

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

Facultad 5, Laboratorio de Investigaciones en Gestión de Proyectos

Facultad 3, Departamento de Ingeniería y Gestión de Software



**DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE LA RELACIÓN RASGOS DE LA
PERSONALIDAD-RENDIMIENTO LABORAL EN PROYECTOS INFORMÁTICOS**

Trabajo final presentado en opción al título de
Máster en Gestión de Proyectos Informáticos

Autor: Ing. Eilys Pacheco Rodríguez

Tutor: DrC. Rafael Rodríguez Puente

Co-tutores: MSc. Elizabeth Rodríguez Stiven

MSc. Nemury Silega Martínez

La Habana, Noviembre de 2014

AGRADECIMIENTOS

A mi padre, donde quiera que esté, por ser mi ángel y darme la fuerza para reponerme ante las dificultades.

A Osni por estar a mi lado en los momentos buenos y malos de la vida. Por compartir mis sueños y brindarme su comprensión e incondicionalidad. Ojalá podamos estar juntos toda la vida.

A mi familia, en especial a mi madre, mi hermana, mi tío y mi abuela que son mi razón de ser. En general a todos, por las preocupaciones que les causé.

A mis tutores, quiero agradecerles especialmente por su ayuda incondicional y los conocimientos que me brindaron. Les debo este resultado pues sin su apoyo no hubiera sido posible realizar esta investigación.

A Yoe, no tengo palabras para expresarte lo agradecida que te estoy.

A Nadia por su invaluable ayuda y ánimo en el momento que más lo necesitaba, gracias a su contribución hoy existe este trabajo.

A mis amigos de Aduana, especialmente a Leo, Eddie, Yorlen, Levis, Yordi, Lauri, Adri. A Isa y Ailia. En fin, a todos por su preocupación constante sobre el avance de esta investigación.

A Aile enormemente agradecida, a Yoansy, a Yadi, a Haydesita, a Yane, a Vanci, a Yanelis, a Rosita y a Liannis por su ayuda a tiempo y su exigencia.

A Carli y a Lili por alarme todo el tiempo para salir juntos de esto. Gracias.

A todos los que formaron parte de esta investigación ya sea por haber respondido aquella enorme encuesta o por preguntar de buena fe si necesitaba ayuda.

A los profesores de la maestría por hacer de mí una profesional mejor preparada. En especial a Yadenis, Sury, Alejandro, Pedro, Michael, Eliana y a Kari.

A la Revolución por regalarnos a todos la posibilidad de ser universitarios y de continuar superándonos.

DECLARACIÓN JURADA DE AUTORÍA

Declaro por este medio que yo Eilys Pacheco Rodríguez, con carnet de identidad 86062619936 soy la autora principal del trabajo final de maestría “Descubrimiento de conocimiento a partir de la relación rasgos de la personalidad-rendimiento laboral en proyectos informáticos”, desarrollada como parte de la Maestría en Gestión de Proyectos Informáticos y que autorizo a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Y para que así conste, firmo la presente declaración jurada de autoría en La Habana a los __ días del mes de _____ del año _____.

Eilys Pacheco Rodríguez

Nombre del Autor

Firma

Rafael Rodríguez Puente

Nombre del Tutor

Firma

RESUMEN

En la actualidad, el éxito de una institución depende en gran medida de los recursos humanos (RRHH) involucrados en sus procesos. Según estudios, el rendimiento de los RRHH es consecuencia del tipo de personalidad que presenta el individuo que labora en la institución. Como parte del progreso de la industria informática, numerosas organizaciones han centrado sus esfuerzos en identificar cuáles son los factores que influyen en el rendimiento laboral de sus RRHH.

