i=1}^{n} |x_i - y_i| \qquad (E4.2)$$

En otras palabras, la distancia Manhattan entre dos elementos es la suma de las diferencias de sus correspondientes componentes (Han, Kamber, & Pei, 2011).

Resultados experimentales

En este apartado se detallan los experimentos realizados para encontrar las competencias y afinidades de cada grupo de trabajo formado por colaboradores y líderes, empleando el algoritmo *k-medias* con distancia Euclídea y distancia Manhattan. Considerando un conjunto de datos caracterizado por 6 variables que definen competencias, una variable de afinidad y 4 variables demográficas, haciendo un total de 11 variables predictoras y 296 observaciones.

Experimento No.	**1**	**Algoritmo aplicado**	***k-medias***
Objetivo	Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 28 grupos		
Conjunto de datos	277 instancias, enfoque didáctico-pedagógico		
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Ejecuciones empleando distancia Euclídea Creación de 30 grupos Empleo de todas las variables (demográficas, competencias, habilidades)			
Resultados	Los gráficos 4.1, 4.2 y 4.3 muestran los resultados obtenidos tras ejecutar el procedimiento descrito.		

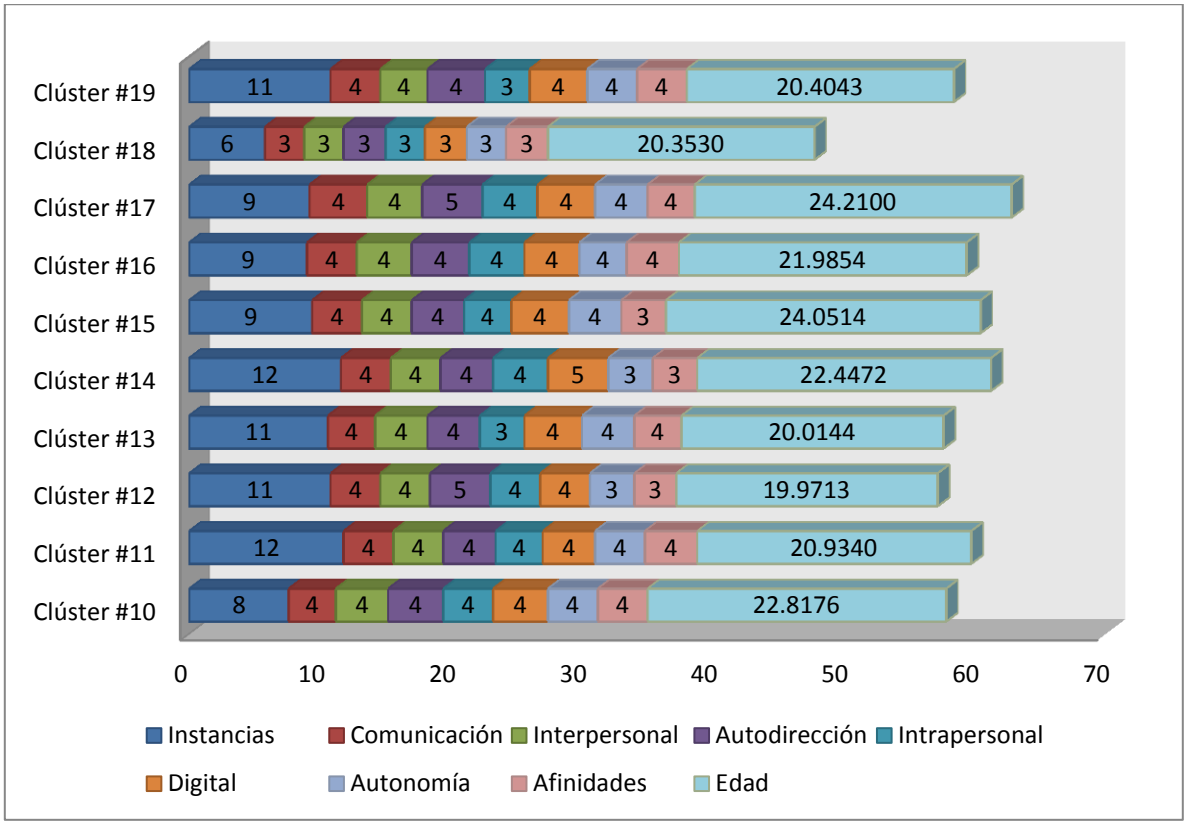


Gráfica 4.1*Cluster* del 0-9 usando distancia Euclídea.

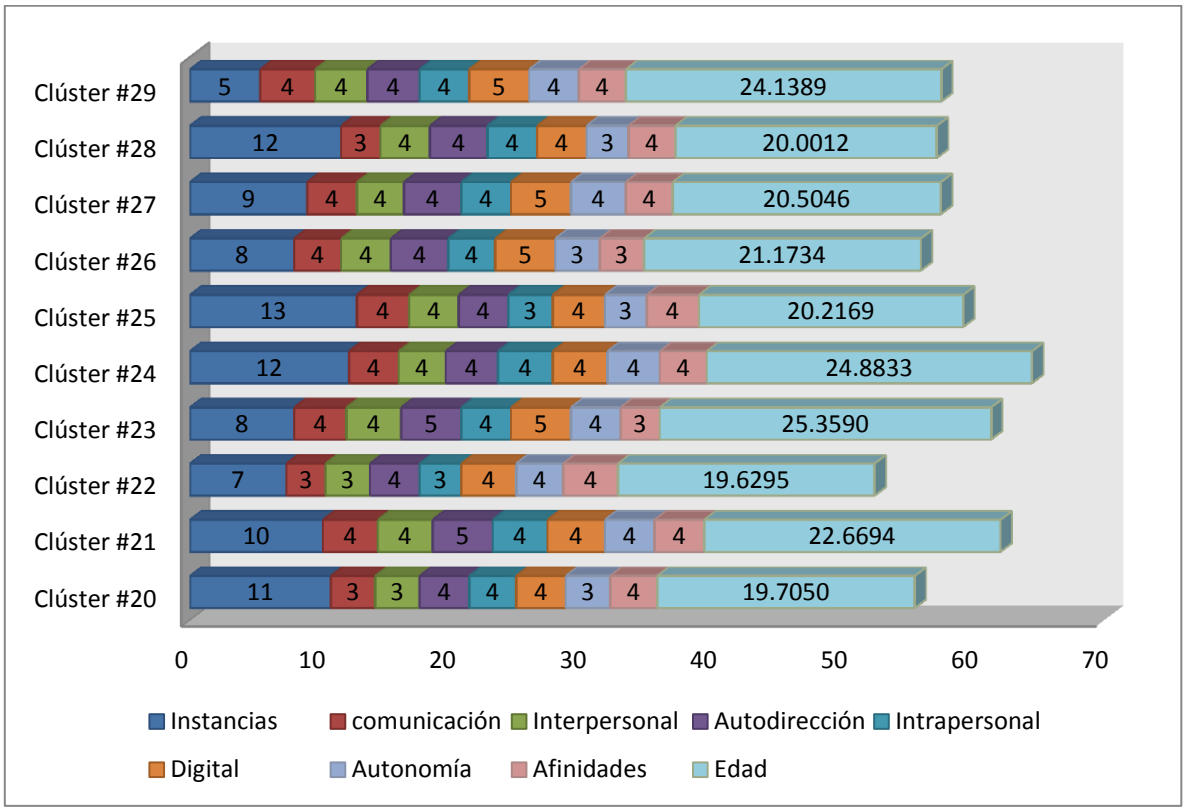
Todas las gráficas mostradas en esta sección tienen la siguiente nomenclatura:

- Los datos se leen de izquierda a derecha y de abajo hacia arriba.
- El primer color (■) se refiere al número de instancias que fueron clasificadas en ese *cluster*
- En los siguientes siete colores están representadas las habilidades (■ Comunicación, ■Interpersonal, ■ Autodirección, ■ Intrapersonal, ■ Digital y ■ Autonomía en el proceso de aprendizaje) y afinidades (■). Recordando que estás fueron valoradas de acuerdo a la escala Likert (5: Totalmente de acuerdo, 4: Parcialmente de acuerdo, 3: Ni de acuerdo ni en desacuerdo, 2: Parcialmente en desacuerdo y 1: Totalmente en desacuerdo).
- Posteriormente, se grafica edad, sexo y la variable dicotómica “trabaja”.
- Ejemplo: Considérese el *cluster* No. 0

- Se tienen 8 instancias; que en promedio tienen 4 en todas las habilidades y en promedio 4 en la afinidad, mientras que la media de edad es de 20.49 años.



Gráfica 4.2 *Cluster* 10-19 usando distancia Euclídea.



Gráfica 4.3 *Cluster* 20-29 usando distancia Euclídea.

Dado que los gráficos mostrados tras la ejecución del algoritmo no muestran resultados concluyentes acerca de las competencias y afinidades que debería tener un grupo de trabajo con igual o semejantes características que un líder, se toma la decisión de realizar la experimentación dividiendo el conjunto de datos en dos grupos, por un lado los colaboradores (277 instancias) y por el otro, los líderes (19 instancias), con el propósito de identificar con mayor certeza las características afines de cada grupo de trabajo.

4.2.1 Experimentación con un enfoque didáctico-pedagógico

Para crear grupos en el contexto educativo, fue necesario hacer una revisión de la literatura acerca del número máximo de estudiantes que pedagógicamente hablando debe tener un grupo, es decir, que deben ser atendidos por un líder.

- González Rufino (2009) menciona que el número de estudiantes por grupo no debe exceder de 10. Mientras que, Maldonado & Giandini (2010) afirman que el número máximo de estudiantes no debe ser superior a 20 estudiantes por grupo.
- El Instituto Nacional de Evaluación Educativa (2013) afirma que el número promedio de estudiantes por grupo debe ser de 15.7, es decir 16. Finalmente, la OCDE (2014) supone que en promedio, un grupo debe estar formado por máximo 23.8 estudiantes, es decir, 24 estudiantes por grupo.

En ese sentido, la figura 4.11 muestra la experimentación que se realizó para conformar grupos de colaboradores, considerando 277 instancias correspondientes a la información de estudiantes de la UPPuebla.

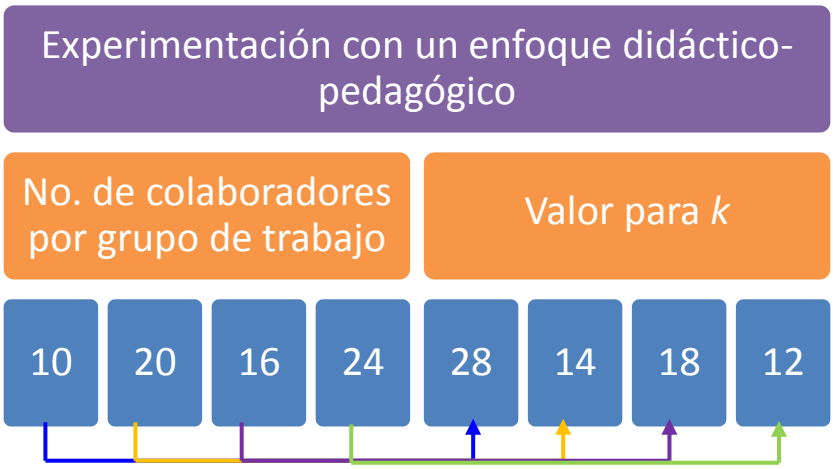
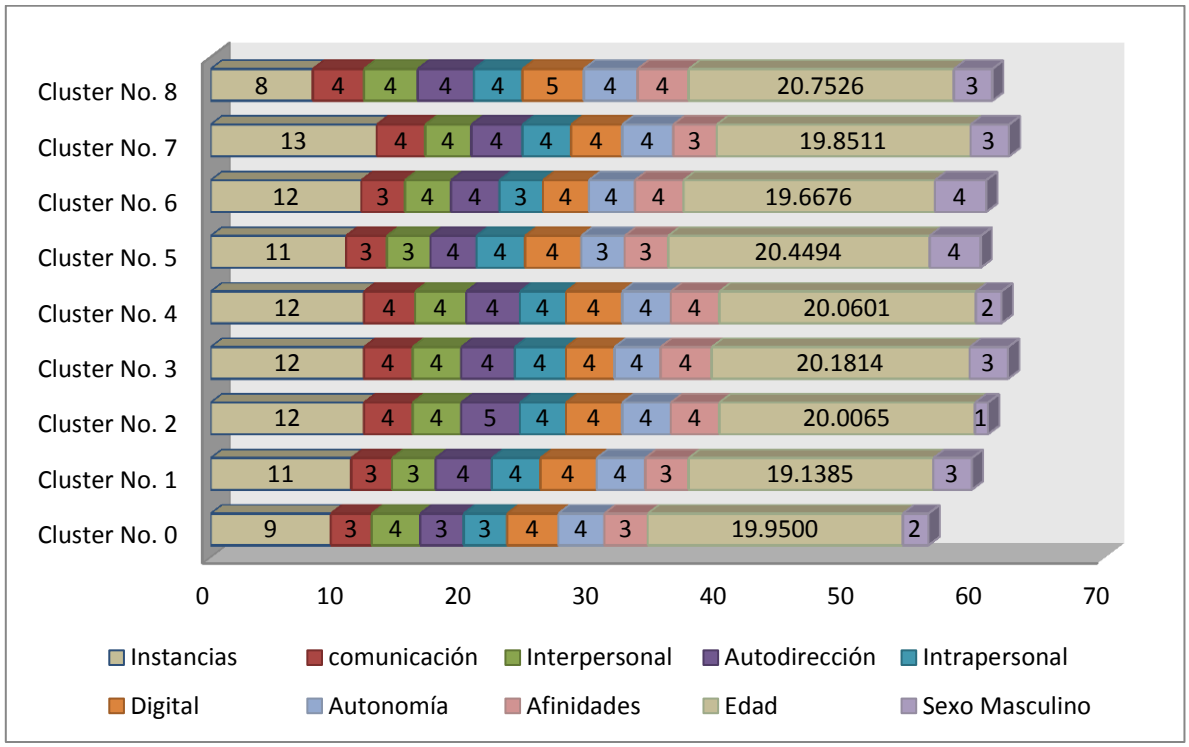


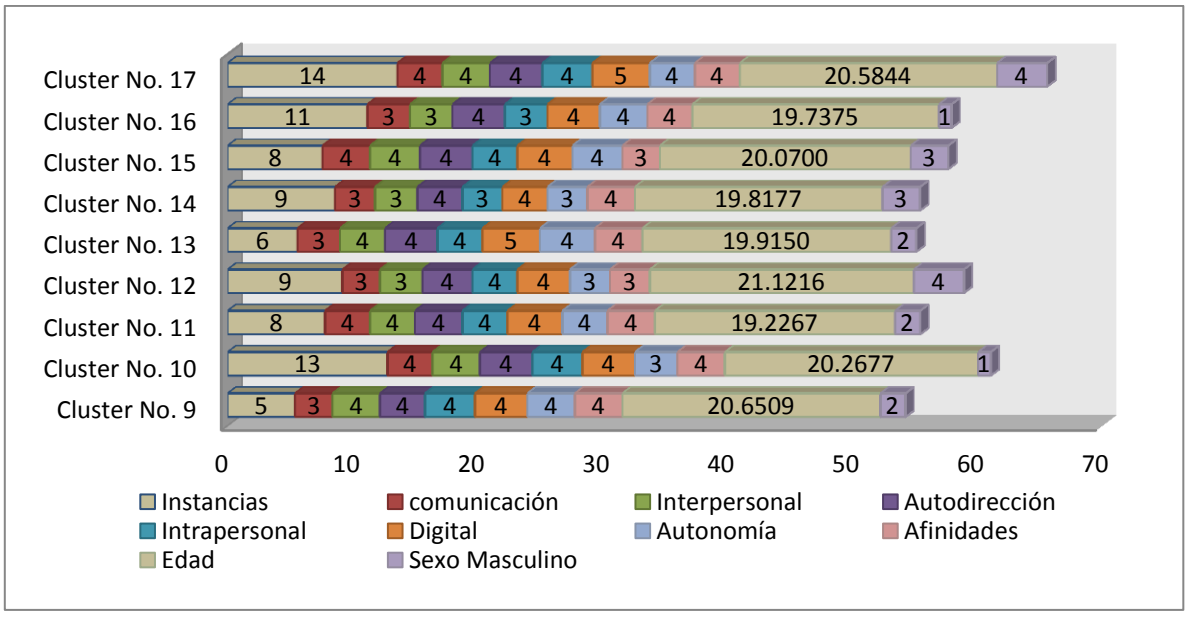
Figura 4.11 Valores que deberá tomar *k* para la aplicación del algoritmo *k-medias*

Experimentación para formar 28 grupos de trabajo

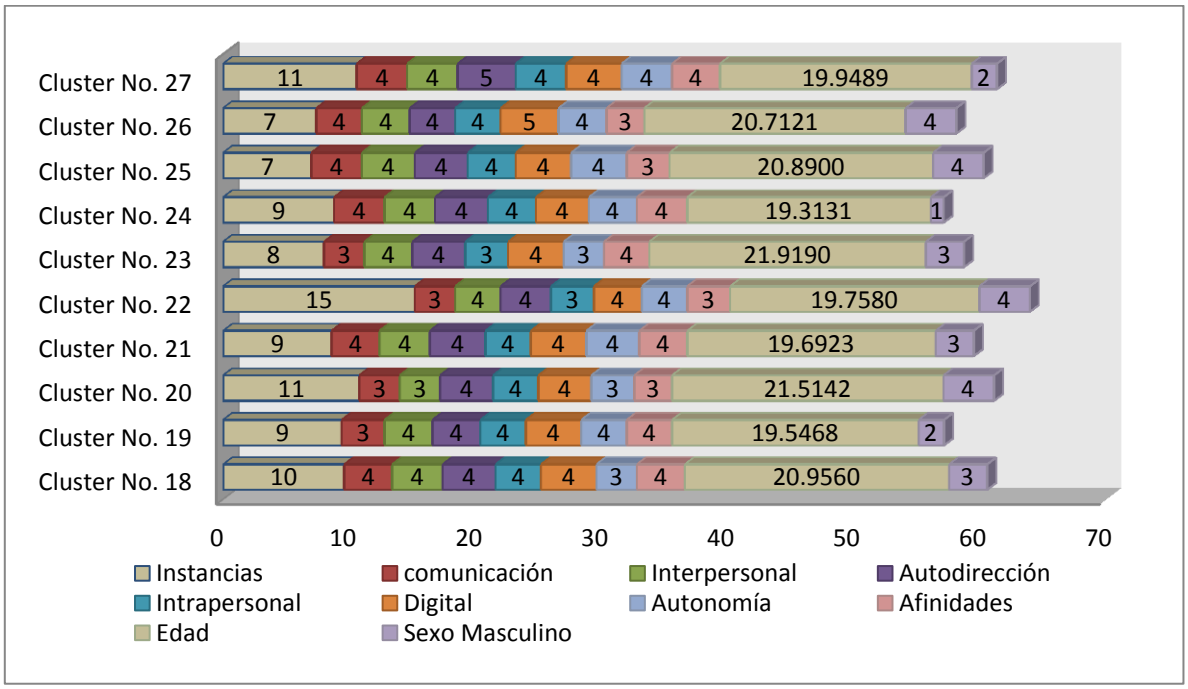
Experimento No.	**2**	**Algoritmo aplicado**	***k-medias***
Objetivo	Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 28 grupos		
Conjunto de datos	277 instancias, enfoque didáctico-pedagógico	**Distancia**	Euclídea
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Creación de 28 grupos de acuerdo a González Rufino (2009). Empleo de todas las variables (demográficas, competencias, habilidades)			
Resultados	Desde el punto de vista pedagógico, *k-medias* formó adecuadamente a 15 grupos que no exceden de 10 integrantes cada uno. Los gráficos 4.4-4.6 muestran los resultados obtenidos tras ejecutar el procedimiento descrito.		



Gráfica 4.4 *Cluster* de colaboradores del 0-8 usando distancia Euclídea.

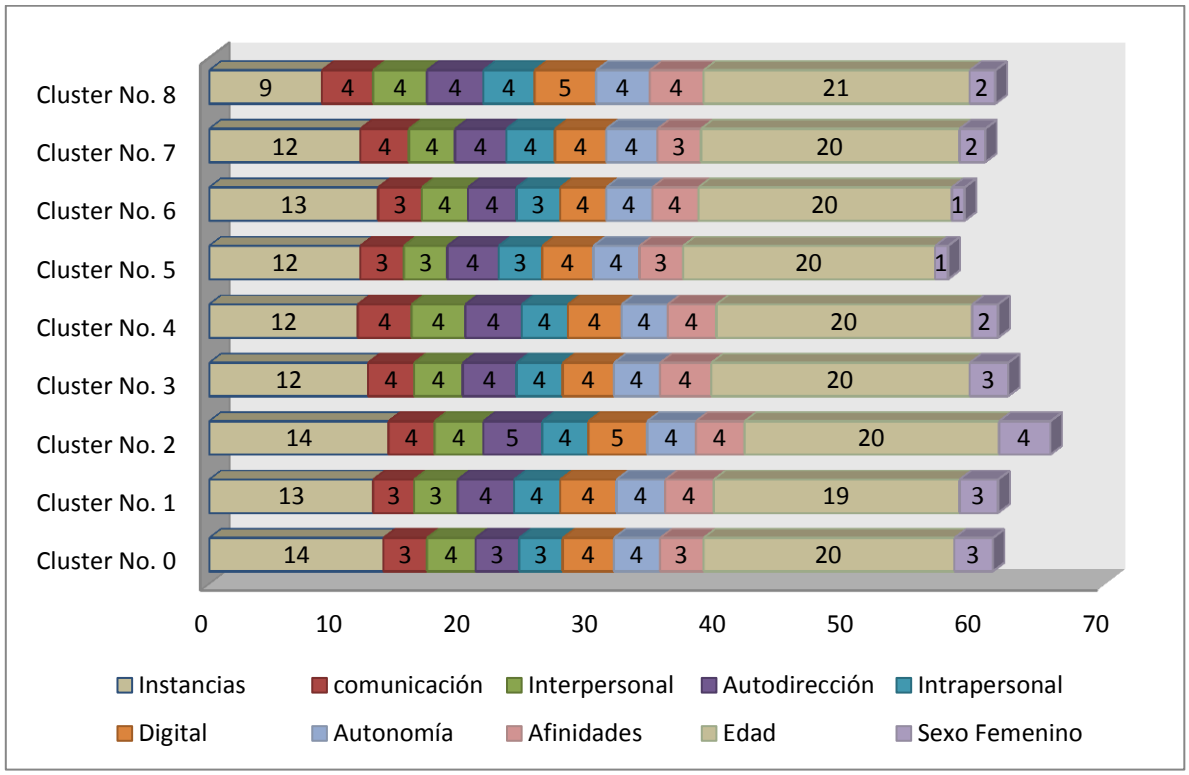


Gráfica 4.5 *Cluster* de colaboradores del 9-17 usando distancia Euclídea.