El objetivo de esta investigación es descubrir conocimiento a partir de la relación existente entre los rasgos de la personalidad y el rendimiento laboral de los RRHH, mediante la aplicación de técnicas de Minería de Datos. Se obtienen como resultado del descubrimiento 20 rasgos relevantes de la personalidad y un conjunto de 76 reglas de decisión para los índices de rendimiento IRHA, IRHE, IRHT e IRHF que expresan la relación entre las variables analizadas: rasgos de la personalidad y rendimiento laboral. Además, se propone un patrón ideal de la personalidad expresado mediante reglas que toman en cuenta los rasgos relevantes identificados. Estos resultados contribuyen a la toma de decisiones en los procesos de la Gestión de los RRHH en el contexto organizacional de proyectos informáticos.

Palabras clave: Minería de Datos, personalidad, recursos humanos, rendimiento.

ABSTRACT

Nowadays, the success or failure of an institution depends heavily on the human resources (HR) involved on its processes. According to several studies, the working staff's performance is a consequence of the type of personality each of its members has. As part of the development of this industry, numerous organizations have focused its efforts on identify the factors that influence on the job performance of HR.

The objective of this research is to discover knowledge based on the relationship between the personality traits and HR's job performance by applying the decision trees Data Mining technique. The results of this research includes the discoveries of 20 relevant personality traits, as well a set of 76 decision rules for the performance indexes IRHA, IRHE, IRHT and IRHF, which express the relationship between the studied variables: personality traits and job performance. Moreover, an ideal personality pattern is proposed, expressed through rules that take into account some relevant traits. These results contribute to the decision making processes of HR management in the organizational context of IT projects.

Key words: data mining, human resources, performance, personality.

ÍNDICE GENERAL

INTRODUCCIÓN	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA SOBRE EL DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE LA RELACIÓN RASGOS-RENDIMIENTO EN PROYECTOS INFORMÁTICOS.....	8
ANÁLISIS BIBLIOMÉTRICO.....	8
RASGOS DE LA PERSONALIDAD DE LOS RRHH	8
<i>Análisis de modelos de la personalidad aplicados en proyectos informáticos</i>	<i>10</i>
EVALUACIÓN DEL RENDIMIENTO LABORAL DE LOS RRHH EN PROYECTOS INFORMÁTICOS.....	13
DESCUBRIMIENTO DE CONOCIMIENTO.....	15
<i>Análisis de técnicas de Minería de Datos</i>	<i>16</i>
<i>Análisis de algoritmos de clasificación de árboles de decisión.....</i>	<i>18</i>
<i>Análisis de herramientas de Minería de Datos.....</i>	<i>20</i>
<i>Análisis de sistemas basados en el conocimiento.....</i>	<i>23</i>
CONCLUSIONES DEL CAPÍTULO	25
CAPÍTULO 2: DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE LA RELACIÓN RASGOS-RENDIMIENTO EN PROYECTOS INFORMÁTICOS	26
PROCESO DE DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE LA RELACIÓN ENTRE LOS RASGOS DE LA PERSONALIDAD Y EL RENDIMIENTO LABORAL DE LOS RRHH	26
<i>Subproceso Base de Casos.....</i>	<i>27</i>
<i>Subproceso Minería de datos</i>	<i>31</i>
<i>Subproceso Descubrimiento</i>	<i>39</i>
CONCLUSIONES DEL CAPÍTULO	44
CAPÍTULO 3: VALIDACIÓN DE LOS RESULTADOS.....	45
ESTRATEGIA DE VALIDACIÓN.....	45
<i>Validación de los rasgos relevantes.....</i>	<i>46</i>
<i>Validación de las reglas de decisión generadas</i>	<i>49</i>
<i>Validación de la aplicación del algoritmo J4.8.....</i>	<i>54</i>

<i>Validación del patrón ideal de la personalidad</i>	55
ANÁLISIS DEL IMPACTO ECONÓMICO Y SOCIAL DE LA PROPUESTA	57
CONCLUSIONES DEL CAPÍTULO	58
CONCLUSIONES	59
RECOMENDACIONES	60
REFERENCIAS BIBLIOGRÁFICAS.....	61
GLOSARIO DE TÉRMINOS	67
ANEXOS	68
ANEXO 1. REGLAS DE DECISIÓN GENERADAS QUE FUERON SELECCIONADAS POR EL NIVEL DE PRECISIÓN PARA LAS CLASES IRHA, IRHF E IRHT.	68
ANEXO 2. ENCUESTA APLICADA A LOS EXPERTOS PARA VALIDAR LA INFLUENCIA DE LOS RASGOS DE PERSONALIDAD EN EL RENDIMIENTO LABORAL.....	73