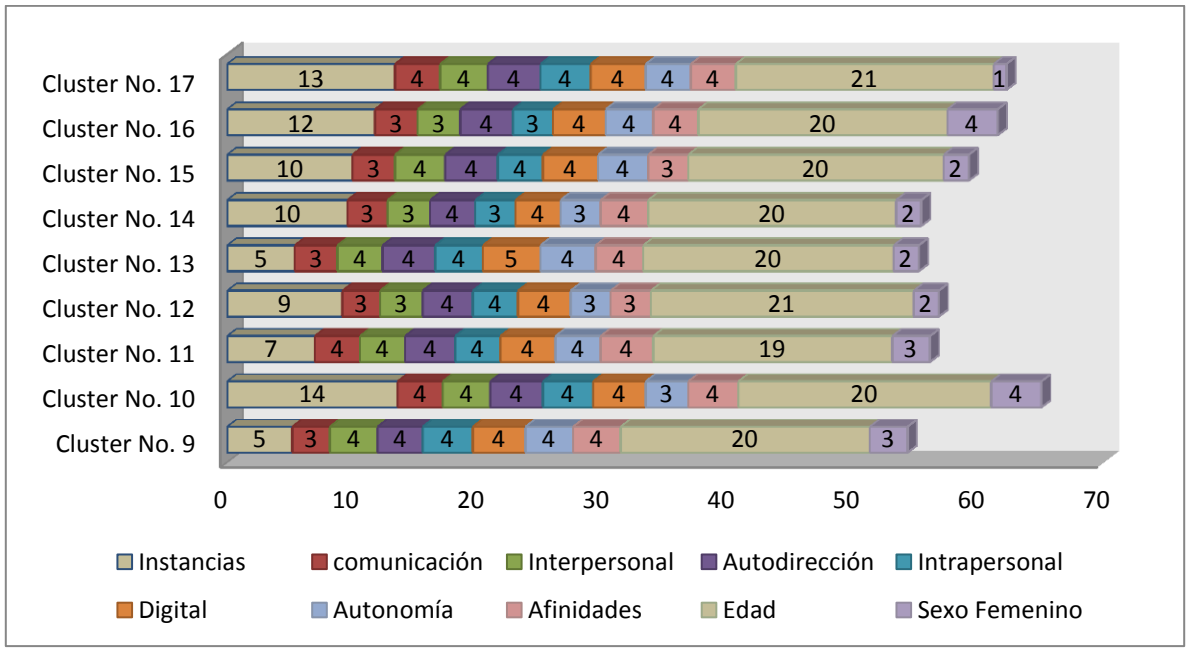


Gráfica 4.6 *Cluster* de colaboradores del 18-27 usando distancia Euclídea.

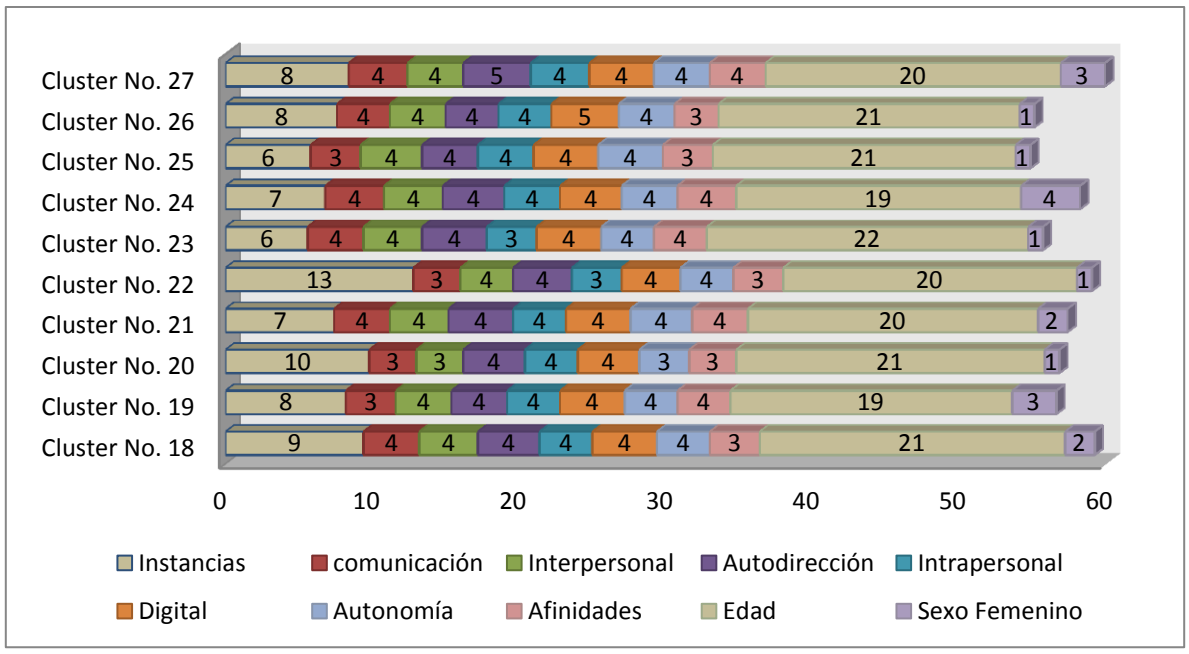
Experimento No.	**3**	**Algoritmo aplicado**	***k-medias***
Objetivo	Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 28 grupos		
Conjunto de datos	277 instancias, enfoque didáctico-pedagógico	**Distancia**	Manhattan
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Creación de 28 grupos de acuerdo a González Rufino (2009). Empleo de todas las variables (demográficas, competencias, habilidades)			
Resultados	El algoritmo *k-medias* creó 16 grupos que pedagógicamente deben conformarse de acuerdo a González Rufino (2009). Los gráficos 4.7-4.9 muestran los resultados obtenidos tras ejecutar el procedimiento descrito.		



Gráfica 4.7 *Cluster* de colaboradores del 0-8 usando distancia Manhattan



Gráfica 4.8 *Cluster* de colaboradores del 9-17 usando distancia Manhattan



Gráfica 4.9 *Cluster* de colaboradores del 18-27 usando distancia Manhattan

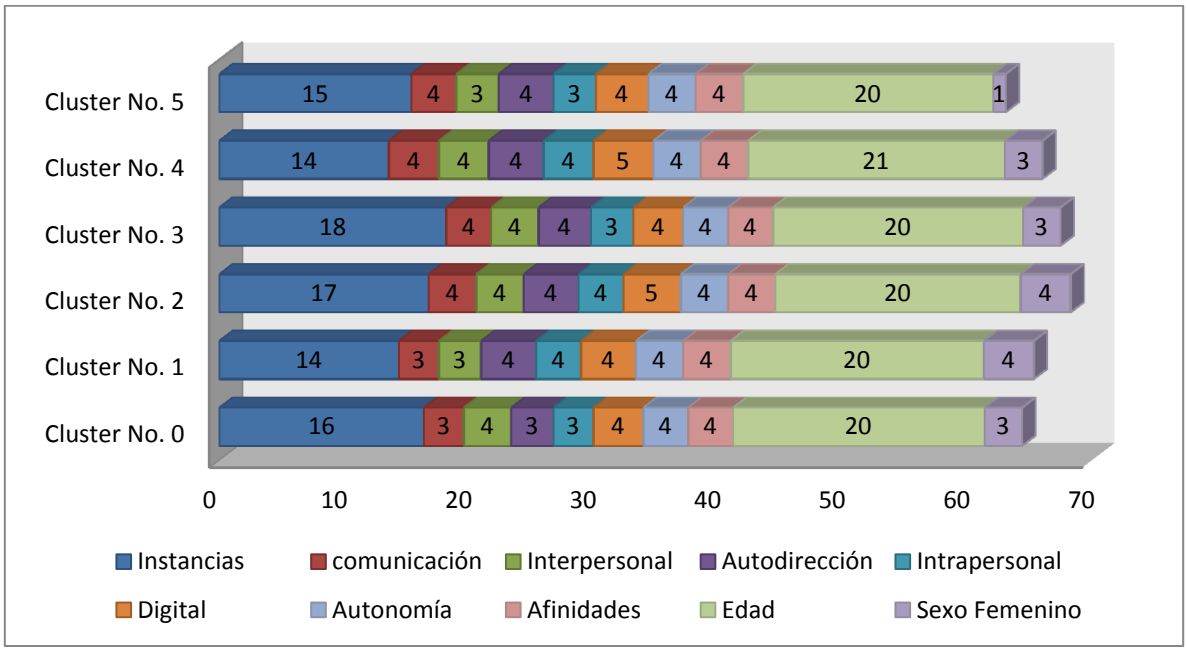
Conclusiones preliminares para formar 28 grupos

Si bien el algoritmo *k-medias* permite realizar agrupación de *n* elementos en función de *x*variables dadas, la distancia Euclídea agrupa a estos elementos (instancias) de acuerdo a la distancia más cercana que existe entre las características de centroide y esta instancia; mientras que la distancia Manhattan agrupa a estas instancias y a todas aquellas instancias que existen en la medida que cumplan con las características del centroide. En otras palabras, la distancia Manhattan al abarcar a un valor más grande, en cuanto a la distancia entre elementos, puede considerar un conjunto mayor de estos elementos que cumplan con las características del centroide.

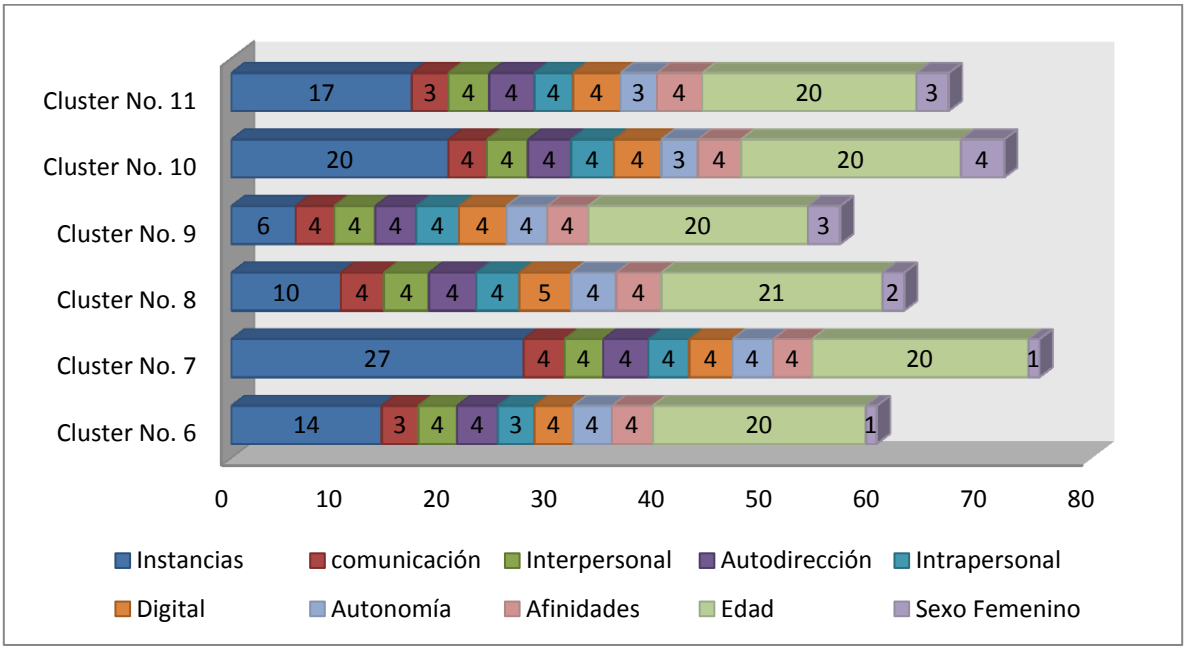
La distancia Manhattan creó 16 *clusters*, mientras que con la distancia Euclídea solamente se crearon 15 *clusters*, que pedagógicamente hablando no exceden los 10 integrantes en cada uno. En otras palabras, *k-medias* devuelve mejores resultados al emplear la distancia Manhattan pues agrupa a un mayor número de elementos que tienen características semejantes con el centroide. El análisis profundo de estos resultados se detalla en la propuesta realizada en los siguientes apartados de este capítulo.

Experimentación para formar18 grupos de trabajo

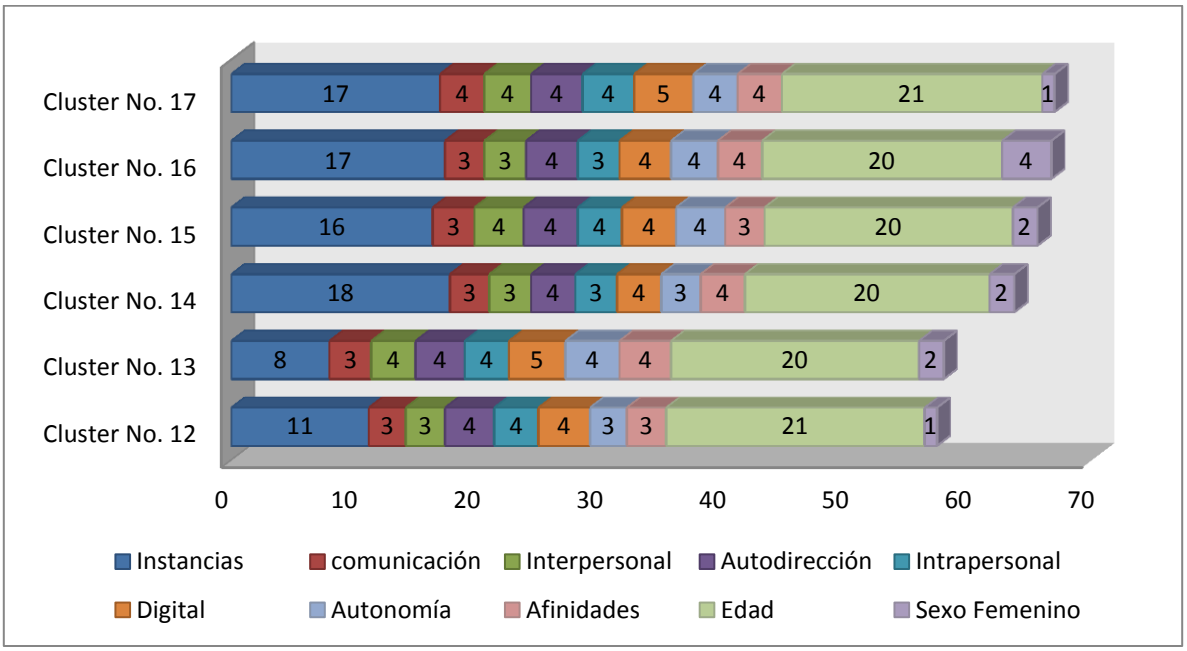
<table>
<tr><td>Experimento No.</td><td>4</td><td>Algoritmo aplicado</td><td colspan="2">k-medias</td></tr>
<tr><td>Objetivo</td><td colspan="4">Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 18 grupos</td></tr>
<tr><td>Conjunto de datos</td><td colspan="2">277 instancias, enfoque didáctico-pedagógico</td><td>Distancia</td><td>Euclídea</td></tr>
<tr><td colspan="5">Procedimiento</td></tr>
<tr><td colspan="5">• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente.
• Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico</td></tr>
<tr><td colspan="5">Excepciones</td></tr>
<tr><td colspan="5">Creación de 18 grupos de acuerdo a el Instituto Nacional de Evaluación Educativa (2013)
Empleo de todas las variables (demográficas, competencias, habilidades)</td></tr>
<tr><td>Resultados</td><td colspan="4">k-medias formó 10 grupos que cumplen con el número máximo de personas correspondiente al no sobrepasar los 16 integrantes. Los gráficos 4.10-4.12 muestran los resultados obtenidos tras ejecutar el procedimiento descrito.</td></tr>
</table>



Gráfica 4.10 *Cluster* de colaboradores del 0-5 usando distancia Euclídea

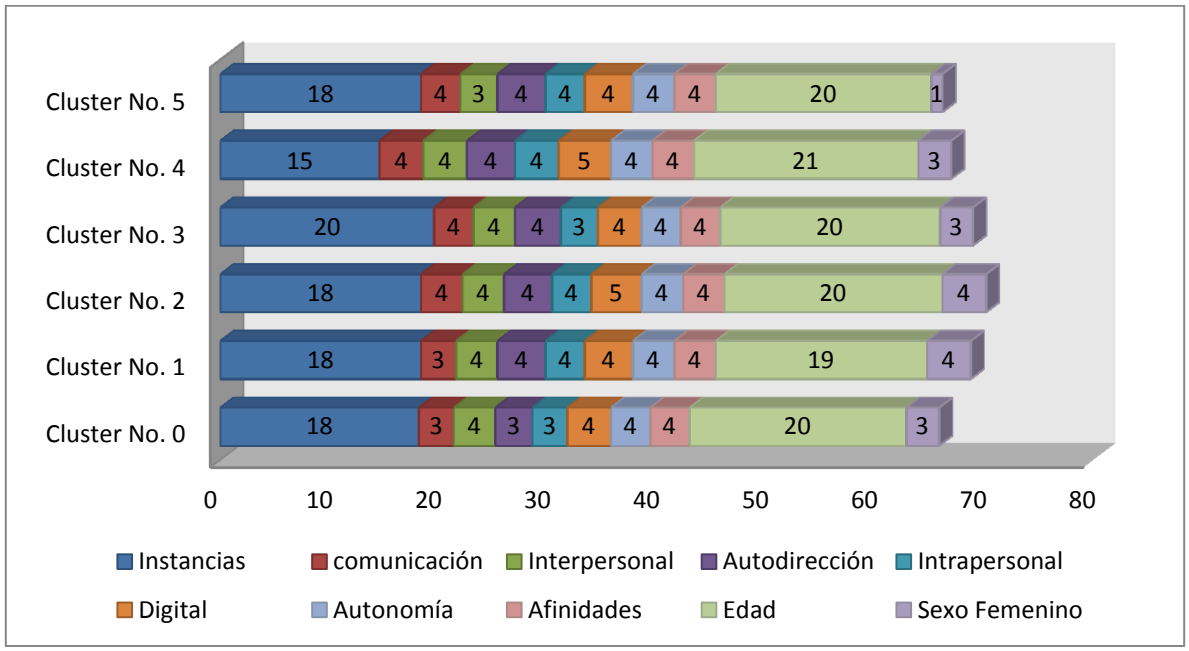


Gráfica 4.11*Cluster* de colaboradores 6-11 usando distancia Euclídea

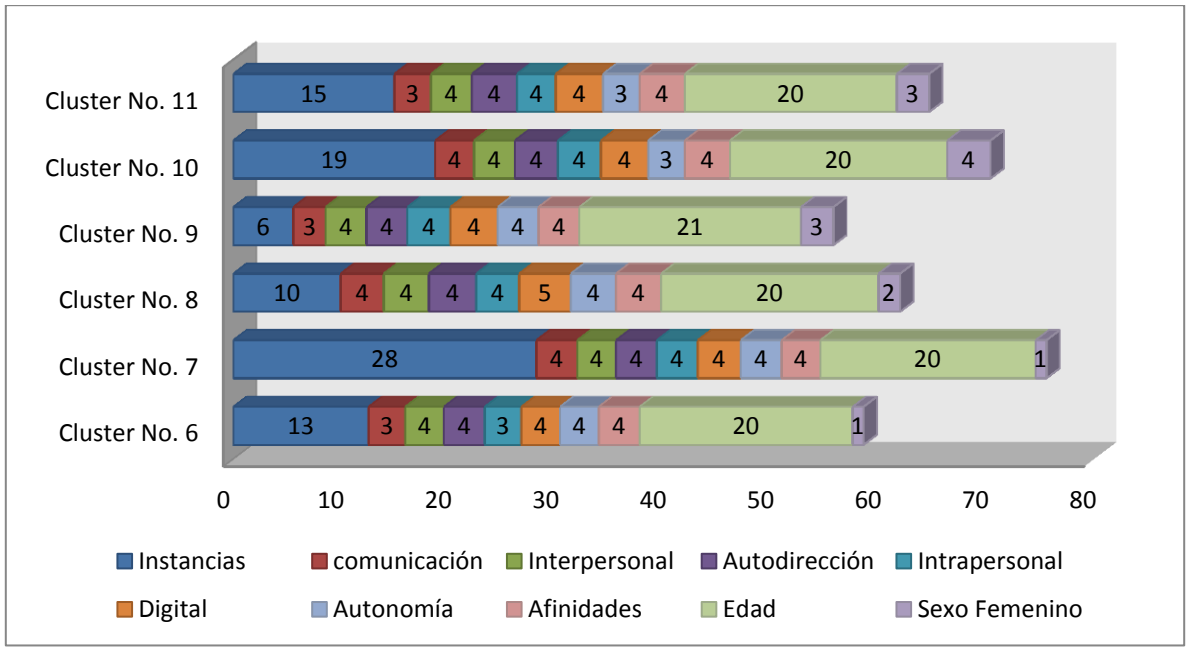


Gráfica 4.12 *Cluster* de colaboradores 12-17 usando distancia Euclídea

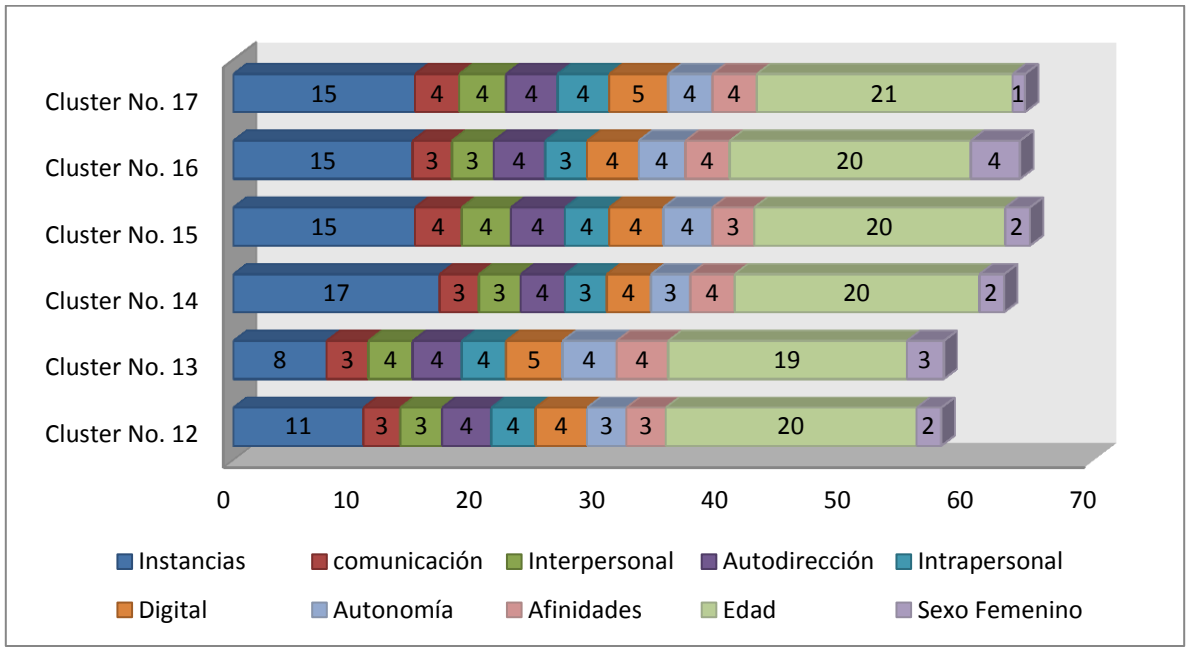
Experimento No.	**5**	**Algoritmo aplicado**	***k-medias***
Objetivo	Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 18 grupos		
Conjunto de datos	277 instancias, enfoque didáctico-pedagógico	**Distancia**	Manhattan
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Creación de 18 grupos de acuerdo a el Instituto Nacional de Evaluación Educativa (2013) Empleo de todas las variables (demográficas, competencias, habilidades)			
Resultados	Al igual que con la distancia Euclídea, *k-medias* formó 10 grupos que cumplen con el número máximo de personas correspondiente al no sobrepasar los 16 integrantes. Los gráficos 4.13-4.15 muestran los resultados obtenidos tras ejecutar el procedimiento descrito.		