INTRODUCCIÓN

Actualmente, las innovaciones tecnológicas y la competitividad entre las empresas informáticas en el mundo se ven en crecimiento diario. Es por esto que en (Eyzaguirre 2003) se plantea que para que exista crecimiento en un mundo globalizado, la competitividad tiene que mejorar permanentemente.

Dadas la diversidad y aceleración actuales del intercambio tecnológico, cultural y de información, la posibilidad de las empresas de competir y superar a sus pares a través de ventajas en cualquiera de estos ámbitos es cada vez menor. En la búsqueda de variables organizacionales que permitan marcar una diferencia, los expertos han concluido que a través de los RRHH, una empresa puede agregar valor a sus operaciones (Chiavenato 1994; Goleman 1999). Por lo cual el éxito de cada organización se ve condicionado a su principal activo tangible, el personal que labora; razón suficiente para centrar sus esfuerzos en contratar RRHH idóneos para obtener altos niveles de rendimiento en la industria.

Son varias las investigaciones donde se reconoce que los RRHH juegan un papel crítico en el éxito o fracaso de un proyecto (Chiavenato 2001; Pinto 2007; García 2008). El estudio de análisis de tendencias en las empresas de Tecnologías de Información, identificó que aproximadamente el 75% de los encuestados, expresan que sus organizaciones planean inversiones en las habilidades de su personal (Standish Group 2008).

Sin embargo el personal sigue siendo el factor menos formalizado en los modelos de procesos y las metodologías de desarrollo de software. Las metodologías se centran más en los aspectos técnicos que en los aspectos humanos. Dentro de los principales problemas mencionados en (Robbins 2004) sobre el área de conocimiento Gestión de RRHH en la Gestión de Proyectos se encuentran:

- Escaso trabajo en equipo.
- Aumento de conflictos que atentan contra la eficiencia.
- Liderazgos ineficientes.
- Desconocimiento de las características propias del personal que labora.
- Asignación inadecuada de tareas y responsabilidades.
- Evaluaciones injustas en la medición del desempeño laboral.

Estos y muchos otros problemas conllevan constantemente a frecuentes fracasos en las organizaciones. Desafortunadamente la mayoría de las veces se construyen respuestas intuitivas a estos problemas y se toman decisiones basadas solamente en percepciones subjetivas (Robbins 2004).

En este sentido, la realidad actual impone la gestión adecuada del talento humano, considerando los procesos a los que ellos son sometidos dentro de las instituciones. En este aspecto los directivos juegan un papel fundamental y en consecuencia, se les debe exigir una actuación permanentemente orientada a alcanzar y mantener la motivación, el interés, el compromiso, la participación y el buen desempeño de sus empleados. Este desempeño laboral constituye un elemento fundamental para el funcionamiento de

cualquier organización, por lo que debe prestársele especial atención dentro del proceso de Gestión de RRHH. La evaluación de dicho rendimiento debe proporcionar beneficios a la organización y a las personas en virtud de contribuir a la satisfacción de los trabajadores para garantizar el alcance de los objetivos institucionales (Pedraza et al. 2010).

A raíz de esto, numerosas organizaciones informáticas utilizan sistemas y modelos de evaluación de los RRHH soportados por las Tecnologías de la Información y las Comunicaciones (TIC). Estos sistemas permiten verificar el nivel del personal teniendo en cuenta sus habilidades de acuerdo al rol que desempeñan, sus fortalezas y debilidades, los resultados, las competencias que presentan así como su rendimiento.