Gráfica 4.13 *Cluster* de colaboradores 0-5 usando distancia Manhattan



Gráfica 4.14 *Cluster* de colaboradores 6-11 usando distancia Manhattan



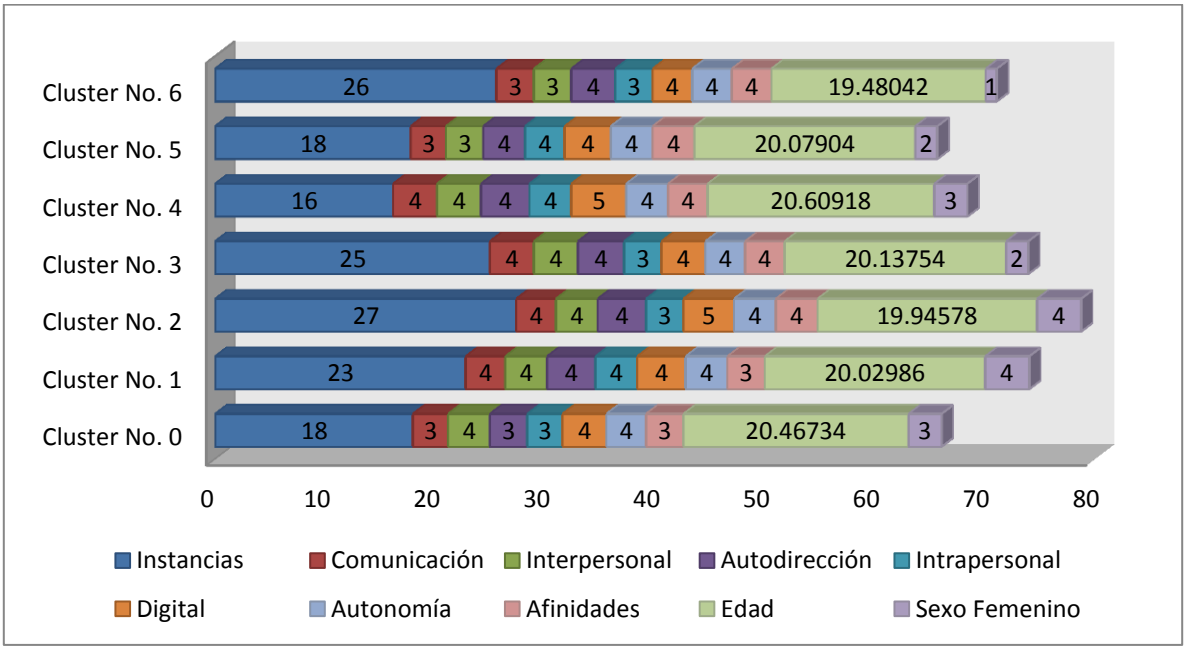
Gráfica 4.15 *Cluster* de colaboradores 6-11 usando distancia Manhattan

Conclusiones preliminares para formar 18 grupos de trabajo

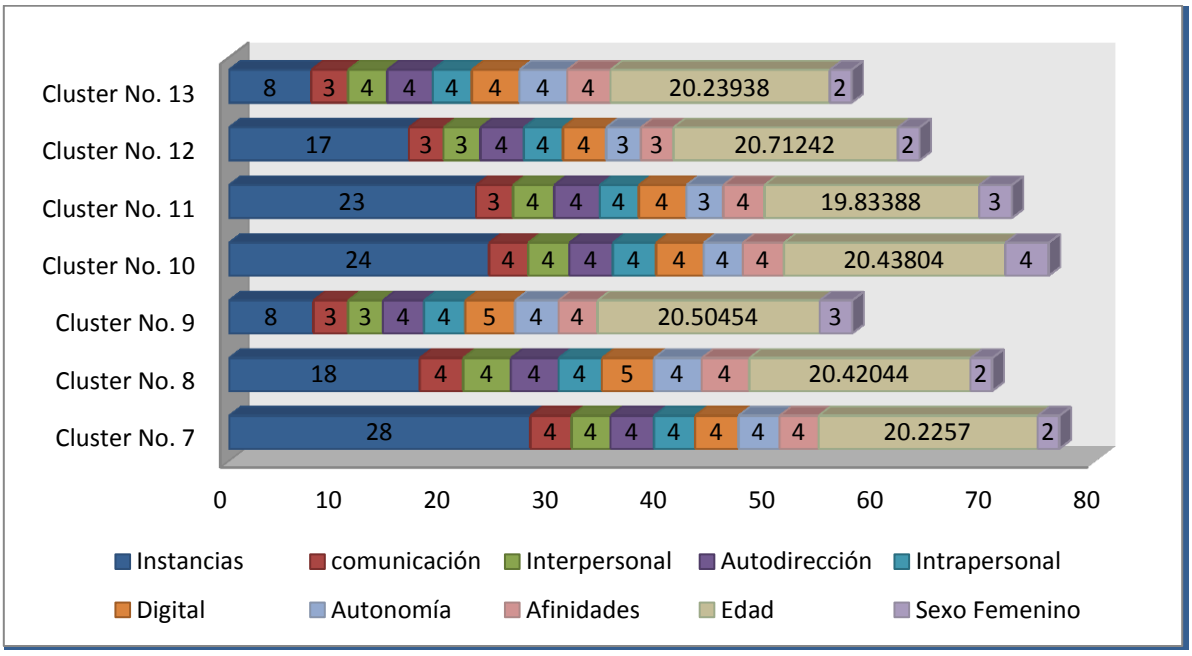
Como se observa en las gráficas anteriores, para *k*=18, el algoritmo al emplear la distancia Manhattan como la distancia Euclídea formó el mismo número de *cluster* que de acuerdo a Maldonado & Giandini (2010) no exceden los 16 integrantes. El análisis profundo de la clasificación de estos *clusters* se detalla en la propuesta realizada en puntos siguientes de este capítulo.

Experimentación para formar 14 grupos de trabajo

Experimento No.	**6**	**Algoritmo aplicado**	***k-medias***
Objetivo	Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 14 grupos		
Conjunto de datos	277 instancias, enfoque didáctico-pedagógico	**Distancia**	Euclídea
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Creación de 14 grupos de acuerdo a Maldonado & Giandini (2010) Empleo de todas las variables (demográficas, competencias, habilidades)			
Resultados	Los gráficos 4.16 y 4.17 muestran los resultados obtenidos tras ejecutar el procedimiento descrito. En donde se observa que solamente 7 *clusters* cumplen con las características requeridas al no exceder sus 20 integrantes.		

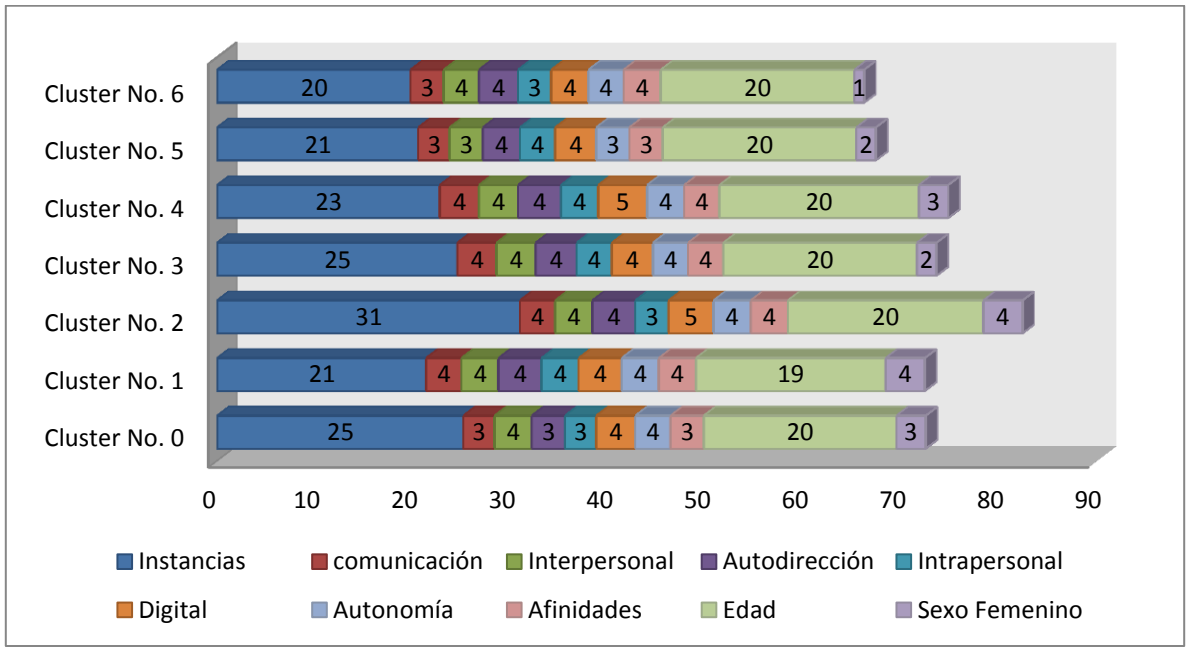


Gráfica 4.16 *Cluster* de colaboradores 0-6 usando distancia Euclídea

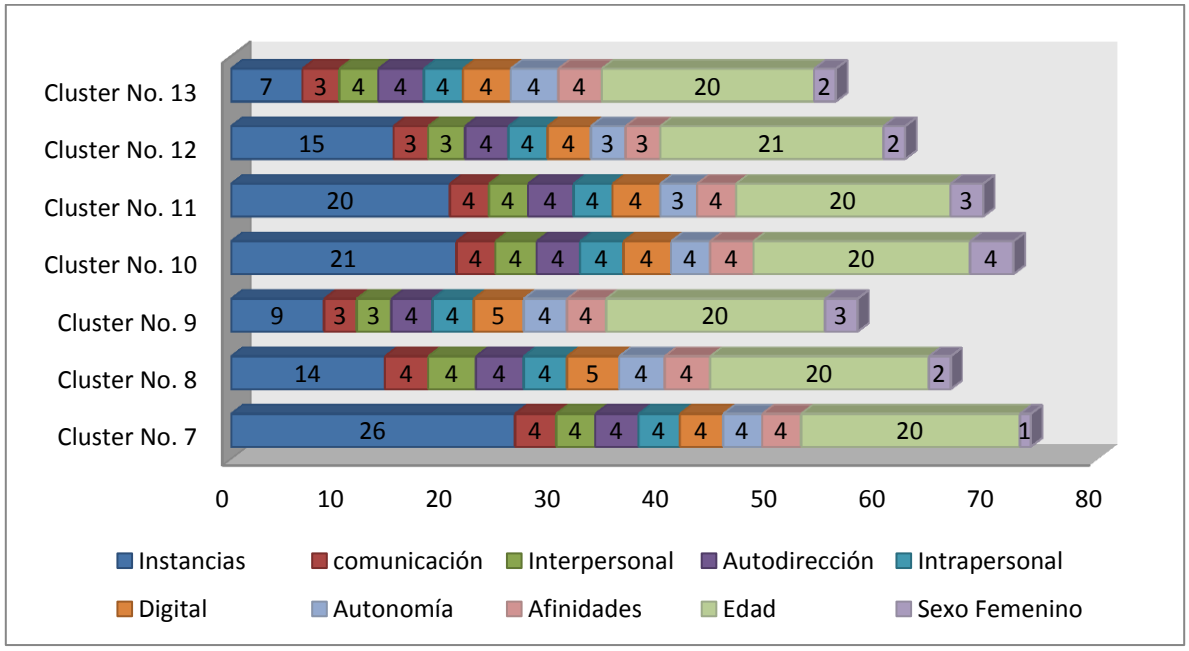


Gráfica 4.17 *Cluster* de colaboradores 7-13 usando distancia Euclídea

Experimento No.	**7**	**Algoritmo aplicado**	***k-medias***
Objetivo	Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 14 grupos		
Conjunto de datos	277 instancias, enfoque didáctico-pedagógico	**Distancia**	Manhattan
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Creación de 14 grupos de acuerdo a Maldonado & Giandini (2010) Empleo de todas las variables (demográficas, competencias, habilidades)			
Resultados	Los gráficos 4.18 y 4.19 muestran los resultados obtenidos tras ejecutar el procedimiento descrito. En donde se observa que solamente 6 *clusters* cumplen con las características requeridas al no exceder sus 20 integrantes.		



Gráfica 4.18 *Cluster* de colaboradores 0-6 usando distancia Manhattan



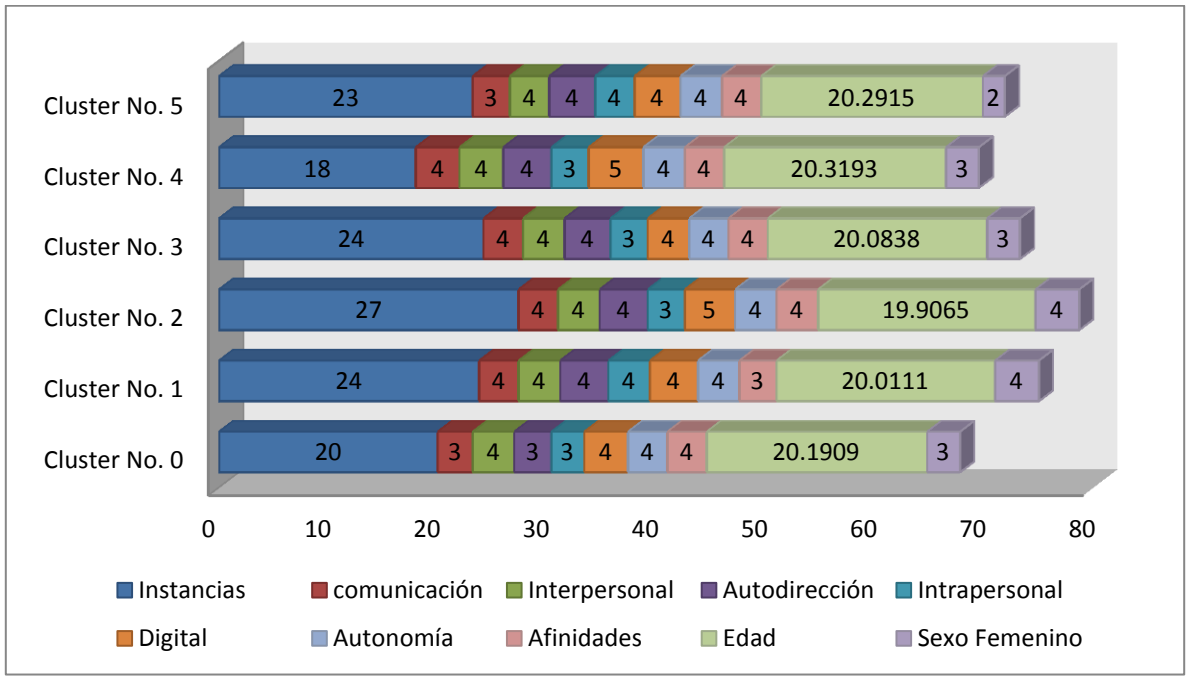
Gráfica 4.19 *Cluster* de colaboradores 7-13 usando distancia Manhattan

Conclusiones preliminares para crear 14 grupos de trabajo

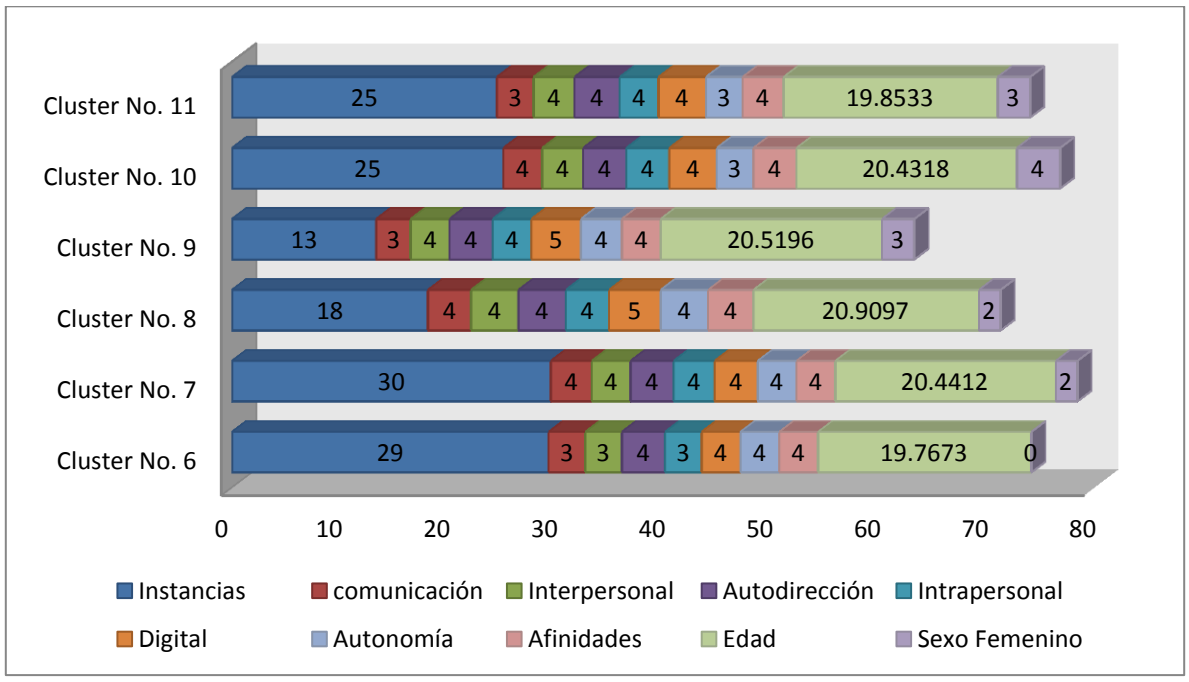
De acuerdo a los resultados obtenidos para *k*=28 y *k*=18 se esperaría que con ambas distancias se crearan el mismo o similar número de *clusters* pedagógicamente adecuados para cualquier valor de *k*. Lo que se comprueba nuevamente con k=14 pues la diferencia entre el número de *cluster* es de 1, esta vez con la distancia Euclídea se logró el mayor número de *clusters* con estas características. El análisis de las competencias y afinidades de cada centroide se detallan en la propuesta realizada en puntos siguientes de este capítulo.

Experimentación para crear 12 grupos de trabajo

<table>
<tr><th>Experimento No.</th><td>8</td><th>Algoritmo aplicado</th><td colspan="2">k-medias</td></tr>
<tr><th>Objetivo</th><td colspan="4">Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 14 grupos</td></tr>
<tr><th>Conjunto de datos</th><td colspan="2">277 instancias, enfoque didáctico-pedagógico</td><th>Distancia</th><td>Euclídea</td></tr>
<tr><th colspan="5">Procedimiento</th></tr>
<tr><td colspan="5">• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente.
• Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico</td></tr>
<tr><th colspan="5">Excepciones</th></tr>
<tr><td colspan="5">Creación de 14 grupos de acuerdo a OCDE (2014)
Empleo de todas las variables (demográficas, competencias, habilidades)</td></tr>
<tr><th>Resultados</th><td colspan="4">Los gráficos 4.20 y 4.21 muestran los resultados obtenidos tras ejecutar el procedimiento descrito. En donde se observa que solamente 7 clusters cumplen con las características requeridas al no exceder sus 24 integrantes.</td></tr>
</table>

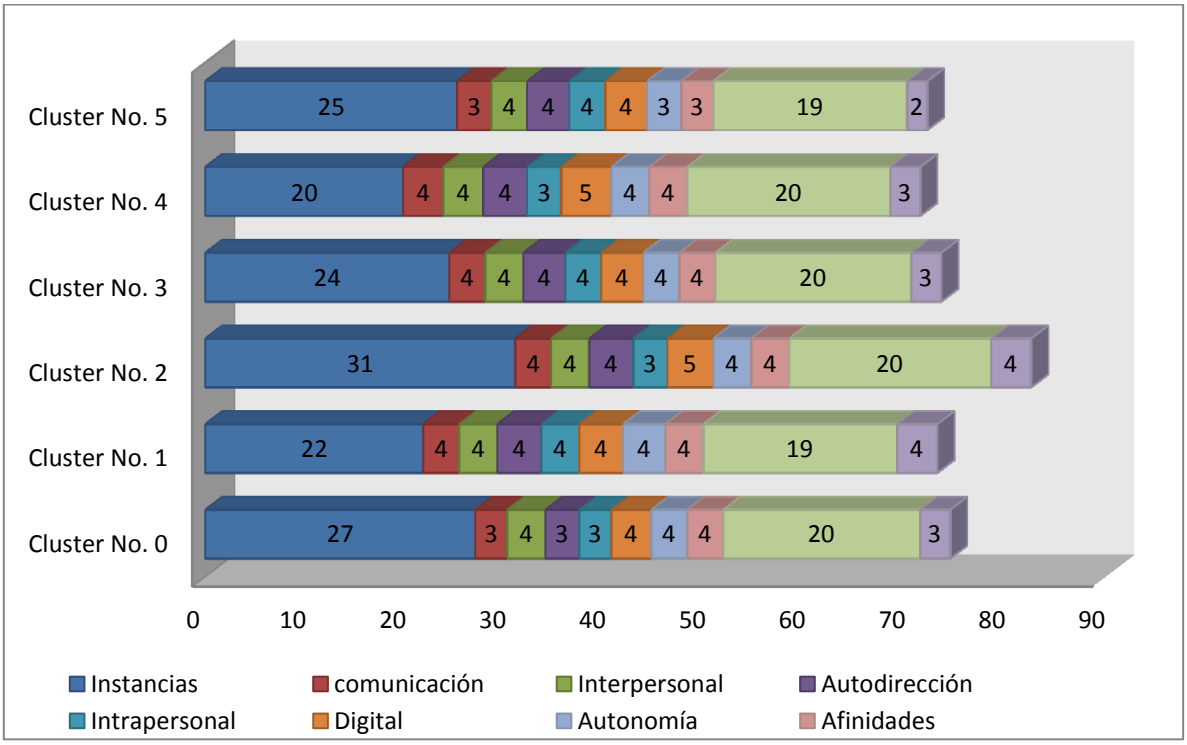


Gráfica 4.20 *Cluster* de colaboradores 0-5 usando distancia Euclídea

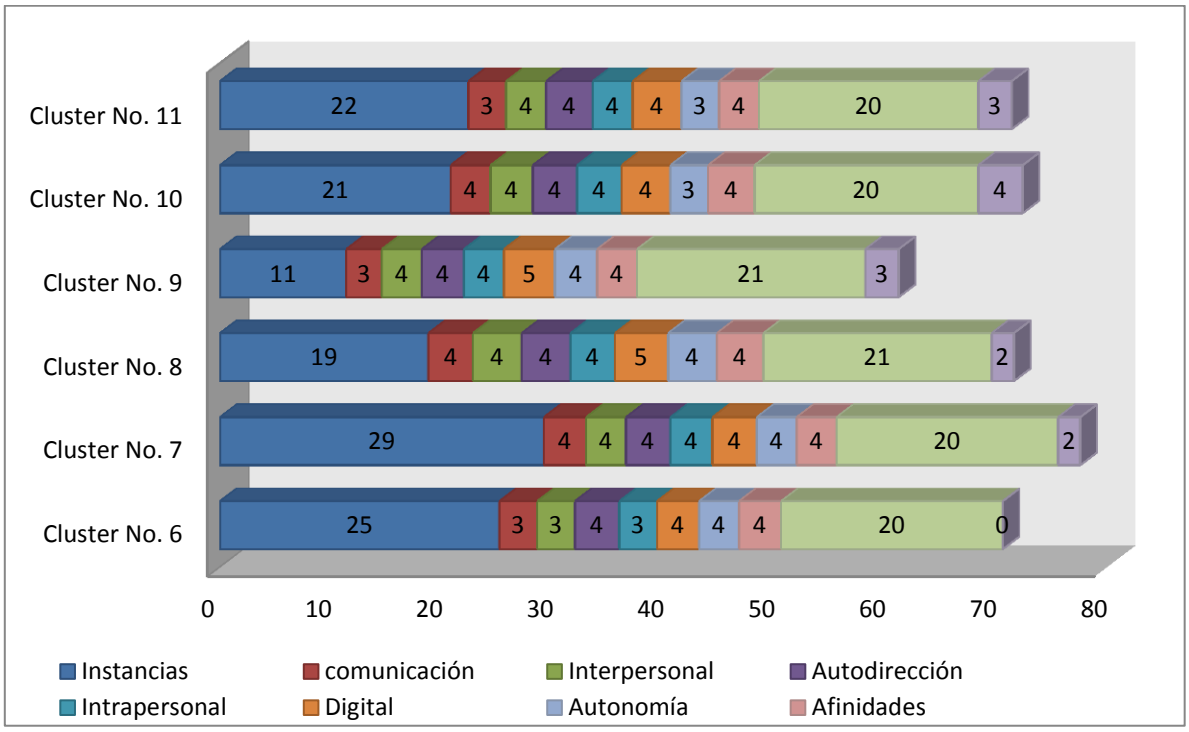


Gráfica 4.21 *Cluster* de colaboradores 6-11 usando distancia Euclídea

Experimento No.	9	Algoritmo aplicado	*k-medias*
Objetivo	Integrar a 277 colaboradores en grupos de trabajo basados en sus competencias y afinidades personales en 14 grupos		
Conjunto de datos	277 instancias, enfoque didáctico-pedagógico	**Distancia**	Manhattan
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Creación de 12 grupos de acuerdo a OCDE (2014) Empleo de todas las variables (demográficas, competencias, habilidades)			
Resultados	Los gráficos 4.22 y 4.23 muestran los resultados obtenidos tras ejecutar el procedimiento descrito. En donde se observa que solamente 7*clusters* cumplen con las características requeridas al no exceder sus 24 integrantes.		



Gráfica 4.22 *Cluster* de colaboradores 0-5 usando distancia Manhattan



Gráfica 4.23 *Cluster* de colaboradores 6-11 usando distancia Manhattan

Conclusiones preliminares para formar 12 grupos de trabajo

Como se notó anteriormente, para los diferentes valores de k y empleando ambas distancias, se obtiene el mismo número de *cluster*s o con diferencia de uno. Sin embargo, la importancia de identificar las características en cuanto a competencias y afinidades que tiene cada centroide, se detallan en la propuesta realizada en puntos siguientes de este capítulo.

Comparativo general de la experimentación desde un enfoque didáctico-pedagógico

En este sentido, la tabla 4.1 muestra un comparativo de los resultados obtenidos, desde el punto de vista didáctico-pedagógico, empleando distancia Euclídea y distancia Manhattan.

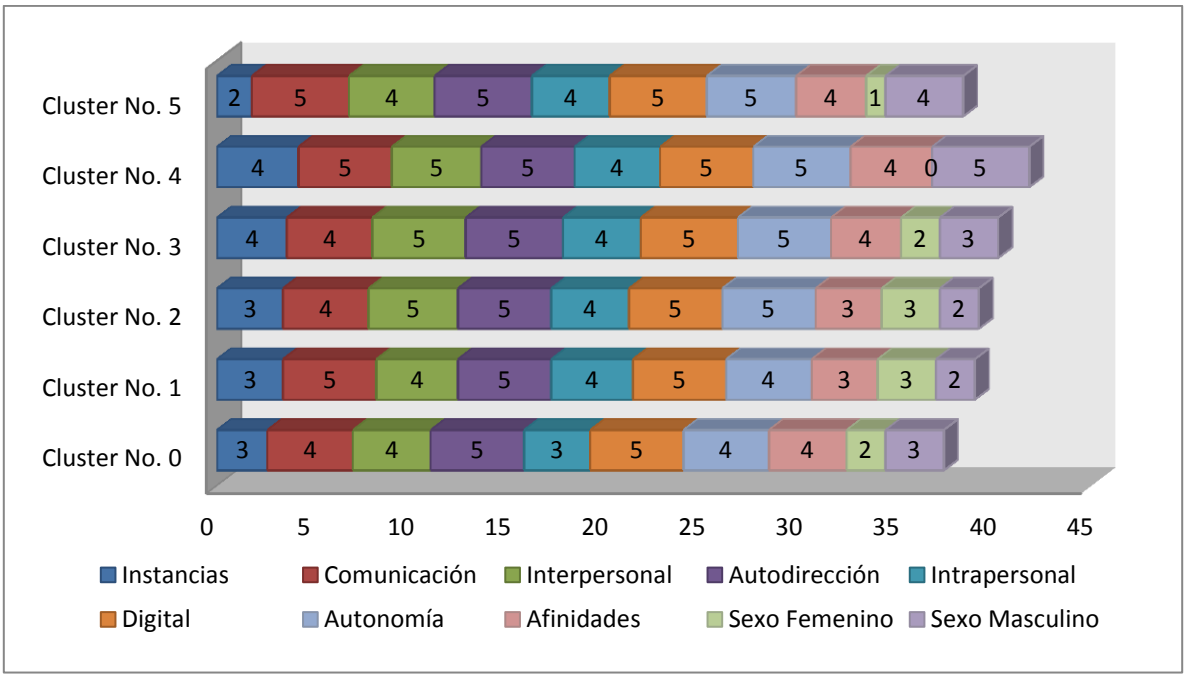
Criterios	Distancia Euclídea				Distancia Manhattan			
	k=28	*k*=18	*k*=14	*k* =12	*k* =28	*k* =18	*k* =14	*k* =12
No. de *cluster*s formados pedagógicamente adecuados	15	10	7	7	16	10	6	6
Promedio de iteraciones	7	6	7	6	7	5	5	6
Promedio de tiempo en construir el modelo (en segundos)	0.062	0.038	0.078	0.02	0.088	0.038	0.03	0.32

Tabla 4.1 Resultados obtenidos con la distancia Euclídea y distancia Manhattan.

La tabla 4.1 indica que en cuanto al número de *clusters* pedagógicamente formados, solamente existe una diferencia de un *cluster* entre la distancia Euclídea y Manhattan respectivamente. Que en cuanto a las iteraciones en promedio el algoritmo *k-medias* al emplear la distancia Euclídea realizó 7 iteraciones en 0.0495 segundos, mientras que con la distancia Manhattan 6 iteraciones en 0.1865; indicando con estos datos que el algoritmo *k-medias* tiene mejor desempeño al utilizar la distancia Euclídea al invertir el menor tiempo aunque realiza el mayor número de iteraciones, mientras que con Manhattan su desempeño es menor aunque realiza un menor número de iteraciones.

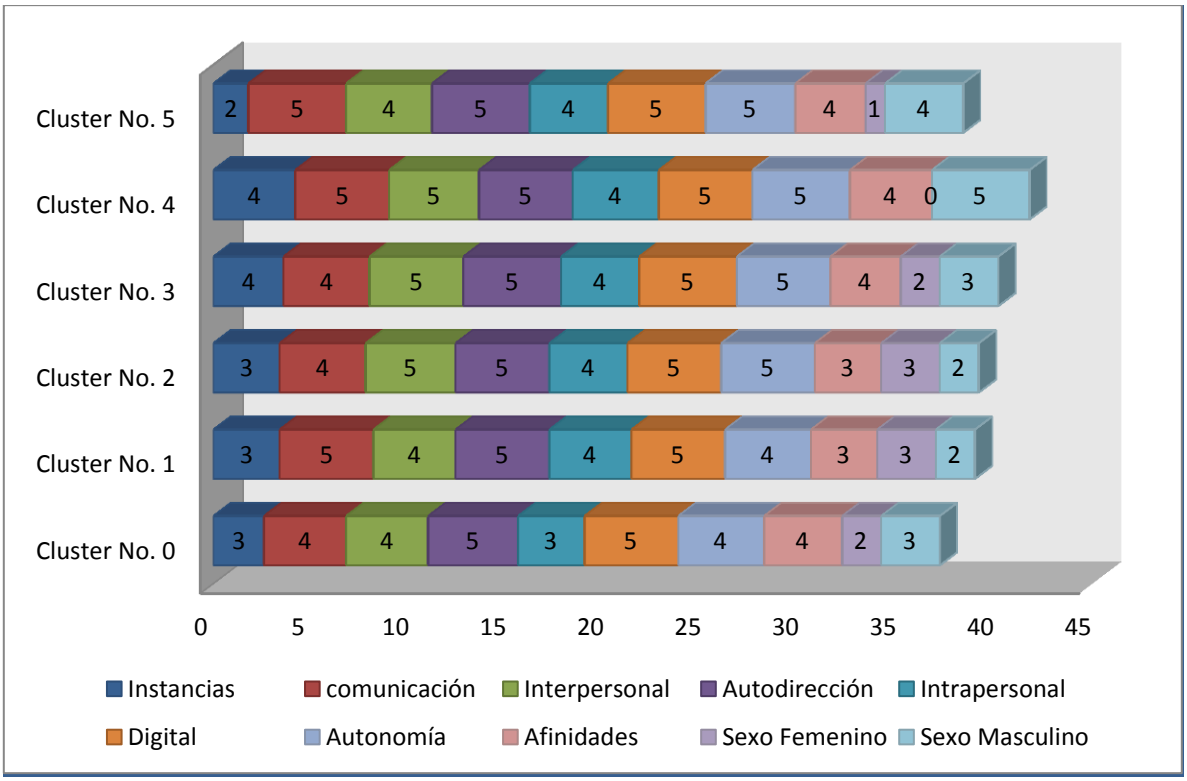
Experimentación para agrupación de líderes

Experimento No.	10	Algoritmo aplicado	*k-medias*
Objetivo	Integrar a 19 líderes basados en sus competencias y afinidades personales		
Conjunto de datos	19 instancias, enfoque didáctico-pedagógico	**Distancia**	Euclídea
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Creación de 6 grupos Empleo de todas las variables (demográficas solamente sexo, competencias, habilidades)			
Resultados	La gráfica 4.24 muestra el resultado obtenido tras ejecutar el procedimiento descrito.		



Gráfica 4.24 *Clusters* de líderes obtenidos empleando distancia Euclídea

Experimento No.	11	Algoritmo aplicado	*k-medias*
Objetivo	Integrar a 19 líderes basados en sus competencias y afinidades personales		
Conjunto de datos	19 instancias, enfoque didáctico-pedagógico	**Distancia**	Manhattan
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Creación de 6 grupos Empleo de todas las variables (demográficas solamente sexo, competencias, habilidades)			
Resultados	La gráfica 4.25 muestra el resultado obtenido tras ejecutar el procedimiento descrito.		



Gráfica 4.25 *Clusters* de líderes obtenidos empleando distancia Manhattan

4.2.2 Experimentación con un enfoque empresarial

Debido a que no existe una fórmula secreta para establecer el número de personas que pueden integrar un equipo de trabajo para realizar un proyecto determinado; la formación de grupos en contextos distintos al ámbito educativo, queda sujeta a lo que se mencionó en el capítulo II, en donde Witten & Frank (2005) señalan que una solución al desconocer el número de *clusters* a formar, consiste en probar diferentes posibilidades y ver cuál es la mejor.

Witten & Frank (2005) proponen una estrategia simple, que consiste en partir de un mínimo determinado y probar hasta un valor máximo fijo, utilizando validación cruzada para encontrar el mejor valor. Es importante considerar que en los datos de entrenamiento la "mejor" agrupación de acuerdo con el criterio de la distancia cuadrada total siempre será elegir tantos grupos como puntos de datos existan.

En este sentido, extrapolando los resultados obtenidos de la experimentación desde un enfoque pedagógico, se realizó una búsqueda de información acerca de las 10 empresas más importantes de México (CNNexpansion, 2014) para identificar cuántos líderes (jefes o directivos en el organigrama) tiene cada empresa y con base en ello, realizar la agrupación en función de sus competencias y afinidades. La tabla 4.2 muestra estos datos.

Empresa	No. de líderes[5]
1. Petróleos Mexicanos[6]	17
2. América Móvil[7]	6
3. *Walmart* de México[8]	26
4. Comisión Federal de Electricidad[9]	13

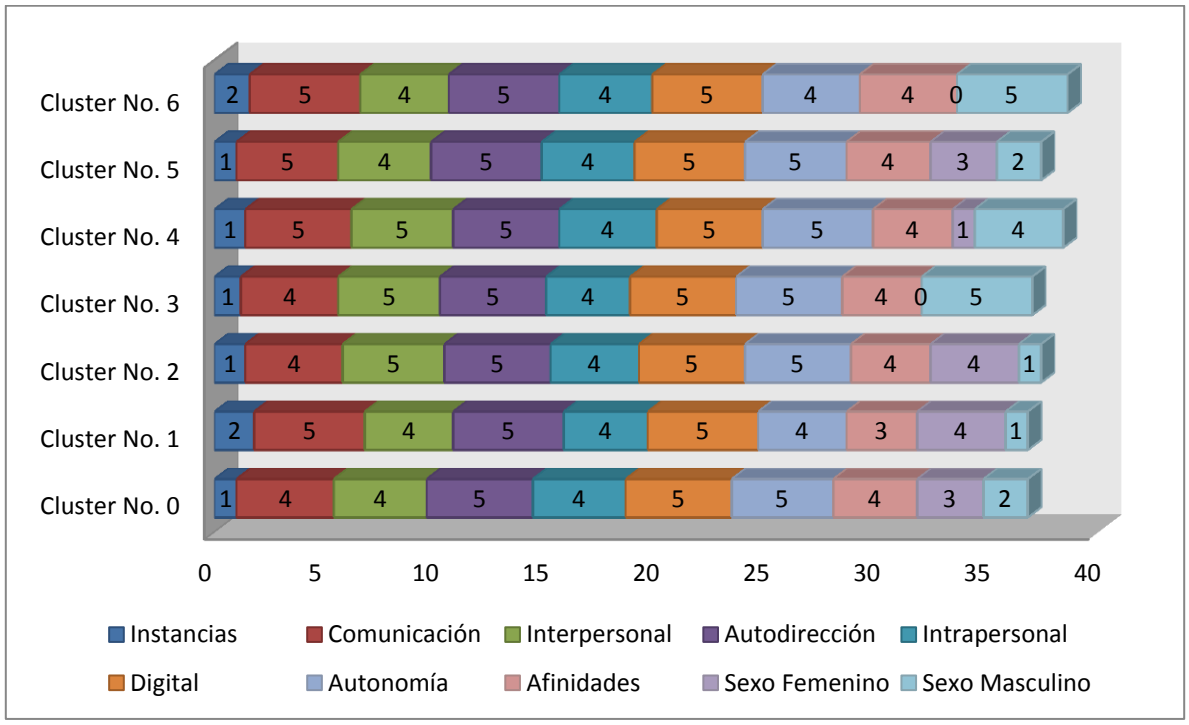
[5] Se consideraron únicamente a los puestos que aparecen en el organigrama de cada empresa citada.
[6] Información obtenida de http://www.pemex.com/acerca/quienes_somos/Paginas/funcionarios.aspx#.VAT5f6OH130
[7] Información obtenida de http://www.americamovil.com.mx/amx/es/cm/about/directory.html?p=1&s=8
[8] Información obtenida de: http://www.theofficialboard.es/organigrama/wal-mart-stores
[9] Información obtenida de: g http://app.cfe.gob.mx/Aplicaciones/QCFE/OrganigramaDigital/

5. FEMSA[10]	9
6. General Motors de México	12
7. Alfa[11]	12
8. Cemex[12]	47
9. Grupo Bimbo[13]	22
10. Chrysler de México[14]	16

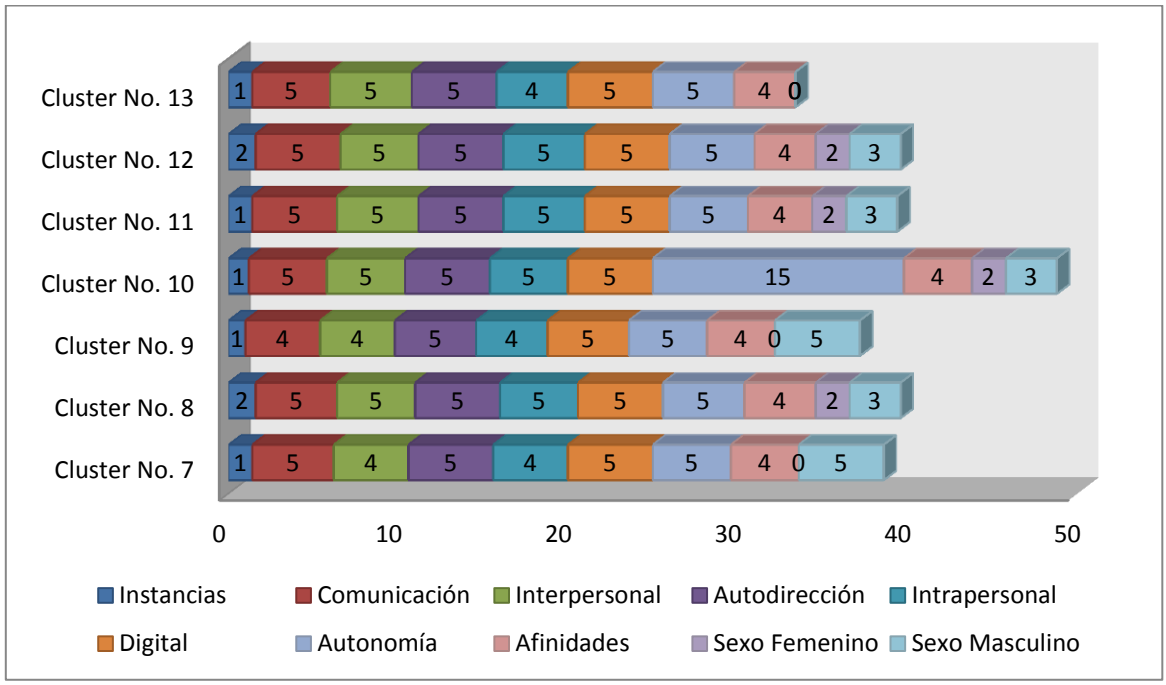
Tabla 4.2 Las diez empresas más importantes de México y sus líderes. (CNNexpansion, 2014)

Experimento No.	**12**	**Algoritmo aplicado**	***k-medias***
Objetivo	Integrar a 19 líderes basados en sus competencias y afinidades personales		
Conjunto de datos	19 instancias, empresarial	**Distancia**	Euclídea
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
De acuerdo a la tabla 4.2 el promedio de los grupos que podrían formarse serían 18 *cluster*s; debido a que solamente se tienen 19 instancias para experimentar; se considera el promedio de las primeras cinco empresas más importantes de México, cuyo valor es 14. Empleo de todas las variables (demográficas solamente sexo, competencias, habilidades)			
Resultados	Las gráficas 4.26 y 4.27 muestran los resultados obtenidos tras ejecutar el procedimiento descrito.		

[10] Información obtenida de: http://ir.femsa.com/mx/management.cfm
[11] Información obtenida de: http://www.alfa.com.mx/NC/equipo-directivo.htm
[12] Información obtenida de: http://www.cemex.com/ES/Inversionistas/EstructuraCorporativa.aspx
[13] Información obtenida de: http://www.grupobimbo.com/es/grupo-bimbo/estructura.html
[14] Información obtenida de : http://www.chryslerdemexico.com.mx/company/team/index.php