Sin embargo, la mayoría de estas instituciones desconoce que un gran número de factores pueden influir en el rendimiento laboral. Los empleados pueden sentirse influenciados negativamente por la microgestión por parte de sus supervisores, por el cambio de labor o simplemente por otros factores que pueden necesitar atención inmediata: los factores emocionales u otros específicos del aprendizaje (Pedraza, Amaya et al. 2010). Los empleados probablemente se sentirán inspirados e influenciados positivamente con equipos de alta calidad y un estilo de gestión del supervisor que sea accesible, sin importar el campo o la industria en la que trabajen (Pascarella y Terenzini 1983).

Varias son las investigaciones centradas en descubrir cuáles factores inciden negativa o positivamente en el rendimiento laboral de un individuo y en qué medida lo hace cada uno de ellos, a fin de implementar las estrategias que incidan en la mejora del rendimiento.

En (Chiavenato 2001) se plantea que el rendimiento está referido "(...) por una parte, a las capacidades y habilidades del individuo para realizar una tarea y por la otra, a las percepciones que él tiene del papel que debe desempeñar en la organización, indicando un enfoque sistémico".

Otros autores (Landeta et al. 2011) señalan que el rendimiento laboral se condiciona por algunos aspectos psicológicos entorno a los antecedentes familiares y educativos; las características personales; los conocimientos y las funciones psicológicas básicas, como son la percepción, la memoria y la conceptualización; además del compromiso por alcanzar las metas.

En (Jano y Ortiz 2005) se han investigado los factores que afectan al rendimiento en las organizaciones, llegando a la conclusión de que uno de ellos es el esfuerzo efectivo realizado por el individuo, así como las habilidades y conocimientos previos a su labor.

Algunos individuos pueden presentar problemas intelectuales, es decir, su capacidad intelectual es inferior a la media, lo que hace que el aprendizaje sea una labor aún más difícil, aunque paradójicamente también existe la posibilidad de que su capacidad intelectual sea superior a la media, es decir, que sea un individuo superdotado, en cuyo caso también se presentará el problema de rendimiento laboral (CEDHAP 2014). A pesar de que el resto de los factores propuestos en (Colectivo de autores Universia 2012) afectan directamente en el rendimiento laboral, existe una estrecha relación con el factor emocional condicionado por las habilidades, pensamientos y comportamientos de un individuo.

Cuando se hace referencia a altos índices de rendimiento en el trabajo se alude al estado afectivo de agrado y altos conocimientos que experimenta un individuo al afrontar su realidad laboral, es decir, el conjunto de condiciones presentes en el ambiente de trabajo. Igualmente, constituye la expresión emocional de la percepción y posee tanto elementos cognitivos como conductuales. Es válido señalar, que son varios los factores que determinan el rendimiento laboral, destacándose entre otros, el trabajo mentalmente desafiante, que permitan el uso de las habilidades cognoscitivas y emocionales propias, apoyo de colegas y compañeros amigables, evaluaciones justas y un supervisor comprensivo, variedad de tareas e idoneidad en la selección, libertad y retroalimentación (Guillén y Bozal 2000; Chiavenato 2001).

En el Centro de Informatización de Entidades (CEIGE) de la Universidad de las Ciencias Informáticas (UCI) se realizó un estudio a 524 individuos seleccionados aleatoriamente, agrupados en: estudiantes, profesores y especialistas de la informática. Este estudio arrojó como resultados que el 82.6 % había sido evaluado con un índice de rendimiento en el trabajo bajo o medio durante el período 2011-2013 mediante la herramienta GESPRO¹.

A raíz de este estudio, se decidió aplicar una encuesta a 50 profesionales del mismo centro que ocupaban las responsabilidades: jefes de equipo, jefes de proyecto y jefes de departamento. Esta encuesta permitió, a través de sus resultados, determinar que existe diversidad de opiniones sobre la evaluación del rendimiento de los subordinados y que se desconoce que la personalidad es otro de los factores que lo determinan. El 83% de los entrevistados afirman que el rendimiento de las personas, entre otros factores, está condicionado mayormente por la experiencia, las competencias específicas en el rol que desempeñan, la motivación y los comportamientos específicos del individuo. Sin embargo, el 17% afirma que el rendimiento se basa solamente en los índices de productividad individual de los miembros de un equipo ante las tareas asignadas, sin importar los factores mencionados anteriormente.