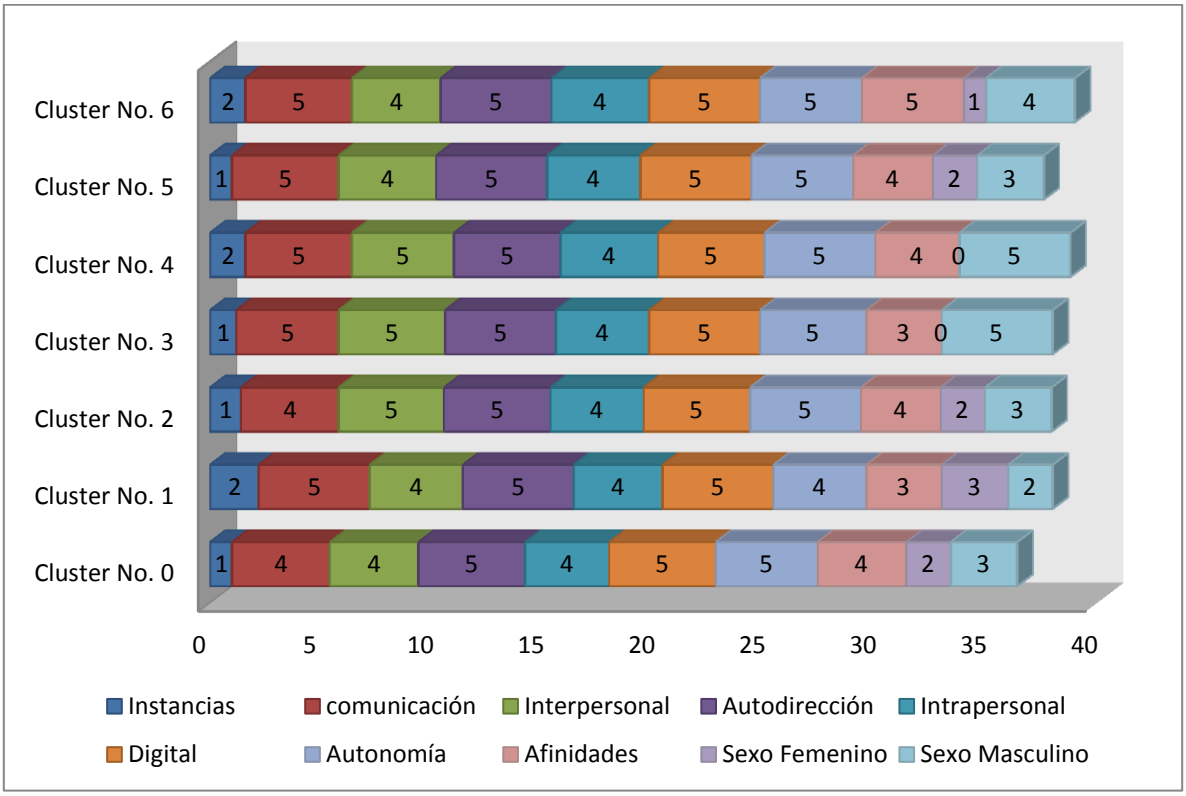


Gráfica 4.26 *Cluster* de líderes 0-6 para k=14 empleando distancia Euclídea

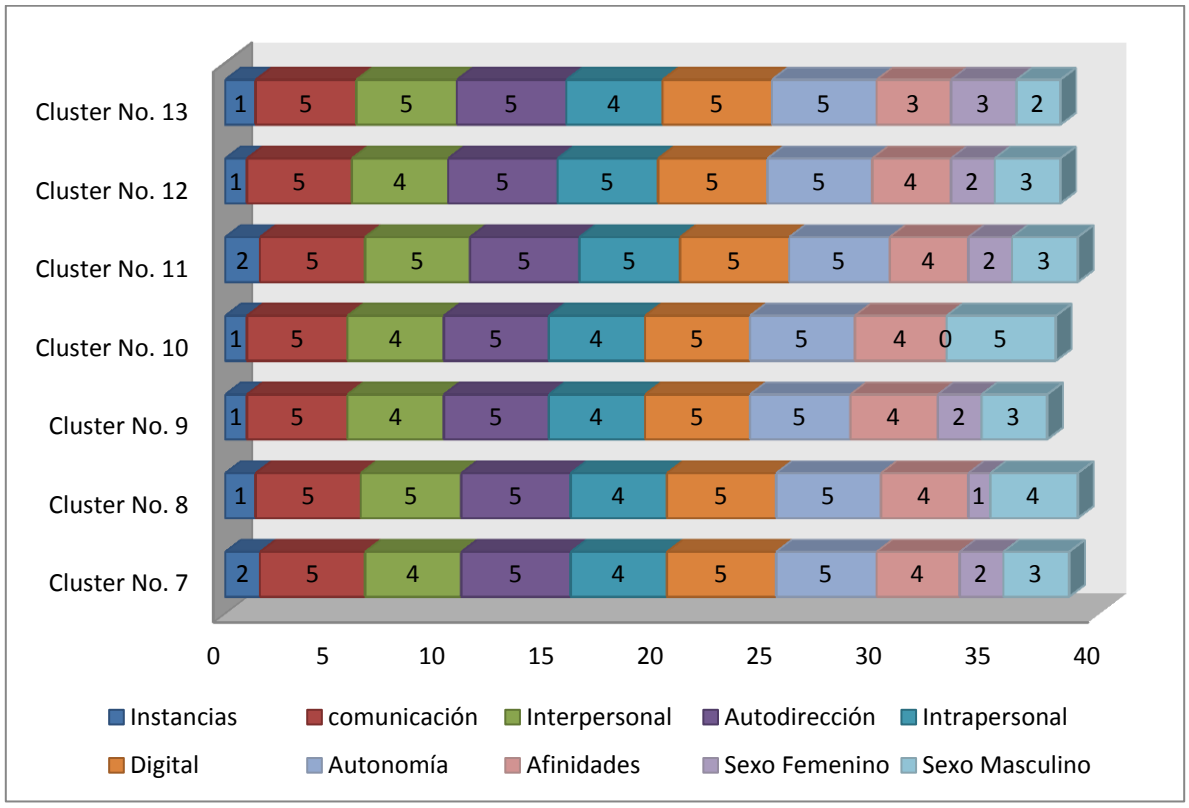


Gráfica 4.27 *Cluster* de líderes 7-13 para k=14 empleando distancia Euclídea

Experimento No.	13	Algoritmo aplicado	*k-medias*
Objetivo	Integrar a 19 líderes basados en sus competencias y afinidades personales		
Conjunto de datos	19 instancias, empresarial	**Distancia**	Manhattan
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
De acuerdo a la tabla 4.2 el promedio de los grupos que podrían formarse serían 18 *cluster*s; debido a que solamente se tienen 19 instancias para experimentar; se considera el promedio de las primeras cinco empresas más importantes de México, cuyo valor es 14. Empleo de todas las variables (demográficas solamente sexo, competencias, habilidades)			
Resultados	Las gráficas 4.28 y 4.29 muestran los resultados obtenidos tras ejecutar el procedimiento descrito.		



Gráfica 4.28 *Cluster* de líderes 0-6 para k=14 usando distancia Manhattan



Gráfica 4.29 *Cluster* de líderes 7-13 para k=14 usando distancia Manhattan

El análisis de las competencias y afinidades de cada *cluster* se detallan en la propuesta realizada posteriormente en este capítulo.

4.3. Método para la identificación automática de características del capital humano usando algoritmos de selección de atributos

El método descrito en este apartado, tiene con fin seleccionar al capital humano con base a sus competencias y afinidades a fin de caracterizar a un líder o un colaborador. Para ello, se emplea el algoritmo selección de atributos usando cinco evaluadores: *CfsSubsetEval*, *GainRatioAttributeEval*, *InfoGainAttributeEval*, *PrincipalComponentes* y *SymetricalUncerAttributeEval*; mismos que aplican diferentes métodos de búsqueda, para la experimentación solamente se emplearon *RandomSearch* (para el primer evaluador) y *Ranker* (para los cuatro evaluadores restantes).

Para seleccionar las habilidades y afinidades de mayor relevancia en el conjunto de datos de líderes y colaboradores se utilizaron cinco métodos de selección de atributos disponibles en Weka en su versión 3.6.11. La selección de atributos incluyó la combinación de un evaluador de atributos y un método de búsqueda(The University of Waikato, 2013).

El primer algoritmo utilizado fue *CfsSubsetEval* en combinación con el método *RandomSearch* que realiza una búsqueda aleatoria en el espacio de subconjuntos de atributos. *CfsSubsetEval* evalúa un subconjunto de atributos considerando la habilidad predictiva individual de cada variable, así como el grado de redundancia entre ellas.

Los cuatro algoritmos restantes son evaluadores de atributos individuales y cada uno se aplicó unido al método *Ranker*, que devuelve una lista ordenada de los atributos según su calidad:

- *GainRatioAttributeEval*: evalúa cada atributo midiendo su razón de beneficio con respecto a la clase.
- *InfoGainAttributeEval*: evalúa los atributos midiendo la ganancia de información de cada uno con respecto a la clase. Previamente discretiza los atributos numéricos.
- *PrincipalComponents:*realiza un análisis de componentes principales y la transformaciónde los datos.
- *SymmetricalUncerAttributeEval:* evalúa el valor de un atributo mediante la medición de la incertidumbre simétrica con respecto a la clase.

Experimentación para identificar las características de líderes

Experimento No.	14	Algoritmo aplicado	Selección de atributos
Objetivo	Encontrar las características más importantes del capital humano		
Conjunto de datos	19 instancias, enfoque empresarial		
Evaluador	*CfsSubsetEval*	**Método de Búsqueda**	*RandomSearch*
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de las cinco ejecuciones y presentarlas en un gráfico			
Excepciones			
Eliminación de las variables demográficas			
Resultados	La figura 4.12 muestra el resultado de la ejecución del algoritmo, mientras que la figura 4.13 muestra el resultado promedio de dicho evaluador		

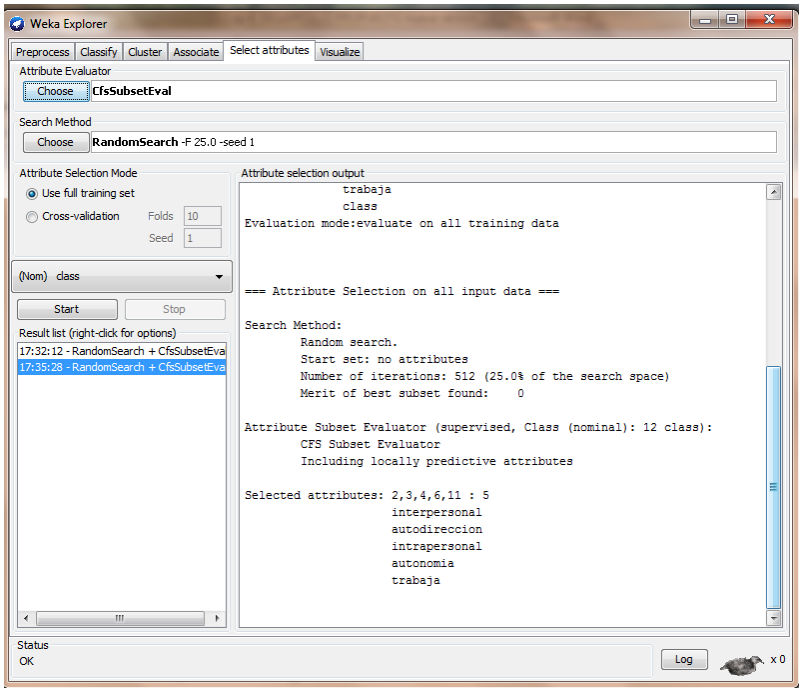


Figura 4.12 Ejecución del evaluador de atributos *CfcsubsetEval*

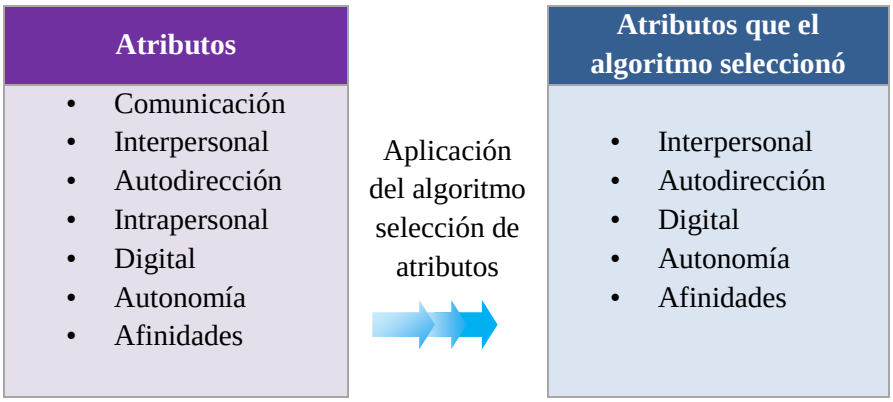


Figura 4.13 Resultados del evaluador de atributos *CfsSubsetEval* para líderes

La figura 4.13 muestra que de los siete atributos, relacionados a 6 competencias y una afinidad, el evaluador *CfcSubsetEval* únicamente seleccionó 5 atributos, los cuales se presentan en orden de prioridad.

Experimento No.	**15**	**Algoritmo aplicado**	**Selección de atributos**
Objetivo	Encontrar las características más importantes del capital humano		
Conjunto de datos	19 instancias, enfoque empresarial		
Evaluador	*GainRatioAttributeEval*, *InfoGainAttributeEval*, *PrincipalComponents* *SymericalUncerAttributeEval*	**Método de Búsqueda**	*Ranker*
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente.			
Excepciones			
Eliminación de las variables demográficas Para los cuatro evaluadores individuales, solamente se realizó una ejecución debido a que no tienen como atributo el valor de la semilla			
Resultados	La figura 4.14 muestra el resultado promedio de cada evaluador		

GainRatioAttributeEval	*InfoGainAttributeEval*	*PrincipalComponents*	*SymericalUncerAttributeEval*
7: Afinidades	7: Afinidades	3: Autodirección	7: Afinidades
3: Autodirección	3: Autodirección	5: Digital	3: Autodirección
2: Interpersonal	2: Interpersonal	4: Intrapersonal	2: Interpersonal
4: Intrapersonal	3: Intrapersonal	6: Autonomía	3: Autodirección
6: Autonomía	6: Autonomía	2: Interpersonal	6: Autonomía
5: Digital	5: Digital	1: Comunicación	5: Digital
1: Comunicación	1: Comunicación	7: Afinidades	1: Comunicación

Figura 4.14 Resultados de los evaluadores de atributos individuales

Experimentación para identificar características de colaboradores

Experimento No.	**16**	**Algoritmo aplicado**	**Selección de atributos**
Objetivo	Encontrar las características más importantes del capital humano		
Conjunto de datos	277 instancias, enfoque empresarial		
Evaluador	*CfsSubsetEval*	**Método de Búsqueda**	*RandomSearch*
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio de ellas y mostrar los resultados			
Excepciones			
Eliminación de las variables demográficas			
Resultados	La figura 4.15 muestra una ejecución del algoritmo, y la figura 4.16 su respectivo resultado promedio del evaluador		

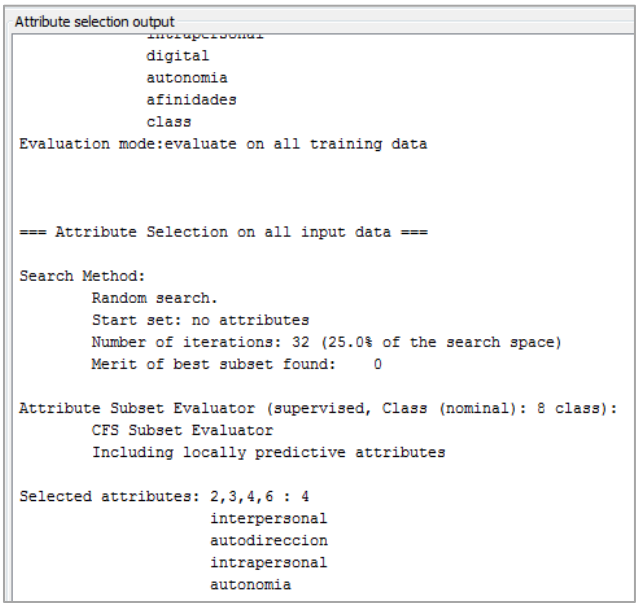


Figura 4.15 Ejecución del evaluador *CfsSubsetEval* para colaboradores

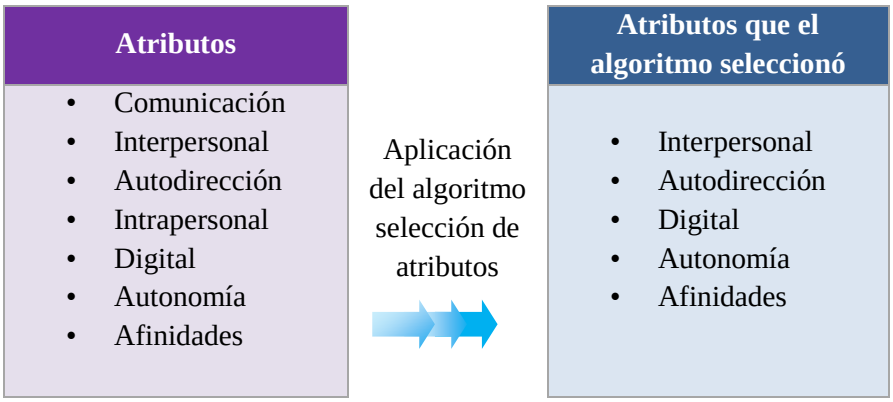


Figura 4.16 Resultados generales del evaluador *CfsSubsetEval* para colaboradores

En la figura 4.16 se muestra la lista de atributos (del lado izquierdo) que se generaron a partir del instrumento H-A y del lado derecho aquellos atributos que el evaluador *CfsSubsetEval* seleccionó como los más importantes y ordenó de acuerdo a su relevancia.

Experimento No.	17	Algoritmo aplicado	Selección de atributos
Objetivo	Encontrar las características más importantes del capital humano		
Conjunto de datos	277 instancias, enfoque empresarial		
Evaluador	*GainRatioAttributeEval* *InfoGainAttributeEval* *PrincipalComponents* *SymericalUncerAttributeEval*	**Método de Búsqueda**	*Ranker*
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente.			
Excepciones			
Eliminación de las variables demográficas Para los cuatro evaluadores individuales, solamente se realizó una ejecución debido a que no tienen como atributo el valor de la semilla			
Resultados	La figura 4.17 muestra el respectivo resultado de cada evaluador		

ainRatioAttributeEval	*InfoGainAttributeEval*	*PrincipalComponents*	*SymericalUncerAttributeEval*
7: Afinidades 3: Autodirección 2: Interpersonal 4: Intrapersonal 6: Autonomía 5: Digital 1: Comunicación	7: Afinidades 3: Autodirección 2: Interpersonal 4: Intrapersonal 6: Autonomía 5: Digital 1: Comunicación	2: Interpersonal 3: Autodirección 6: Autonomía 5: Digital 4: Intrapersonal 1: Comunicación 7: Afinidades	7: Afinidades 3: Autodirección 2: Interpersonal 4: Intrapersonal 6: Autonomía 5: Digital 1: Comunicación

Figura 4.17 Resultados de los evaluadores de atributos individuales para colaboradores

La figura 4.17 muestra la lista de atributos ordenados de acuerdo a importancia de cada uno de los evaluadores aplicados al conjunto de datos de colaboradores.

4.4. Método para la clasificación de individuos aplicando el algoritmo árboles de decisión

En este apartado se describe el método para identificar el atributo más importante que caracteriza a un líder y colaborador; esto a partir de la aplicación del algoritmo árboles de decisión y con ello obtener un árbol que permita reconocer aquellos atributos sobresalientes que identifiquen a un líder de un colaborador.

Se utilizó el algoritmo árboles de decisión C4.5 (método J48 en el WEKA) para identificar la competencia más importante a considerar en un líder y colaborador; así como para clasificar entre estos dos tipos de roles. Este algoritmo de aprendizaje fue aplicado al conjunto de datos de líderes y colaboradores; mismo que suman 296 instancias. El algoritmo C4.5 fue seleccionado para este estudio, puesto que trata de apuntar directamente hacia los atributos relevantes y de ignorar los irrelevantes(Han, Kamber, & Pei, 2011); así como se trata de una herramienta más descriptiva que puede ser muy útil en la toma de decisiones.

Experimento No.	**18**	**Algoritmo aplicado**	**Árboles de decisión**
Objetivo	Caracterizar a colaboradores en función de sus habilidades y afinidad personal		
Conjunto de datos	296 instancias, enfoque empresarial		
Procedimiento			
• Realizar cinco ejecuciones del algoritmo, con diferentes valores de semilla elegidos heurísticamente. • Obtener el promedio y mostrar el árbol obtenido			
Excepciones			
Eliminación de las variables demográficas			
Resultados	En las cinco ejecuciones se obtuvo el mismo árbol (véase figura 4.18) En la tabla 4.3 se muestra la matriz de confusión y en la tabla 4.4 se muestran los valores de desempeño del algoritmo		

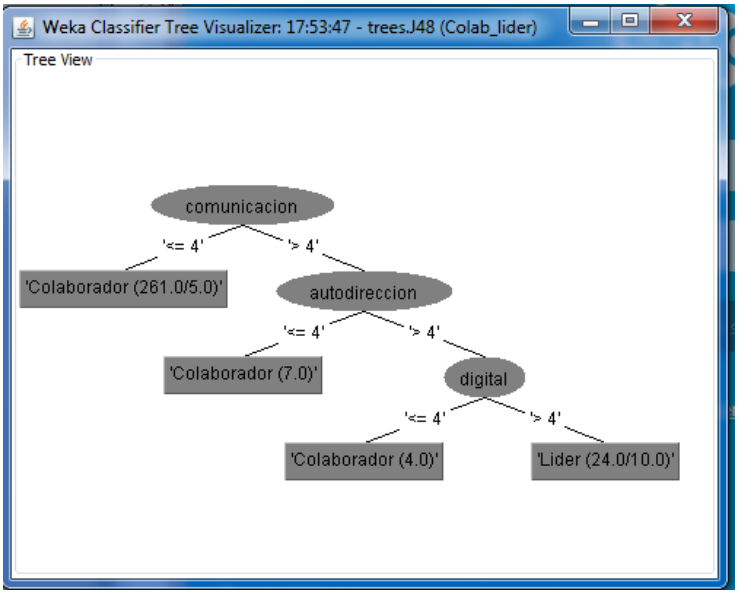


Figura 4.18 Resultados del evaluador árboles de decisión

En la figura 4.18 se muestra el árbol obtenido tras la ejecución del algoritmo árboles de decisión. Es importante mencionar que en las cinco ejecuciones realizadas, se obtuvo el mismo árbol. El cual indica que la característica más importante es la competencia "comunicación" y que a partir de ella se derivan valores menores a 4 (parcialmente de acuerdo) que definen a un colaborador y competencias "autodirección, digital" mayores a 4 que definen a un líder. El análisis completo de este árbol se describe en la propuesta realizada posteriormente.

En la tabla 4.3 se muestra la matriz de confusión de cada una de las ejecuciones del algoritmo. Se observa que en la primera ejecución clasificó correctamente, de las 19 instancias identificadas como líderes, solamente a 6 y el resto fueron clasificados como colaboradores. En tanto, que de las 277 instancias identificadas como colaboradores, clasificó correctamente a 268 instancias como tales y el resto como líderes. En general, para las cinco ejecuciones, las clasificaciones correctas se encuentran en la diagonal, es decir, en las columnas marcadas con color verde.

	Líder	Colaborador	Instancias correctamente clasificadas
Líder	**6**	13	274
Colaborador	9	**268**	
Líder	**4**	15	271
Colaborador	10	**267**	
Líder	**3**	16	272
Colaborador	8	**269**	
Líder	**5**	14	273
Colaborador	9	**268**	
Líder	**3**	16	272
Colaborador	8	**269**	

Tabla 4.3. Matriz de confusión de las cinco ejecuciones del algoritmo árboles de decisión

La tabla 4.4 muestra el promedio de las cinco ejecuciones en las medidas de desempeño: exactitud, precisión, *recall* y medida F. Las cuales determinan el porcentaje de instancias correctamente clasificadas.

Clase	Exactitud	Precisión	Recall	Medida F
Líder	92.02704	0.3178	0.2212	0.2596
Colaborador		0.9478	0.9684	0.9578

Tabla 4.4 Desempeño obtenido por el algoritmo árboles de decisión al promediar las cinco ejecuciones

4.5. Propuesta para la mejora de procesos en la toma de decisiones

De acuerdo a Checa Hinojo & Portillo García (2014)no se tiene una vía o manera mágica para la mejora de un proceso empresarial, ni se puede emplear uno que funcione en un proceso determinado para todos los procesos de la empresa. "*Un proceso será óptimo si es*

eficiente en su ejecución y sobresaliente en su eficacia. Es decir, cumple con los objetivos planificados con un máximo aprovechamiento de los recursos empleados".

Basado en lo anterior, en esta sección se describen las propuestas para mejorar algunos procesos de la toma de decisiones; particularmente enfocados en innovar la inteligencia de negocios desde un punto de vista pragmático en el área de capital humano.

4.5.1 Proceso para la integración de equipos de trabajo

La formación de equipos de trabajo dentro de una empresa es un asunto crítico y relevante, pero lo es aún más para las pequeñas y medianas compañías que en esta modalidad de trabajo encuentran una gran ventaja competitiva. Trabajar en equipo puede potenciar exponencialmente los límites de productividad y eficiencia de cada individuo que colabora para un fin común (SoyEntrepreneur.com, 2014).

La consultora Hay Group (2013) afirma que los equipos efectivos avanzan con sus tareas mucho más rápido y permiten que la empresa resista tiempos difíciles de una forma eficiente. Propone algunos criterios para formar equipos de trabajo altamente eficaces, tales como: liderazgo efectivo, estructura y límites, definición de objetivos y roles, selección de integrantes, establecer una visión común y generar compromisos.

Dado que uno de los objetivos que se pretende cumplir en este trabajo de tesis es formar grupos de trabajo basados en sus habilidades y afinidades. Hay Group (2013), propone que para seleccionar a las personas con las aptitudes más adecuadas para conformar el equipo, se deben identificar primero cuáles son las competencias requeridas para desempeñar satisfactoriamente los roles que se definieron.

Considerar que cada persona tiene talentos y habilidades que, si son bien canalizados, potenciarán el trabajo del grupo. Y que al momento de integrar el equipo es preciso elegir, en la medida de lo posible, a los miembros que sean diferentes entre sí, pero que haya compatibilidad. De allí que en este estudio se han considerado a las afinidades personales

para encontrar dicha compatibilidad personal si fuera el caso que no existiese compatibilidad basada en sus competencias profesionales.

En este sentido, la propuesta que se hace para integrar equipos de trabajo de acuerdo a sus habilidades y afinidades personales aplicando el algoritmo *k-medias*; cuya representación gráfica está basada en BPMN[15] (por sus siglas en inglés *Business ProcessModel and Notation*) es la siguiente:

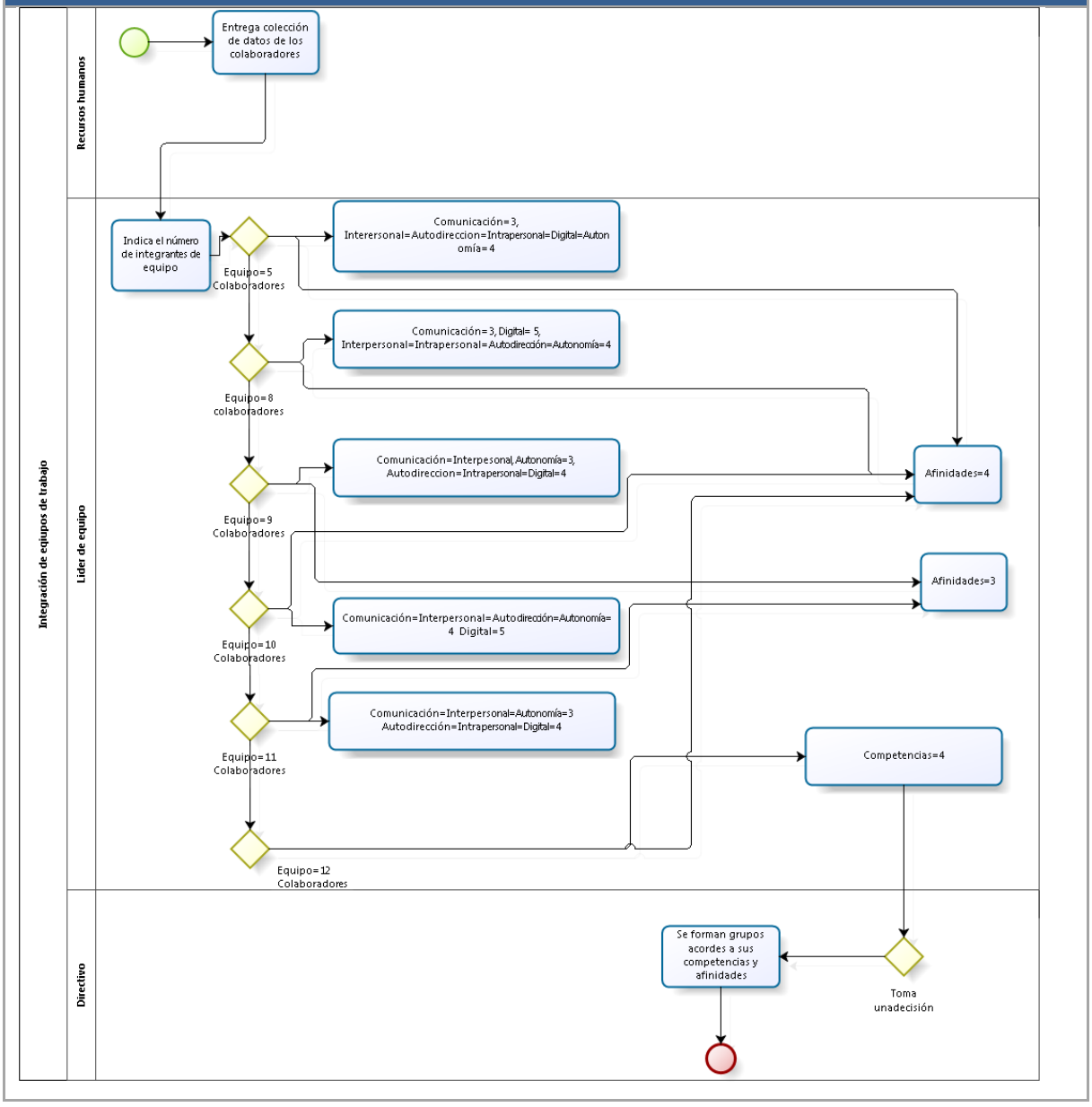
Descripción del proceso: Se busca establecer una secuencia de pasos para identificar las competencias y afinidades del capital humano a fin de integrarlos adecuadamente a un equipo de trabajo y lograr que éste sea eficiente; coadyuvando a crear inteligencia de negocios.

- El departamento de RH (Recursos Humanos) de cada empresa es el encargado de compartir información al resto de los departamentos que permita identificar verazmente sus competencias y afinidades; a través de instrumentos adecuados, como es el caso del H-A.
- Esta información, es compartida con el líder de equipo para identificar las competencias y afinidades para formar grupos eficaces (SoyEntrepreneur.com, 2014).

[15]Es una notación gráfica estandarizada que permite el modelado de procesos de negocio, en un formato de flujo de trabajo (workflow). El principal objetivo es proporcionar una notación estándar que sea fácilmente legible y entendible por parte de todos los involucrados e interesados del negocio (*stakeholders*). Entre estos interesados están los analistas de negocio (quienes definen y redefinen los procesos), los desarrolladores técnicos (responsables de aplicar los procesos) y los gerentes y administradores del negocio (quienes monitorizan y gestionan los procesos). En síntesis BPMN tiene la finalidad de servir como lenguaje común para cerrar la brecha de comunicación que frecuentemente se presenta entre el diseño de los procesos de negocio y su implementación. Tomado de (OMG , 2014)

Nombre del proceso	Integrar equipos de trabajo	Tipo de proceso	Proceso de apoyo horizontal

Diseño del proceso objetivo



- En función del número de integrantes a conformar el equipo, el líder busca en su proceso cuáles son las competencias y afinidades que debe tener cada uno.
  - Ejemplo:

- Supóngase que se desea integrar un equipo de 8 colaboradores para desarrollar un proyecto sobre desarrollo de software.
- Se busca en el proceso y se observa que, estos 8 colaboradores deben tener un 3 en comunicación, es decir, que su competencia no está claramente definida. Deben tener en las competencias: interpersonal, autodirección, intrapersonal y autonomía un valor de 4, es decir, que tienen una tendencia a ser eficaces en estas competencias.
- Si se desea asegurar que dicho equipo sea completamente eficaz, entonces es importante considerar sus afinidades personales, y en este rubro, todos los colaboradores deberán tener un 4, en otras palabras, deben poseer similares gustos personales y pasatiempos.

o Al tener las competencias definidas y si se desea, también las afinidades, entonces el directivo está listo para tomar una decisión de conformar grupos con estas características.

Para *clusters* menores a cinco personas, iguales a seis y siete o mayores a 13 personas; las competencias y afinidades fueron totalmente diferentes; por lo que existen muy diversas formar para formar grupos en función de sus competencias y afinidades personales.

4.5.2 Proceso para la selección de líderes

De acuerdo a George, Sims, McLean, & Mayer (2011) los líderes auténticos saben que la clave para lograr una organización exitosa es contar con líderes facultados en todos los niveles, incluyendo aquellos con los que no se relacionan directamente. Es por ello, que en esta sección se plantea un proceso para la selección de líderes basada en sus competencias y afinidades.

En el estudio realizado por Goleman (2011) muestra que el intelecto es un factor clave en el desempeño sobresaliente, que las habildiades cognitivas son especialmente importantes pero que la que inteligencia emocional desempeña un papel cada vez más trascendental en los niveles superiores de la organización.

Como se ha mencionado anteriormente, el instrumento H-A engloba habilidades cognitivas como autonomía en el proceso de aprendizaje e inteligencia emocional como habilidades intrapersonales e interpersonales, dada la importancia que conlleva elegir a un líder con estas características. A continuación se presenta la propuesta para seleccionar a líderes con habilidades cognitivas e inteligencia emocional empleando el algoritmo selección de atributos.

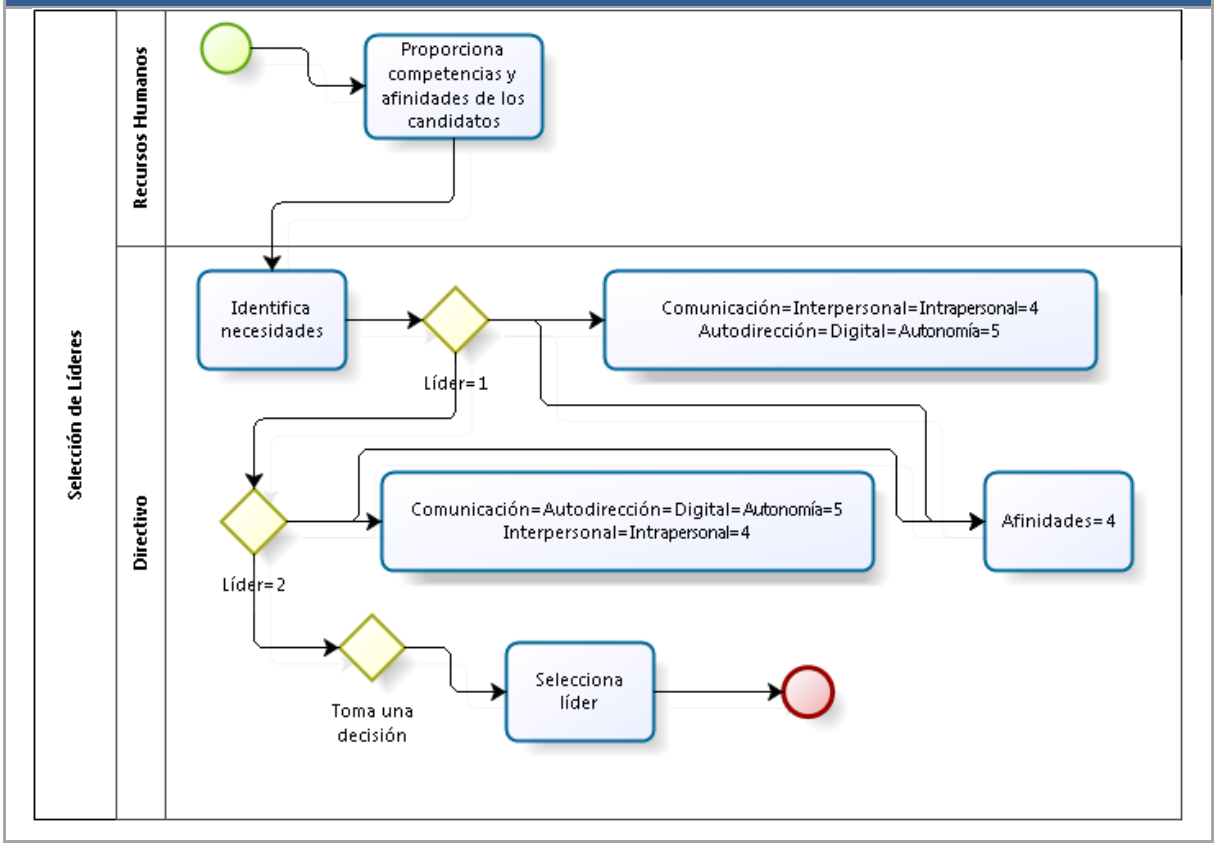
Descripción del proceso

- El departamento de RH es el encargado de proporcionar información (competencias y afindides) que ayude a la selección de líderes
- El directivo identifica las necesidades en cuanto al número de líderes que se desean elegir
    - Si solamente es un líder entonces deberá considerar que en sus compentecias de comnunicación, interpersonal, intrapersonal tenga una tendencia a la excelencia y que en la competencia de autdirección, digital y autonomía deberá tener calificación de excelente.
    - Si son dos líderes que se deben elegir entonces es necesrio considera que en sus habilidades de comunicación, autodirección, digital y autonomía sean excelentes y que en sus habildiades interpersonal e intrapersonal tengan tendecia a la excelencia.
    - Para ambos casos, es importante considerar sus afinidades en cuanto a los colaboradores, pues como se afirma en Leadership Business Group (2014) y Goleman (2011) para lograr un equipo eficaz es importante la convivencia

en grupo. En este sentido, el líder debiera tener una afinidad=4 con sus futuros colaboradores.

Nombre del proceso	Selección de líderes	Tipo de proceso	Proceso de apoyo mixto

Diseño del proceso objetivo



Debido a que el conjunto de datos para realizar la experimentación con líderes fue pequeño, tras la aplicación del algoritmo se identifican que existen múltiples formas de caracterizar a más de dos líderes.

4.7.3 Caracterización y clasificación del capital humano

En la figura 4.19 se muestran las características que deben tener un líder y un colaborador en función de sus competencias y afinidades personales; esto a partir de los resultados obtenidos tras la aplicación del algoritmo selección de atributos. Esta tabla puede ser considerada para la selección, también, de otro tipo de puestos dentro de una empresa o dentro de una institución educativa.

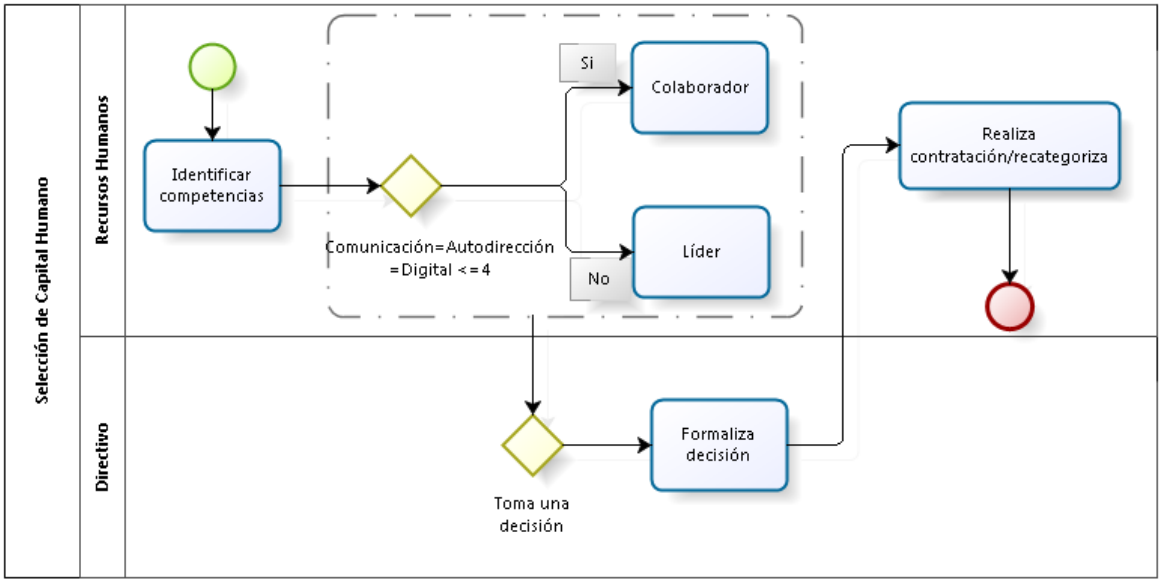
CARACTERIZACIÓN DEL CAPITAL HUMANO	
Líder	**Colaborador**
1. Habilidad intrapersonal 2. Autodirección 3. Habilidad digital 4. Autonomía en el proceso de aprendizaje 5. Afinidades con sus colaboradores	1. Autodirección 2. Habilidad digital 3. Autonomía en el proceso de aprendizaje 4. Afinidades con sus compañeros
LISTA ORDENADA DE CARACTERÍSTICAS DEL CAPITAL HUMANO SEGÚN SU IMPORTANCIA	
1. Afinidades 2. Autodirección 3. Habilidad interpersonal 4. Habilidad intrapersonal 5. Autonomía en el proceso de aprendizaje 6. Habilidad digital 7. Habilidad comunicativa	

Figura 4.19 Caracterización del capital humano

A continuación se presenta el proceso para la caracterización del capital humano en función de sus competencias y afinidades personales, esto tras la aplicación del algoritmo árboles de decisión.

Nombre del proceso	Caracterización de capital humano	Tipo de proceso	Proceso de apoyo mixto

Diseño del proceso objetivo



Descripción del proceso

- El departamento de recursos humanos es el encargado de identificar competencias de los colaboradores a través del instrumento H-A
  - Identifica si la competencia comunicativa, autodirección y habilidad digital tiene un valor menor o igual a 4, si esto ocurre entonces se puede caracterizar como un colabrador
  - Si dichas habilidades tienen un valor mayor que cuatro, es decir, 5. Entonces este individuo es un líder.
  - Posterior a la caracterización, el departamento envía la información a los directivos
- Los directivos o jefes de cada departamento son quienes deciden qué puesto dar, si de colaborador o líder
  - Se formaliza la decisión y se envía a recursos humanos
- El departamento de recursos humanos es el encargado de notificar el puesto o recategorización y finaliza el proceso.

En el capítulo siguiente se realiza el análisis de la propuesta presentada.

Referencias del capítulo

Checa Hinojo, E. J., & Portillo García, J. A. (2014). *Dirección de la actividad empresarial de pequeños negocios o microempresas.* Málaga, España: IC Editorial.

CNNexpansion. (2014). *Top 10 del Ranking 2014*. Recuperado el Septiembre de 2014, de 500 empresas más importantes de México 2014: http://www.cnnexpansion.com/rankings/interactivo-las-500/2014

Coordinación de Universidades Politécnicas. (2012). *Modelo Educativo*. Obtenido de Coordinación de Universidades Politecnicas: http://politecnicas.sep.gob.mx/

George, B., Sims, P., McLean, A. N., & Mayer, D. (2011). Descubra su auténtico liderazgo. *Harvard Business Review, 89*(11), 10-17.

Goleman, D. (2011). ¿Qué hace un líder? *Harvard Business Review, 89*(11), 42-50.

González Rufino, C. (2009). Tutoría universitaria y aprendizaje por competencias. ¿Cómo lograrlo? *Revista Electrónica Interuniversitaria de Formación del Profesorado, 12*(1), 181-204.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques.* Amsterdam: Morgan Kaufmann.

Hay Group. (2013). *Senior Leadership Teams*. Obtenido de "What it Takes to Make them Great": http://www.haygroup.com/ww/downloads/details.aspx?id=1192

Instituto Nacional de Evaluación Educativa. (2013). *PANORAMA DE LA EDUCACIÓN: Indicadores de la OCDE 2013, Informe Español.* Madrid, España: MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE.

Leadership Business Group. (2014). *Lo que hacen los líderes eficaces* . Obtenido de Grupo Empresarial: http://www.leadership-bg.com/index.php?option=com_content&id=345:o-que-fazem-os-lideres-eficazes&catid=71&Itemid=86&lang=es

Maldonado, Á., & Giandini, V. (2010). La conformación de un Sistema Tutorial para el Curso de Nivelación a Disancia de la Facultad de Ingeniería de la UNLP: decisiones, dificultades e impacto. *TE&ET | Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*(4), 47-53.

OCDE. (2014). *Resultados TALIS 2013 (Teaching And Learning International Survey)*. Obtenido de OCDE Education: http://www.oei.es/noticias/spip.php?article14148&debut_5ultimasOEI=15

OMG . (2014). *Business Process Model and Notation*. Obtenido de Home: http://www.bpmn.org/

SoyEntrepreneur.com. (2014). *5 consejos para formar equipos efectivos*. Obtenido de Recursos Humanos: http://www.soyentrepreneur.com/26530-5-consejos-para-formar-equipos-efectivos.html

The University of Waikato. (2013). *Weka 3: Data Mining Software in Java*. Obtenido de Machine Learning Group at the University of Waikato: http://www.cs.waikato.ac.nz/ml/weka/

Universidad Politécnica de Puebla. (2013). *Oferta Educativa*. Obtenido de Universidad Politécnica de Puebla: http://serpaguppue.uppuebla.edu.mx/

Witten, I., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2a. ed.). New Zeland: Elsevier.

Capítulo V

Análisis de resultados

INTRODUCCIÓN

En este capítulo se realiza el análisis de los resultados obtenidos y mostrados en el capítulo anterior. En función de la propuesta presentada que dan cumplimiento a cada uno de los objetivos específicos planteados al inicio de este trabajo de investigación; que llevan consigo el cumplimiento del objetivo general.

5.1. Análisis de los resultados

Como se describió en capítulos anteriores, en este trabajo de investigación, se considera a la inteligencia de negocios desde un punto de vista pragmático; es decir, asociándolo directamente con las tecnologías de la información y considerándolo como el conjunto de metodologías, aplicaciones, prácticas, capacidades y tecnologías que permiten reunir, depurar y transformar datos en sistemas transaccionales e información desestructurada o estructurada para su análisis directo, enfocadas en la administración o creación de información que ayude a los usuarios de una organización tomar mejores decisiones (Curto Díaz, 2012).

En este sentido, se planteó el desarrollo de métodos aplicando algoritmos de aprendizaje automático para que a través de algunos procesos empresariales se mejore (optimice) la toma de decisiones directivas y con ello crear una ventaja competitiva en la inteligencia de negocios. Entiéndase como optimización de procesos "*Un proceso será óptimo si es eficiente en su ejecución y sobresaliente en su eficacia. Es decir, cumple con los objetivos planificados con un máximo aprovechamiento de los recursos empleados"* (Stentoft Arlbjørn, 2010) y (Kress, 2010).

En este apartado se analizan detalladamente los resultados obtenidos tras la aplicación del algoritmo *k-medias*, selección de atributos y árboles de decisión.

Debido a que no existe una forma para determinar el número de personas que debe integrar un grupo de trabajo; la experimentación con el algoritmo *k-medias* se realizó desde dos aristas: por un lado, considerando el enfoque didáctico-pedagógico y por el otro el lado, el enfoque empresarial.

Para el primer enfoque, el conjunto de datos se basó en información obtenida de profesores y estudiantes de la Universidad Politécnica de Puebla; mientras que para el segundo enfoque, se consideraron a las primeras diez empresas más importantes del País de acuerdo a lo publicado por CNNexpansion en enero de 2014.

En la experimentación, para ambos enfoques se usó la distancia Euclídea y Manhattan con el fin de determinar el mayor número de individuos coincidentes con el centroide. La propuesta derivada de esta experimentación radica en un proceso para integrar individuos a equipos de trabajo "similares" a fin de lograr con esto que el equipo sea eficiente y por consiguiente, lograr mayor productividad en la empresa; en este sentido, todas las competencias fueron importantes, así como las afinidades personales.

Mientras que los resultados para determinar la características de un líder, sobresalen las competencias: interpersonal, autodirección, digital y autonomía; destituyendo a las afinidades personales. Lo que deja entrever que un buen liderazgo no está en función de las afinidades personales que pudiese tener con su equipo de trabajo; que a diferencia de los colaboradores si es un factor importante que determina un buen ambiente de trabajo y por lo tanto, mayor productividad.

Con los algoritmos de selección de atributos, se aplicaron cinco evaluadores y dos métodos de búsqueda. Tras la experimentación se pudieron reconocer las competencias más importantes que deben clasificar al capital humano en líder o colaborador, según sus competencias y afinidades. Éstos algoritmos, en contraposición con el algoritmo árboles de decisión, sí consideró importante las afinidades personales entre cada individuo listándola como principal, seguida de las competencias: autodirección, interpersonal, intrapersonal,

autonomía, digital y comunicación; de acuerdo a su orden de importancia. Esta lista es importante, puesto que admite asignar actividades acordes a sus competencias y si se desea aumentar la productividad, entonces también será necesario considerar las afinidades personales que éste tenga con el grupo de trabajo.