A partir de lo anteriormente planteado se enuncia el **Problema** de la investigación:

El desconocimiento de la relación existente entre los rasgos de la personalidad y el rendimiento laboral de los RRHH en proyectos informáticos de la UCI, está afectando la gestión del conocimiento.

Para solucionar este problema, se define como **Objeto de investigación**: descubrimiento de conocimiento.

Para dar solución al problema generado se perfiló el siguiente **Objetivo general**:

Descubrir conocimiento a partir de la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH en proyectos informáticos de la UCI, mediante la aplicación de técnicas de Minería de Datos para aumentar la gestión del conocimiento.

Tributan al logro del Objetivo general los siguientes **Objetivos específicos**:

¹ Paquete de Herramientas de Gestión de Proyectos utilizada en la UCI. <http://gespro.ceige.prod.uci.cu/>

- Elaborar el marco teórico de la investigación sobre el descubrimiento de conocimiento, técnicas de Minería de Datos, creación de bases de conocimientos, rasgos de personalidad de los RRHH y la evaluación del rendimiento laboral en proyectos informáticos.
- Construir las bases de conocimientos a utilizar en la investigación a partir de datos reales documentados en el GESPRO.
- Aplicar la técnica de árboles de decisión, para descubrir la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH.
- Interpretar los resultados obtenidos a partir de la aplicación de la técnica de árboles de decisión, para expresar el conocimiento descubierto mediante la generación de reglas de decisión.
- Validar los resultados de la investigación a partir del estudio de casos y el criterio de expertos.

Centrando el desarrollo de esta investigación en el siguiente **Campo de acción**: descubrimiento de conocimiento a través de técnicas de Minería de Datos.

Tipo de investigación:

Se utiliza la investigación exploratoria y posteriormente correlacional, ya que se parte del objeto esencial de aumentar el grado de familiaridad con fenómenos relativamente desconocidos, poco estudiados o novedosos, para luego medir cómo se relacionan o vinculan dichos fenómenos entre sí (o si no se relacionan).

Las preguntas que guiarán la investigación son:

- ¿Cuáles son los índices de rendimiento laboral obtenidos por un conjunto de individuos que pertenecen a proyectos informáticos?
- ¿Cuáles son los rasgos de la personalidad que más influyen por cada índice de rendimiento laboral establecido?
- ¿Qué asociación existe entre los rasgos de la personalidad y los índices de rendimiento laboral?

El diseño de la investigación es no experimental, ya que no se construyen grupos ni situaciones sino que se observan situaciones o grupos existentes. En particular, es una investigación no experimental correlacional o causal ya que se pretende encontrar la relación entre las variables: rasgos de la personalidad y rendimiento laboral de los RRHH. La estrategia de validación de la investigación emplea el criterio de expertos y el estudio de casos, así como los instrumentos entrevistas y encuestas.

Para darle cumplimiento a los objetivos trazados se utilizaron los siguientes métodos de investigación:

Métodos teóricos:

- Hipotético-deductivo: permitirá formular las preguntas de la investigación para descubrir conocimiento a partir de la relación entre los rasgos de la personalidad y rendimiento laboral de los RRHH.

- Analítico-sintético: utilizado para descomponer el problema de investigación en elementos, profundizar en su estudio y luego sintetizarlos en la solución propuesta.
- Histórico-lógico: se empleó en el estudio crítico de los modelos de evaluación de la personalidad y en el descubrimiento de conocimiento, así como de los algoritmos de clasificación más relevantes en la Minería de Datos de acuerdo a la literatura.

Métodos empíricos:

- Medición: se utilizó para obtener información numérica acerca de las evidencias recopiladas, en función de la cantidad, el aporte de cada una de ellas a la definición de los tipos de personalidad y el nivel de la calidad del dato.
- Entrevista: se realizaron entrevistas a los equipos de dirección de los RRHH escogidos para la validación. Este método permitió obtener información sobre la problemática que se aborda en esta investigación y precisar el problema a resolver.
- Encuesta: se utilizó primeramente para determinar la problemática de esta investigación. Luego se empleó para conocer las opiniones y valoraciones de los expertos sobre los rasgos de personalidad relevantes en el rendimiento laboral y para validar los resultados obtenidos en la investigación a partir de su aplicabilidad. Se utilizó además el instrumento NEO-PI del modelo Cinco Factores de Personalidad para evaluar los rasgos de personalidad de los individuos del CEIGE que formaron parte de la investigación.
- Criterio de Expertos: para validar los datos obtenidos por el algoritmo J4.8 referente a los rasgos de la personalidad más relevantes y la aplicabilidad del patrón ideal de la personalidad propuesto en la investigación.

Para el desarrollo de la investigación se tuvo en cuenta los métodos estadísticos árboles de decisión y generación de reglas de clasificación.

Aporte de la investigación:

- Evaluación de los rasgos de la personalidad de los 250 individuos que formaron parte de esta investigación.
- Listado de los rasgos más relevantes para cada los índices de rendimiento IRHA, IRHE, IRHT e IRHF en el estudio realizado.
- Proceso de descubrimiento de conocimiento, a partir de la utilización de árboles de decisión, para determinar la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH en proyectos informáticos.
- Reglas que expresan la relación entre los rasgos relevantes de la personalidad y el rendimiento laboral de los RRHH.

- Base de conocimiento construida a partir de la evaluación de los rasgos de la personalidad y los índices de rendimiento obtenidos.
- Patrón Ideal de la personalidad.

Avales y publicaciones obtenidas:

- [Eilys Pacheco Rodríguez, Osnier Ramírez Alea][“Modelo de Gestión del Conocimiento para el Departamento Desarrollo de Soluciones para la Aduana”][UCIENCIA 2012, II Taller de Sistemas de Gestión de la Información y el conocimiento][Habana][Cuba][febrero 2012][ISBN 978-959-286-019-3].
- [Mailen Edith Escobar Pompa, Eilys Pacheco Rodríguez][“Diseño de una Base de Casos para la evaluación de competencias de equipos a partir de evidencias”][UCIENCIA 2012, VI Taller de Inteligencia Artificial][Habana][Cuba][febrero 2012][ISBN 978-959-286-019-3].
- [Eilys Pacheco Rodríguez, Yadenis Pinero Pérez, Meylin Rodríguez Socarrás, Osnier Ramírez Alea, Mailen Edith Escobar Pompa][“Modelo de Evaluación del rendimiento de los estudiantes en los proyectos de desarrollo de software del CEIGE”][UCIENCIA 2012, II Taller de Gestión de Proyectos][Habana][Cuba][febrero 2012].
- [Eilys Pacheco Rodríguez, Yadenis Pinero Pérez, Meylin Rodríguez Socarrás, Mailen Edith Escobar Pompa][“Modelo de Evaluación del aporte de los indicadores de rendimiento a las competencias genéricas de los estudiantes del CEIGE.”][XVI Convención Científica de Ingeniería y Arquitectura CUJAE 2012, I Congreso Internacional de Ingeniería Informática y Sistemas de Información CIIISI’12][Habana][Cuba][noviembre 2012].
- [Eilys Pacheco Rodríguez, Osnier Ramírez Alea, Maurice Cabrejas Martínez, Yaniselis Sánchez Hormigó][“Procedimiento para crear proyectos educativos desde la producción”][XVI Convención Científica de Ingeniería y Arquitectura CUJAE 2012, II Simposio TIC y Aprendizaje en las Ciencias Técnicas][Habana][Cuba][noviembre 2012].
- [Eilys Pacheco Rodríguez, Yoendry Betancourt, Katerin Martínez] [Minería de Datos para la evaluación de los tipos de personalidad en los RRHH en proyectos informáticos][8va Peña Tecnológica 2013].
- [Ailenis Macias, Eilys Pacheco, Yadira Calimano][“La evaluación en el proceso de enseñanza aprendizaje”][Universidad 2014 UCI].
- [Eilys Pacheco Rodríguez][“Modelo de gestión del conocimiento para el Departamento Aduana”][CIIPRO 2012, Mar del Plata, Argentina 3er Congreso Iberoamericano de Proyectos][ISBN: 978-987-1312-49-8].
- [Eilys Pacheco Rodríguez][“Modelo de evaluación del rendimiento en los proyectos de desarrollo de software”][CIIPRO 2012, Mar del Plata, Argentina 3er Congreso Iberoamericano de Proyectos][ISBN: 978-987-1312-49-8].