El algoritmo árboles de decisión, permitió caracterizar al capital humano, particularmente en líder o colaborador, en función de sus competencias y afinidades. No obstante, el árbol generado solamente consideró que un líder debe tener puntajes superiores a 4 en las competencias: comunicación (la más importante), autodirección y habilidad digital; mientras que las características para un colaborador es que su puntaje sea menor a 4 en las mismas competencias que para el líder. En contraposición con la lista generada por el algoritmo selección de atributos; para caracterizar al capital humano no es importante la afinidad que éstos pudiesen tener entre sí.

La experimentación con los algoritmos propuestos, contribuye significativamente a establecer procesos que impactan en el nivel estratégico, operativo y directivo de las empresas; al proponer una forma diferente de seleccionar al capital humano.

En general, los tres procesos propuestos por medio de los métodos automáticos generados a través de los algoritmos aplicados, proveen una forma innovadora para asignar actividades acordes a las competencias del individuo, para desarrollar perfiles de puestos, para obtener mejores resultados al asignar un líder acorde al grupo de trabajo y retos a enfrentar. Finamente, contribuyen a la toma de decisiones sin sesgo, de forma rápida y sencilla; pues cada clúster o árbol puede ser adaptado a los requerimientos que el puesto, actividad, proyecto o empresa demanden.

5.2 Contribuciones originales del trabajo de investigación

Como se describió en el capítulo II, el aprendizaje automático ha sido poco estudiado y aplicado en el contexto educativo y de negocios; es por ello que las contribuciones originales de este trabajo son destacadas al ofrecer como herramienta tecnológica su aplicación para resolver problemas derivados de los procesos en la toma de decisiones. Estas contribuciones son:

- Aplicar la tecnología de aprendizaje automático para mejorar los procesos en la toma de decisiones
- Aplicar diferentes formas para caracterizar, seleccionar y clasificar el capital humano
    - El algoritmo k-medias ayuda a formar grupos de individuos, para este estudio, generando grupos de trabajo homogéneos en cuanto a sus competencias (comunicación, digital, intrapersonal, interpersonal, autonomía) y afinidades personales.
    - El algoritmo selección de atributos ayuda a enlistar aquellas características más importantes de un conjunto de datos; para este caso, aquellas competencias que debiera tener un líder o un colaborador.
    - El algoritmo árboles de decisión, favorece la clasificación del capital humano al colocar en la cima del árbol aquella característica más importante, para este caso, la competencia de comunicación.
- Modelar procesos a través de los resultados obtenidos de la experimentación con los algoritmos: Proceso para la integración de equipos de trabajo, proceso para la selección de líderes y proceso para la caracterización y clasificación del capital humano.

Con todo esto, la principal contribución de esta investigación es ayudar a las empresas u organizaciones a resolver los retos actuales que la sociedad demanda; pues al aplicar la tecnología de aprendizaje automático en la mejora de procesos de la toma de decisiones se crean propuestas valiosas como:

- Acceso y análisis de datos de manera rápida para tomar decisiones de negocio a nivel operativo, directivo y estratégico
- Mejor toma de decisiones, sin sesgo, basadas en información exacta
- Una solución diferente en la integración y modelación de los datos, de modo que sea confiable y exacta
- Mejora de los procesos relacionados al capital humano
- Productividad en el procesamiento de información
- Reducción de procesos manuales

5.3 Impacto del trabajo de investigación

El impacto de este trabajo de investigación considera varias aristas, que Aleixandre Mendizábal, Gómez González, & Moñux Chércoles (2003) sugieren debe contemplar un proyecto, a saber: sistema de innovación, económico, social, empleo y ambiental.

Sistema de innovación

El impacto en este contexto está relacionado con los recursos del negocio, es decir, con el capital humano al poder caracterizarlo y asignarle actividades de acuerdo a sus competencias y afinidades con el grupo de trabajo; siendo esta una forma novedosa para la selección de personal al potencializar la productividad del personal y por ende, de la empresa.

Al emplear tecnología de aprendizaje automático, se crea una cultura tecnológica al mostrar una forma diferente de reclutar al capital humano, asignarle actividades coherentes con sus competencias y afinidades.

Al tomar decisiones sin sesgo por medio de la tecnología, se genera una nueva alternativa de uso de la misma y se le da una dimensión valorativa que conlleva evitar los cambios traumáticos al sustituir las viejas costumbres por la tecnología de aprendizaje automático.

Crear una cultura empresarial diferente, en donde las competencias y afinidades del capital humano son base importante para la productividad del negocio.

Finalmente, se genera una cultura tecnológica en la sociedad, a través de una dimensión cognitiva al promover una cultura científica y tecnológica mediante la aplicación de tecnologías, como es el caso de la tecnología de aprendizaje automático, poco estudiadas o aplicadas en contextos de negocios y educación.

Impacto económico

Particularmente, el impacto económico de un proyecto está relacionado con la potenciación de las PyMES al ser competitiva basada en su capital humano y por ende, en su producción. En este sentido, este proyecto de investigación al ser implementado en la industria puede generar mejores fuentes de empleo, capital humano capacitado, competitivo y auto-motivado al contratarse bajo sus habilidades y afinidades en función del puesto a desarrollar.

Además, al lograr mayor productividad a través de su capital humano capacitado, la empresa genera mayores ganancias, mayor competitividad en el mercado y por ende, mayor oferta de empleo.

Impacto social

La propuesta presentada tiene un socio-diseño al incluir, como parte importante en los grupos de trabajo, las afinidades personales; que contribuyen sensiblemente a fomentar las relaciones sociales de manera fácil y rápida.

Al aplicar la TAA en la empresa, se considera a una empresa inclusiva; pues los puestos de trabajo serían acordes a las competencias que cada individuo posee.

Si una empresa agrupa a personas en función de sus competencias y afinidades personales con el grupo de trabajo, se forman personas felices y por ende, se logra una sociedad feliz;

quizá esto reduzca varios problemas sociales como la delincuencia y violencia.

Empleo

El impacto en este contexto está relacionado con los aspectos cualitativos (transformación), es decir con el efecto cualitativo que conlleva la aplicación de la tecnología de aprendizaje automático (TAA) en el negocio o IES, pues a partir de esta investigación se pueden generar mejores perfiles de puesto al integrar al capital humano en un puesto o actividad acorde a sus competencias y afinidades personales con el grupo de trabajo, derivando con esto, quizá, el riesgo laboral de ausentismos, rotación frecuente de personal, insatisfacción laboral, entre otros.

A las empresas o IES, el impacto de aplicar la TAA conllevaría a tener un capital humano competitivo y productivo y por consiguiente, elevar la productividad, ganancias y reconocimiento en otros estratos sociales, económicos y empresariales.

Impacto ambiental

El impacto ambiental de este trabajo de investigación está relacionado con el eco-diseño del proceso al proporcionar de forma automática información al usuario de la organización y propiamente, al consumidor del proceso o producto. Al mismo tiempo, el impacto ambiental se ve reflejado en el menor consumo de materiales y ahorro de energía al emplear tecnologías de aprendizaje automático cuyo procesamiento se realiza en segundos a diferencia de las tecnologías comúnmente empleadas (ERPs) para procesar y analizar información.

Referencias del capítulo

Aleixandre Mendizábal, G., Gómez González, F. J., & Moñux Chércoles, D. (2003). Desarrollo de una Guía de Evaluación de Impacto Social para Proyectos de I+D+I. *Revista Iberoamericana de Ciencia, Tecnología, Sociedad e Innovación, Enero-Abril*(6), http://www.oei.es/revistactsi/numero5/articulo4.htm.

Curto Díaz, J. (2012). *Introducción al Business Intelligence.* Barcelona, España: UOC.

Kress, M. (2010). *Intelligent Business Process Optimization for the Service Industry.* Demand: KIT Scientific Publishing.

Stentoft Arlbjørn, J. (2010). *Business Process Optimization.* Denmark: Academica.

Checa Hinojo, E. J., & Portillo García, J. A. (2014). *Dirección de la actividad empresarial de pequeños negocios o microempresas.* Málaga, España: IC Editorial.

Conclusiones y Trabajo Futuro

INTRODUCCIÓN

En este apartado se detallan las conclusiones obtenidas tras la realización del trabajo de tesis, así mismo, se puntualiza el cumplimiento de cada uno de los objetivos y por consiguiente del objetivo general. Finalmente, se expone el trabajo a futuro que se pretende realizar para continuar con la investigación a fin de proporcionar herramientas diferentes que mejoren asertivamente la toma de decisiones de los directivos.

Conclusiones

Recordando que la problemática que dio origen a este trabajo de tesis, se fundamentaba en la pregunta ¿Cómo procesar grandes de información y analizarlos de manera objetiva en el menor tiempo? A través de la experimentación realizada, se ha encontrado una solución y respuesta al planteamiento. Una solución es emplear técnicas de aprendizaje automático para la extracción, depuración, procesamiento y análisis de la información de manera sencilla, rápida y confiable; que conlleva a la reducción de tiempo al momento de analizarlos y transformarlos en información entendible para los usuarios y directivos de las de las organizaciones.

En este sentido, se responde a las preguntas que los directivos se plantean:

¿Cómo encontrar, analizar y evaluar más posibilidades en menos tiempo? Empleando tecnología de aprendizaje automático que genera más alternativas e información procesada

acorde a los requerimientos de la organización; promoviendo con ella una toma de decisiones objetiva, rápida y confiable basada en tecnología.

¿Cómo hacer elecciones más asertivas en los negocios? Empleando algoritmos de aprendizaje automático puesto que dan soporte a la toma de decisiones y que favorece el desarrollo de la inteligencia de negocios.

Es así como la tecnología de aprendizaje automático favorece la generación de esta inteligencia de negocios al explorar grandes bases de datos, de manera automática o semiautomática, a que actúe como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva; en este contexto, al aumentar su productividad a través de la agrupación del capital humano en equipos homogéneos basados en sus competencias y afinidades personales, al fortalecer el liderazgo mediante la selección del líder adecuado para un grupo de trabajo y al caracterizar a sus líderes identificando aquellas competencias que los hacen diferentes de un colaborador.

A continuación se detalla el cumplimiento de cada uno de los objetivos específicos.

O.E. No. 1: Extraer información automáticamente para optimizar la toma de decisiones a través de aprendizaje automático logrando asertividad en el proceso

Para cumplir con este objetivo se realizó la extracción de la información a través de los algoritmos k-medias, selección de atributos y árboles de decisión. El primero para formar grupos de individuos basados en sus competencias y afinidades, el segundo para identificar las competencias sobresalientes tanto de líderes como de colaboradores y finalmente, el tercero para obtener aquellas características que debe cumplir un líder y un colaborador. La asertividad en el proceso de la toma de decisiones se logra a través de la aplicación de la tecnología de aprendizaje automático, mediante el empleo de estos tres algoritmos al extraer y procesar la información de manera confiable, sencilla, rápida y objetiva.

O.E. No. 2: Integrar equipos de trabajo de acuerdo a sus habilidades y afinidades personales aplicando el algoritmo k-medias

Este objetivo se cumplió al experimentar con diferentes conjuntos de datos y obtener las competencias que caracterizan a cada grupo de trabajo con un número mayor a 2 integrantes para colaboradores y menor a 2 para líderes.

O.E. No. 3 Identificar a grupos de líderes a través de sus competencias y afinidades personales aplicando el algoritmo *k-medias* para la mejora del desempeño de los directivos y del personal clave en las empresas

Por medio del algoritmo *k-medias* se lograron identificar las competencias que debe tener un líder, ayudando con esto a mejorar tanto el desempeño de los líderes como de sus colaboradores.

O.E. No. 4 Identificar automáticamente las características más relevantes del capital humano para asignar con efectividad actividades acordes a su perfil usando selección automática de atributos

El algoritmo selección de atributos identificó que las competencias más relevantes del capital humano deben ser: Afinidades, Autodirección, Interpersonal, Intrapersonal, Autonomía, Digital, Comunicación; es decir, las afinidades son las más relevantes y se cree que esto fomente una mayor colaboración entre los integrantes del grupo de trabajo y por ende, mayor productividad. Además, agruparlos bajo estas características ayudaría al departamento de recursos humanos y a los líderes a asignarles actividades acordes a cada perfil.

O.E. No. 5 Clasificar a individuos como líderes o colaboradores en relación a sus competencias y afinidades usando el algoritmo árboles de decisión

El algoritmo árboles de decisión ayudó a obtener las mejores características que deben tener un líder y un colaborador. A saber, un líder se caracteriza por tener una excelente comunicación, una autodirección desarrollada y una habilidad digital sobresaliente. Mientras que un colaborador, debe tener estas mismas competencias, solo que a un nivel medio a bajo, es decir, una comunicación, una autodirección y una habilidad digital con tendencia media a baja.

O.E. No. 6 Modelar procesos estratégicos para optimizar la toma de decisiones en la clasificación y selección del capital humano

La aplicación de los algoritmos *k-medias*, selección de atributos y árboles de decisión, permitieron la selección y caracterización del capital humano a través de tres métodos automáticos que favorecen la optimización de la toma de decisiones, mediante el modelado de tres procesos estratégicos llamados: Creación de grupos de trabajo, selección de líderes y caracterización del capital humano. Particularmente este objetivo, conlleva al cumplimiento del objetivo general de investigación.

En definitiva, la aplicación de la tecnología de aprendizaje automático logra generar inteligencia de negocios al aplicar herramientas tecnológicas que contribuyan a analizar, procesar y depurar información de manera rápida, sencilla y confiable. En este sentido, la aplicación de los algoritmos k-medias, selección de atributos y árboles de decisión, permiten tomar decisiones sin sesgo; pues ofrecen de manera gráfica la información procesada que facilita el proceso, particularmente relacionado con el capital humano.

En general, los tres procesos propuestos por medio de los métodos automáticos generados a través de los algoritmos aplicados, proveen una forma innovadora para asignar actividades acordes a las competencias del individuo, para desarrollar perfiles de puestos, para obtener

mejores resultados al asignar un líder acorde al grupo de trabajo y retos a enfrentar. Finamente, contribuyen a la toma de decisiones sin sesgo, de forma rápida y sencilla; pues cada clúster o árbol puede ser adaptado a los requerimientos que el puesto, actividad, proyecto o empresa demanden.

Trabajo a futuro

Existen muchas maneras de enriquecer este trabajo de investigación, a continuación se mencionar algunas actividades futuras que contribuirían a generar mejores beneficios de la aplicación de la tecnología de aprendizaje automático en el contexto de negocio y educativo.

- Desarrollar instrumentos de evaluación basados en las características de los puestos genéricos de una empresa con el fin de brindar mejores métodos para la selección y caracterización del capital humano
- Ampliar el conjunto de datos sobre líderes para ofrecer otras alternativas de su selección y coherentes con las actividades que se desempeñarán
- Ampliar las competencias a evaluar, de tal forma, que se incluyan todas aquellas necesarias para desempeñar un puesto en particular, ya sea como líder o colaborador
- Aplicar cada uno de los procesos propuestos para estimar su efectividad tanto en la industria como en las instituciones de educación superior
- Aplicar métodos de sobre o sub-muestreo de datos para balancear los conjuntos de datos
- Desarrollar una herramienta de software que implemente los métodos propuestos como apoyo a los directivos que facilite la toma de decisiones

Anexos

INTRODUCCIÓN

En esta sección se encuentra toda la información que complementa este trabajo de investigación, así como la producción científica derivada de ello.

ANEXO 1. CÁLCULO DEL ALFA DE CRONBACH

Para obtener la validación del instrumento denominado “Perfiles” se realizaron los siguientes pasos.

Paso 1: Mediante la ayuda de una hoja de cálculo, ubicar en las columnas el número de preguntas y en las filas el número de participantes.

Paso 2: Capturar las respuesta de cada participante en las respectivas columnas.

Paso 3: Calcular la sumatoria de cada una de las filas, es decir, de cada participante.

Paso 4: obtener las varianza de cada una de las sumatorias obtenidas en el paso 3.

VALIDACIÓN DEL INSTRUMENTO "PERFILES" POR MEDIO DEL ALFA DE CRONBACH

	ítems																														Suma	Varian
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	S	V
E1	4	3	1	2	2	2	3	2	2	2	2	2	2	2	1	1	1	1	2	2	2	2	3	2	3	5	5	2	3	4	4	5	4	2	85	1.348
E2	3	2	3	1	2	3	3	3	1	1	1	2	1	1	2	2	2	3	2	2	2	1	3	1	1	2	3	1	3	2	2	1	1	2	65	0.628
E3	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	50	0.257
E4	3	1	2	2	2	3	3	2	2	2	2	2	2	2	2	3	3	2	2	2	2	2	2	2	1	1	1	1	2	1	1	3	1	1	65	0.447
E5	2	2	1	2	1	3	2	2	3	2	3	1	2	1	1	1	2	2	2	1	2	3	2	2	4	4	4	2	3	4	1	2	5	3	77	1.11
E6	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	4	4	2	4	2	2	2	4	76	0.428
E7	2	1	2	1	1	1	2	2	2	1	1	1	2	2	1	1	1	2	2	2	1	1	2	2	2	1	1	3	3	3	2	1	3	2	57	0.468
E8	2	2	3	1	2	1	2	2	2	2	2	1	2	3	2	2	1	1	2	2	1	1	2	2	1	1	2	2	2	1	1	3	1	2	59	0.382
E9	1	1	2	2	3	3	1	3	1	1	1	1	2	2	1	1	1	1	4	1	3	1	3	2	1	3	2	1	1	3	4	1	3	3	64	1.016
E10	2	3	2	2	2	2	2	2	2	1	2	2	2	1	1	1	1	2	1	1	1	2	2	1	2	2	4	2	4	2	1	4	2	4	67	0.817
E11	2	2	2	2	1	3	2	2	3	2	2	3	2	3	2	1	1	2	3	4	3	2	2	3	2	3	4	2	3	2	2	2	4	4	82	0.674
E12	2	2	3	2	3	3	3	3	3	2	2	2	2	2	2	2	2	2	1	2	2	2	2	3	1	3	3	1	2	3	3	4	2	1	77	0.504
E13	2	2	3	2	2	2	2	1	1	2	2	2	4	3	5	1	1	2	3	2	2	2	3	2	2	2	3	4	3	3	1	2	4	1	78	0.941
A1	2	1	1	1	1	1	1	1	1	1	1	2	2	2	3	3	3	2	2	1	1	1	2	2	2	5	5	1	5	5	5	5	5	5	81	2.546
A2	2	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	1	1	2	5	1	4	4	4	4	4	58	1.487
A3	3	1	3	1	2	3	2	1	1	1	1	2	2	2	1	1	3	1	2	2	1	1	1	3	1	1	4	4	3	2	2	2	2	2	64	0.834
	ítems			Estudiantes				Asesores				Varianza de ítems									Varianza total													109.3	13.89

Paso 5: Calcular el Alfa de *Cronbach* de acuerdo a la fórmula $\alpha = \frac{K}{K-1}\left[1 - \frac{\sum V_i}{V_t}\right]$

Paso 6: Considerar que el Alfa de Cronbach debe estar entre 0 y 1 para que un instrumento sea válido. Para el caso particular del instrumento "perfiles" se obtuvo un valor de

Alfa de Cronbach	0.899355801

indicando que dicho instrumento tiene la validez suficiente para evaluar cada una de las variables en cuestión.

Es importante comentar que este procedimiento debe aplicarse a cada una de las pruebas piloto que se realizan al administrar el instrumento.

ANEXO 2. INSTRUMENTO PROPUESTO

FOLIO: _____

CUESTIONARIO PARA IDENTIFICAR AFINIDADES Y HABILIDADES

Introducción: La Universidad Autónoma de Chiapas y la Universidad Politécnica de Puebla trabajan en un estudio que servirá para identificar el perfil de los estudiantes con el objetivo de asignarles un tutor acorde a las características del grupo. Solicitamos de su ayuda para contestar este cuestionario.

Indicaciones: Le pedimos que conteste el cuestionario con la mayor sinceridad posible. No hay respuestas correctas ni incorrectas. Todas las preguntas tienen cinco opciones de respuesta, elije solamente la que mejor describa lo que consideres apropiado. Por último, le pedimos evitar dejar preguntas sin contestar.

Información demográfica

Edad:_____ Sexo: M F Estado civil: _____ Trabaja: SI NO

CUESTIONARIO

5: Totalmente de acuerdo, **4:** Parcialmente de acuerdo, **3:** Ni en acuerdo ni en desacuerdo, **2:** Parcialmente en desacuerdo, **1:** Totalmente en desacuerdo

Ítem	5	4	3	2	1
1. Tengo buena construcción gramatical para redactar reportes y ensayos					
2. Tengo habilidad de presentación, discusión y argumentación					
3. Represento fácilmente mis ideas con mnemotecnia (diagramas, esquemas, presentaciones, etc.)					
4. Explico e interpreto fácilmente la realidad					
5. Comparto mis ideas con otros de manera sencilla					
6. Produzco ideas originales que permiten crear e innovar					
7. Aplico fácilmente conceptos, valores y herramientas en la realidad natural o social					
8. Propongo alternativas para solucionar problemas y selecciono fácilmente las opciones viables					

9. Enfrento problemas y los supero con facilidad					
10. Considero que tengo autonomía intelectual y moral					
11. Realizo actos con responsabilidad ética, social y ambiental					
12. Demuestro de manera oral, escrita o física las cualidades propias					
13. Defino necesidades de aprendizaje y busco satisfacerlas con el máximo provecho y mínimo esfuerzo					
14. Suelo realizar planes para alcanzar metas realistas					
15. Utilizo recursos tecnológicos e informáticos para facilitar mi aprendizaje o actividad laboral					
16. Aprovecho recursos digitales de forma responsable, pertinente y ágil en mi actividad académica					
17. Encuentro información de interés de forma rápida y sencilla					
18. Organizo mi tiempo adecuadamente para realizar actividades académicas o personales					
19. Tengo la capacidad de auto motivarme a pesar de la dificultad de ciertas actividades					
20. Tengo responsabilidad y compromiso para realizar cada una de las actividades planeadas					
21. Tengo una mente abierta para las diferentes opiniones que se generan en el grupo					
22. Realizo una autoevaluación constante de mis progresos					
23. Aplico constantemente estrategias para evitar mis deficiencias					
24. Disfruto de ir al cine/teatro/museo					
25. Practico o veo algún deporte					
26. Asisto frecuentemente a espectáculos					
27. Me distraigo realizando alguna actividad dentro de casa					
28. Disfruto de participar frecuentemente en redes sociales					
29. Frecuentemente, opto por cenar fuera de casa					
30. Disfruto de los paseos al aire libre					
31. Me gusta ir al spa/gimnasio					
32. Visito frecuentemente a un familiar o amigo					

Gracias por su tiempo y participación.

ANEXO 3 PRODUCCIÓN CIENTÍFICA

1. Urbina Nájera, A. B., de la Calleja, J., Vega Lebrún, C. A., López Maldonado, N., & Pico González, B. (2014). Desarrollo y validación de un instrumento para identificar perfiles de tutorados y tutores de la modalidad virtual. *CAFVIR 2014* (págs. 227-234). Antigua Guatemala: ESVIAL.

Referencias

[1] Aldiss, B., Watson, I., & Spielberg, S. (2001). Artificial Intelligence A.I. Los Ángeles, California, Estados Unidos.

[2] Aleixandre Mendizábal, G., Gómez González, F. J., & Moñux Chércoles, D. (2003). Desarrollo de una Guía de Evaluación de Impacto Social para Proyectos de I+D+I. Revista Iberoamericana de Ciencia, Tecnología, Sociedad e Innovación, Enero-Abril(6), http://www.oei.es/revistactsi/numero5/articulo4.htm.

[3] Alhah, S., Abu Hammad, A., Samhouri, M., & Al-Ghandoor, A. (2011). Modeling stock market exchange prices using artificial neural networks: a study of amman stock exchange. Jordan Journal of Mechanical and Industrial Engineering, 5(5), 439:446.

[4] Álvarez González, M. (2008). La tutoría académica en el Espacio Europeo de la Educación Superior. Revista Interuniversitaria de Formación del Profesorado, 22(1), 71-88.

[5] Álvarez, J., & Zwir, I. (2001). Aprendizaje Automático (AA). Obtenido de brains and machines: http://www-2.dc.uba.ar/materias/aa/aa.html#Qu%C3%A9%20significa

[6] Alzate-Medina, G. M., & Peña-Barrero, L. B. (2010). La tutoría entre iguales: una modalidad para el desarrollo de la escritura en la Educación Superior. Universitas Phsycologica, 9(1), 123-138.

[7] Amaya Amaya, J. (2010). Toma de decisiones gerenciales: Métodos cuantitativos para la administración. Bogotá, Colombia: ECOE EDICIONES.

[8] ANUIES. (2011). Programas institucionales de tutoría una propuesta de la ANUIES 3a Edición. México, D.F.: ANUIES.

[9] ANUIES-Universidad Pedagógica Nacional. (2004). La innovación en la educación superior: Documento estratégico. México, D.F.: ANUIES.

[10] Arciniega, S., Del Rosario, M., Calderón, B., & M. L. (2006). Validez y confiabilidad del estudio socieconómico. México, DF.: UNAM.

[11] Arias Delgado, L. P. (2011). Módulo: Cerebro y Aprendizaje. Chile: Fundación universitaria del Área Andina.

[12] Asimov, I., Silverberg, R., & Kazan, N. (1999). Bicentennial Man. Los Ángeles, California, Estados Unidos.

[13] Barrón Tirado, M. C. (2009). Docencia universitaria y competencias didácticas. Perfiles educativos, XXXI(125), 76-87.

[14] Bautista, G., Borgues, F., & Forés, A. (2006). Didáctica universitaria en entornos virtuales de enseñanza-aprendizaje. Madrid, España: Narcea.

[15] Benson, P. (2007). Autonomy and its role in learning. Springer International Handbooks of Education, 15, 733-745.

[16] Bensoussan, B. E., & Fleisher, C. S. (2013). Analysis without paralysis: 12 tools to make better strategic decision. USA: Pearson Education.

[17] Berberena González, V. H. (2011). El Patrón de Lealtad de Clientes: una ventaja competitiva sostenible. Obtenido de Negociación Comercial: negociacioncomercial.com.mx/archivos/archivo_107.pdf

[18] Berberena González, V. H. (2011). El Patrón de Lealtad de Clientes: una ventaja competitiva sostenible. Obtenido de Negociación Comercial: negociacioncomercial.com.mx/archivos/archivo_107.pdf

[19] Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. USA: O´Really Media, Inc.

[20] Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. USA: O´Really Media, Inc.

[21] Bishop, C. M. (2007). Pattern recognition and Machine Learning. Singapore: Springer.

[22] Bustillo Porro, V. (2009). Nuevas Tecnologías de la información: Herramientas para la educación. Teoría de la Educación, 3(10), http://campus.usal.es/~teoriaeducacion/rev_numero_06/n6_art_bustillo.htm. Obtenido de Ediciones Universidad de Salamanca: http://campus.usal.es/~teoriaeducacion/rev_numero_06/n6_art_bustillo.htm

[23] Cabrera Dokú, K., & González F., L. E. (2006). Currículo universitario basado en competencias. Barranquilla, Colombia: Ediciones Uninorte.

[24] Cano González, R. (2009). Tutoría universitaria y aprendizaje por competencias ¿Cómo lograrlo? REIFOP: Revista Electrónica Interuniversitaria de Formación del Profesorado, 12(1), 181-204.

[25] Cervantes H., V. (2005). Interpretaciones del coeficiente Alpha de Cronbach. Avances en medición, 3, 9-28.

[26] Checa Hinojo, E. J., & Portillo García, J. A. (2014). *Dirección de la actividad empresarial de pequeños negocios o microempresas.* Málaga, España: IC Editorial.

[27] Cid Sabucedo, A., Pérez Abellás, A., & Sarmiento Campos, J. A. (2011). La tutoría en el Practicum. Revisión de la literatura. Revista de educación, 354(enero-abril), 127-154.

[28] CNNexpansion. (2014). Top 10 del Ranking 2014. Recuperado el Septiembre de 2014, de 500 empresas más importantes de México 2014: http://www.cnnexpansion.com/rankings/interactivo-las-500/2014

[29] Conforth, M., & Meng, Y. (2008). An Artificial Neural Network Based Learning Method for Mobile Robot Localization . Robotics, Automation and Control (pp. 494-504). Viena, Austria: I-Tech.

[30] Cooper, M. M., & Sandí-Ureña, S. (2009). Design and Validation of an Instrument to Assess Metacognitive Skilfulness in Chemistry Problem Solving. Journal of Chemical Education, 86(2).

[31] Coordinación de Universidades Politécnicas. (2012). Modelo Educativo. Obtenido de Coordinación de Universidades Politécnicas: http://politecnicas.sep.gob.mx/

[32] Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tesis. Psychometrika, 16(3), 297-335.

[33] Cruz Bejarán, R. M. (2010). Espacio Docente: Portal de recursos para el docente innovador. Obtenido de: El estudiante virtual: http://docentes.unibe.edu.do/noticias_x.aspx?idx=81

[34] Cruz, J. A., & Wishant, D. S. (2007). Applications of machine learning in cancer prediction and prognosis. Cancer Informatics, 2.

[35] Curto Díaz, J., & Conera i Carat, J. (2010). Introducción al Business Intelligence. Barcelona: Editorial UOC.

[36] David, F. R. (1997). Concepto de Administración Estratégica. México, D.F.: Pearson Education.

[37] De La Calleja, J., Benitez, A., Medina, M. A., & Fuentes, O. (2011). Machine learning from imbalanced data sets for astronomical object classification. SoCPaR (pp. 435-439). Dalian, China: IEEE Xplore Digital Library.

[38] De La Calleja, J., Huerta, G., & Fuentes, O. (2010). The imbalanced problem in morphological galaxy classification. CIARP (pp. 533-540). Sao Paulo, Brasil: Springer Verlag.

[39] De Quiroga, A. (2004). El proceso educativo según Paulo Freire y Enrique Pichon-Riviére. San Pablo, Brasil: Plaza y Váldes S.A. de C.V.

[40] Díaz Narváez, P. (2009). Metodología de la investigación y bioestadística. Santiago de Chile: RIL Editores.

[41] Dirección General de Desarrollo de la Gestión e Innovación Educativa-SEP. (2013 julio). Secretaría de Educación: Gobierno del Estado de Jalisco. Obtenido de: Modelo de Gestión Educativa Estratégica: Programa Escuelas de Calidad: http://portalsej.jalisco.gob.mx/sites/portalsej.jalisco.gob.mx.programa-escuelas-calidad/files/pdf/mgee.pdf

[42] Drucker, P. F., Hammond, J., Raiffa, H., & Argyris, C. (2001). On Decision Making. Boston: Harvard Bussines Review.

[43] Duvivier, R. J., Van Dalen, J., Van Der Vleuten, C. P., & Scherpbier, A. J. (2009). Teacher perceptions of desired qualities, competencies and strategies for clinical skills teachers. Medical Teacher, 31(7), 634-641.

[44] Engedy, I. (2009). Artificial neural network based mobile robot navigation. Intelligent Signal Processing, 2009.WISP 2009 (pp. 241-246). Budapest, Hungria: IEEE Xplore Digital Library.

[45] Er, O., Yumusak, N., & Temurtas, F. (2010). Chest disease diagnosis using ANNS. Expert systems with application, 37(12), 7648-7655.

[46] Ferrell, O. C., Hirt, G. A., & Ferrel, L. (2010). Introducción a los negocios en un mundo cambiante. Colorado, CA: McGraw Hill.

[47] Freitas, A. A. (2002). Data Mining and Knowledge Discovery with Evolutionary Algorithms. The Netherlands: Springer-Verlag.

[48] Fujita, H., & Revetria, R. (2012). New treds in software methodologies, tools and techniques. Proceedings of the Eleventh SoMeT_I2. Amsterdam: IOS Press.

[49] García Fernández, L. A. (2004). Usos y aplicaciones de la inteligencia artificial. La ciencia y el hombre, XVII(3), http://www.uv.mx/cienciahombre/revistae/vol17num3/articulos/inteligencia/.

[50] García Manjón, V. J., & Pérez López, M. C. (2008). Espacio Europeo de Educación Superior, competencias profesionales y empleabilidad. Revista Iberoamericana de Educación, 46(9), 1-12.

[51] García Morate, D. (2006). Weka en castellano. Obtenido de Diego García Morate: http://www.metaemotion.com/diego.garcia.morate/

[52] García Pérez, S. L. (2006). Importancia de la tutoría en la vida universitaria. Pacific Circle Consortium-Conference (págs. 1-15). México, D.F.: Latinamerican Faculty of Social Sciences, Mexican Campus.

[53] George, B., Sims, P., McLean, A. N., & Mayer, D. (2011). Descubra su auténtico liderazgo. Harvard Business Review, 89(11), 10-17.

[54] Goleman, D. (2011). ¿Qué hace un líder? Harvard Business Review, 89(11), 42-50.

[55] González Rufino, C. (2009). Tutoría universitaria y aprendizaje por competencias. ¿Cómo lograrlo? Revista Electrónica Interuniversitaria de Formación del Profesorado, 12(1), 181-204.

[56] Hamilton, H. (2009). Site Map of Course Notes. Obtenido de Computer Science 831: Knowledge Discovery in Databases: http://www2.cs.uregina.ca/~dbd/cs831/index.html

[57] Hamilton, L. (2014). Six Novel Machine Learning Applications. Obtenido de Forbes: http://www.forbes.com/sites/85broads/2014/01/06/six-novel-machine-learning-applications/

[58] Han, J., Kamber , M., & Pei, J. (2011). Data mining: concepts and techniques. Amsterdam: Morgan Kaufmann.

[59] Harvard business essentials. (2006). Toma de decisiones para conseguir mejores resultados. EE.UU.: Deusto.

[60] Hay Group. (2013). Senior Leadership Teams. Obtenido de "What it Takes to Make them Great": http://www.haygroup.com/ww/downloads/details.aspx?id=1192

[61] Heckerling, P. S., Canaris, G., Flach, S. D., Tape, T. G., Wigton, R. S., & Gerber, B. S. (2007). Predictors of urinary tract infection based on ANNS & genetic algorithms. International journal of Medical Informatics, 76(4), 289-296.

[62] Hellriegel, & Slocum. (2009). Comportamiento Organizacional. México, D.F.: Cengage Learning Editores. From http://books.google.com.mx/books?hl=es&lr=&id=__g324XjZNwC&oi=fnd&pg=PR25&dq=toma+de+decisiones+gerenciales&ots=7k7_vdYHXk&sig=yPHU5MSjY57MjLy7h5FNw83bgp4#v=onepage&q=toma%20de%20decisiones%20gerenciales&f=false

[63] Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). Metodología de la investigación. México, DF.: Mc Graw Hill/Interamericana Editores.

[64] Hill, C. W., & Jones, G. R. (2009). Administración Estratégica. México, D.F.: Mc Graw Hill.

[65] Hitt, M. A., Black, J. S., & Porter, L. W. (2014). Management. Edinburg Gate, Harlow: Pearson Education Limited.

[66] Instituto Nacional de Evaluación Educativa. (2013). PANORAMA DE LA EDUCACIÓN: Indicadores de la OCDE 2013, Informe Español. Madrid, España: MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE.

[67] Isik, L., Leibo, J. Z., & Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. Front. Comput. Neurosci, 6(37).

[68] Kadhim A-Shayea, Q. (2011). Artificial neuronal networking in medical diagnosis. International journal of computer science Issues, 8(2).

[69] Kelly, P. K., & Gorín, J. (1999). Las Tecnicas para la Toma de Decisiones en Equipo: Guia Practica para Obtener Buenos Resultados. Ediciones Granica SA.

[70] Kress, M. (2010). Intelligent Business Process Optimization for the Service Industry. Demand: KIT Scientific Publishing.

[71] Krogerus, M., & Tschäppeler, R. (2011). The decision book. London, UK: Profile Books LTD.

[72] Leadership Business Group. (2014). Lo que hacen los líderes eficaces . Obtenido de Grupo Empresarial: http://www.leadership-bg.com/index.php?option=com_content&id=345:o-que-fazem-os-lideres-eficazes&catid=71&Itemid=86&lang=es

[73] Ledesma, R., Molina Ibañez, G., & Valero Mora, P. (2002). Análisis de consistencia interna mediante Alfa de Cronbach: un programa basado en gráficos dinámicos. Psico-USF, 7(2), 143-152.

[74] M.P. van der Aalst, W. (2011). Process Mining: Discovery, Conformance and Enhancement of Business Processes (Google eBook). London, UK: Springer-Verlag.

[75] Maldonado, Á., & Giandini, V. (2010). La conformación de un Sistema Tutorial para el Curso de Nivelación a Disancia de la Facultad de Ingeniería de la UNLP: decisiones, dificultades e impacto. TE&ET | Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología(4), 47-53.

[76] Manpowerve. (2012 йил Octubre). Marpowerve Blog. From La importancia de identificar habilidades y logros: http://manpowerve.blogspot.mx/2012/10/la-importancia-de-identificar.html

[77] McCormack, L., Bann, C., Squiers, L., Berkman, N. D., Squire, C., Schillinger, D., Hibbard, J. (2010). Measuring Health Literancy: A Pilot Study of a New Skills-Based Instrument. Journal of Heatlth Communication: International Perspectives, 15(S2), 51-71.

[78] Méndez del Río, L. (2006). Más allá del Business Intelligence: 16 experiencias de éxito. Barcelona: Gestión 2000.

[79] Mitchel, T. M. (1997). Machine Learning. Singapore: Mc Graw Hill.

[80] Mitchell, T. M. (2000). Decision Tree Learning. Obtenido de Washington State University: http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CD0QFjAA&url=http%3A%2F%2Fwww.cs.princeton.edu%2Fcourses%2Farchive%2Fspr07%2Fcos424%2Fpapers%2Fmitchell-dectrees.pdf&ei=QAY7U-ngA4Xo2QXw7oCwAQ&usg=AFQjCNGMko1QbJWpXS8K8lzSZ

[81] Monahan, G. E. (2000). Management Decision Making. Cambridge, UK.: Cambridge University Press.

[82] Montero Lorenzo, J. M. (2007). Minería de Datos: Técnicas y herramientas. Madrid, España: Thomson Ediciones Paraninfo S.A.

[83] Muñoz Pérez, J. (2010). Inteligencia computacional inspirada en la vida. Málaga, España: Servicio Publicaciones UMA.

[84] Narciso, G. N. (2008). La función tutorial en la Universidad en el actual contexto de la Educación Superior. Revista Interuniversitaria de Formación del Profesorado, 22(1), 21-48.

[85] Nuevas Tecnologías. (2011). Fernández Editores. Obtenido de La elección de una carrera o un trabajo: http://www.tareasya.com.mx/index.php/tareas-ya/secundaria/formacion-civica-y-etica/el-individuo/1728-Habilidades,-aptitudes-e-intereses.html

[86] OCDE. (2014). Resultados TALIS 2013 (Teaching And Learning International Survey). Obtenido de OCDE Education: http://www.oei.es/noticias/spip.php?article14148&debut_5ultimasOEI=15

[87] OMG. (2014). Business Process Model and Notation. Obtenido de Home: http://www.bpmn.org/

[88] Pérez, P., De La Calleja, J., Medina, M. A., & Benitez, A. (2012). Aplication of machine learning to classify dialetic retinopathy. SPPRA, 146-153.

[89] Pontificia Universidad Católica de Argentina. (2012). Pontificia Universidad Católica de Argentina. Obtenido de Importancia de la tutoría y de una formación: http://www.uca.edu.ar/index.php/site/index/es/uca/facultades/buenos-aires/ingenieria/alumnos/tutorias/

[90] Prawda, J. (2004). Métodos y modelos de investigación de operaciones: modelos determinísticos. México: Limusa.

[91] Prieto Herrera, J. E. (2013). Investigación de mercados. Bogotá-Colombia: Ecoe Editores.

[92] Real Academia Española. (2013). Diccionario de la Lengua Española. Obtenido de afinidad: http://lema.rae.es/drae/?val=afinidad

[93] Render, B., Stair, R. J., & Hanna, M. E. (2006). Métodos cuantitativos para los negocios. Prentice Hall.

[94] Repetto, E., & Beltrán, S. (2009). Formación en competencias socioemocionales. Madrid, España: La Muralla.

[95] Rico García, M. G., & Sacristán Navarro, M. (2012). Fundamentos empresariales. Madrid: ESIC, Editorial.

[96] Rodríguez, A. (s.f.). La función directiva. México D.F.: UNAM.

[97] Rodríguez, R., & Mislej, E. (2013). Aprendizaje Automático - Machine Learning. Obtenido de Departamento de Computación- Universidad de Buenos Aires: http://www.dc.uba.ar/materias/aa/2013/cuat1

[98] Roman, J. D., & Ferrández, M. (2008). Liderazgo y coaching. LibrosEnRed.

[99] Rué, J. (2009). El aprendizaje autónomo en educación superior. Barcelona, España: Narcea Ediciones.

[100] Russell, S., & Norvig, P. (2009). Artificial Intelligence: A Modern Approach (3era. ed.). Prentice Hall.

[101] Rzaev, R., Aliev, E., Akbarov, R., & Askerov, N. (2013). Estimation of the Efectiveness of Regional Investment Projects by Fuzzy Conclusion Method. In A.

M. Gil-Lafuente, L. Barcellos-Paula, J. M. Merigó-Lindahl, F. A. Silva-Marins, & A. C. De Azevedo-Ritto, Decision Making Systems in Business Administration (pp. 27-37). Singapur: World Scientific Publishing.

[102] Sabherwal, R., & Becerra-Fernandez, I. (2011). Business Intelligence: Practices, Technologies and Management. United States of America: Wiley.

[103] Salles J., A. A. (2011). Lean production and bussiness efficiency: An artificial neural network analysis in auto parts companies. Techonology management conference (ITMC) (pp. 855-863). Sao Paulo: IEEE Xplore Digital Library.

[104] Sanz de Acedo Lizarraga, M. L. (2010). Competencias congnitivas en educación superior. Madrid, España: Narcea.

[105] Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). JRTE: Journal of Research on Technology in Education, 42(2), 123-149.

[106] Schmitz, C. (2010). LimeSurvey. From The project: https://www.limesurvey.org

[107] Serra de la Figuera, D. (2004). Métodos cuantitativos para la toma de decisiones. España: Ediciones Gestión 2000.

[108] Shawe-Taylor, J. (2006). Machine learning research topics and application field. Obtenido de Centre for Computational Statistics and Machine Learning: http://www.csml.ucl.ac.uk/courses/msc_ml/?q=node/37

[109] Sinnexus. (2012). Business Intelligence. Obtenido de ¿Qué es Business Intelligence?: http://www.sinnexus.com/business_intelligence/

[110] Slade, S. (1994). Goal-based decision making: an interpersonal model. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers.

[111] SoyEntrepreneur.com. (2014). 5 consejos para formar equipos efectivos. Obtenido de Recursos Humanos: http://www.soyentrepreneur.com/26530-5-consejos-para-formar-equipos-efectivos.html

[112] Sprenger, M. (2010). Brain-Based Teaching in the Digital Age. Aurora, USA: Assn for Supervision & Curricu.

[113] Stentoft Arlbjørn, J. (2010). Business Process Optimization. Denmark: Academica.

[114] Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. EUA.: Addison-Wesley.

[115] The University of Waikato. (2013). Weka 3: Data Mining Software in Java. Obtenido de Machine Learning Group at the University of Waikato: http://www.cs.waikato.ac.nz/ml/weka/

[116] Tirenni, G., Kaiser, C., & Herrmann, A. (2007). Applying decision trees for value-based customer relations management: Predicting airline customers' future values. Database Marketing & Customer Strategy Management, 14(2), 130–142.

[117] Torres Pérez, V., Sánchez García, J., & Ramírez Gutiérrez, D. (2014). Los 6 pecados capitales en la inteligencia de negocios. Obtenido de IPADE: http://www.ipade.mx/editorial/Pages/articulo-los-6-pecados-capitales-en-la-inteligencia-de-negocios.aspx

[118] Tufféry, S. (2011). Data Mining and Statistics for Decision Making. United Kingdom: John Wiley & Sons.

[119] UNACH. (2010). Manual de modalidades de la tutoría. Tuxtla Gutiérrez: UNACH.

[120] Universidad Politécnica de Puebla. (2011). Certificaciones/ ISO 9000. Obtenido de: Universidad Politécnica de Puebla: http://serpaguppue.uppuebla.edu.mx/ISO9000.php

[121] Universidad Politécnica de Puebla. (2011). Oferta Educativa. Obtenido de: Universidad Politécnica de Puebla: http://serpaguppue.uppuebla.edu.mx/

[122] Universidad Politécnica de Puebla. (2013). Oferta Educativa. Obtenido de Universidad Politécnica de Puebla: http://serpaguppue.uppuebla.edu.mx/

[123] Universidad Politécnica de Puebla. (2013). Informe de Labores 2012-2013. Puebla: Universidad Politécnica de Puebla.

[124] Urbina Nájera, A. B., de la Calleja, J., Vega Lebrún, C. A., López Maldonado, N., & Pico González, B. (2014). Desarrollo y validación de un instrumento para identificar perfiles de tutorados y tutores de la modalidad virtual. CAFVIR-2014 (págs. 227-234). Antigua Guatemala: CAFVIR.

[125] Vaello Orts, J. (2009). El profesor emocionalmente competente: Un puente sobre "aulas" turbulentas. Barcelona, España: Grao.

[126] Vercellis, C. (2009). Business Intelligence: Data mining and optimization for decision making. United Kingdom: Wiley.

[127] Vintar, J., Goldsman, A., & Asimov, I. (2004). I Robot. Los Ángeles, California, Estados Unidos.

[128] Wikipedia. (2014). Supervised Learning. Obtenido de Wikipedia: http://en.wikipedia.org/wiki/Supervised_learning

[129] Wikipedia. (2014). Unsupervised Learning. Obtenido de Wikipedia: http://en.wikipedia.org/wiki/Unsupervised_learning

[130] Williams, S., & Williams, N. (2007). The profit impact of Business Intelligence. San Francisco, CA.: Elsevier.

[131] Winograd, M., Fernández Lamarra, N., & Farrow, A. (1998). Herramientas Para la Toma de Decisiones en América Latina Y El Caribe: Indicadores Ambientales Y Sistemas de Información Geográfica. CIAT.

[132] Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. San Francisco, CA.: Morgan Kaufmann Publishers (Elsevier).

[133] Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. San Francisco, CA: ELSEVIER.

[134] Yokono, J. J., & Poggio, T. (2009). Object Recognition Using Boosted Oriented Filter Based Local Descriptors. IEEJ Transactions on Electronics, Information and Systems, 129(5), 806-811.