- [Eilys Pacheco Rodríguez, Yoendry Betancourt Peña][“Propuesta de aplicación del algoritmo GENRUL bajo el enfoque de aprendizaje supervisado en el análisis de la influencia de los rasgos de la personalidad en el rendimiento”][9na Peña Tecnológica 2014][ISBN 978-959-286-027-8].
- [Eilys Pacheco Rodríguez, Yoendry Betancourt Peña, Osnier Ramírez Alea][“Influencia de los rasgos en el rendimiento laboral de los RRHH en proyectos informáticos”][Conferencia Científica UCI][La Habana, Cuba][2014].
- [Eilys Pacheco Rodríguez, Yoendry Betancourt Peña, Osnier Ramírez Alea][“Algoritmo J4.8 en la búsqueda de las relaciones entre rasgos de la personalidad y rendimiento laboral en proyectos informáticos”][5º Congreso Iberoamericano de Ingeniería de Proyectos][Loja, Ecuador][2014].

Estructura de la tesis:

El presente trabajo se ha estructurado en tres capítulos. En el primer capítulo se realiza el marco teórico de la investigación sobre el descubrimiento de conocimiento, técnicas de Minería de Datos, creación de bases de conocimientos basadas en casos, rasgos de la personalidad de los RRHH e indicadores para evaluar el rendimiento laboral. Se resume la experiencia nacional e internacional en relación a los modelos de evaluación de la personalidad. Además, se analizan y describen técnicas y algoritmos de limpieza de datos.

En el segundo capítulo se detalla el proceso de descubrimiento de conocimiento propuesto, así como los subprocesos que lo componen. Además, se seleccionan los rasgos relevantes que más influyen en el rendimiento laboral de los RRHH y se generan las reglas como parte del descubrimiento. Este conocimiento se obtiene a partir de la construcción e interpretación de los árboles de decisión tras aplicar el algoritmo J4.8 de Minería de Datos.

En el tercer capítulo se valida el conocimiento que se descubre en el proceso de descubrimiento de conocimiento y el patrón ideal de la personalidad a través de su aplicación en estudio de casos y mediante el método de criterio de expertos. Por otro lado, se evalúa el nivel de confiabilidad presentado por el algoritmo en la clasificación de los datos a clasificar.

Por último se establecen las conclusiones, se emiten recomendaciones, se relacionan las referencias bibliográficas y se incluyen los anexos que facilitan la comprensión de la tesis.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA SOBRE EL DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE LA RELACIÓN RASGOS-RENDIMIENTO EN PROYECTOS INFORMÁTICOS

En el presente capítulo se realiza un estudio de los conceptos asociados a la investigación referente a los rasgos y tipos de personalidad de los RRHH a través del uso de la bibliografía consultada, asumiendo una posición crítica. Se caracterizan modelos que evalúan varios rasgos de la personalidad que se han considerado para actividades de evaluación en el ámbito del software, así como los instrumentos que utilizan. Además, se analizan indicadores para evaluar el rendimiento laboral y las experiencias de su aplicación en la Gestión de los RRHH en proyectos informáticos.

Se describen a su vez técnicas de Minería de Datos, profundizándose en los árboles de decisión. Se realiza una revisión de las tendencias sobre la construcción de las bases de casos, así como las principales técnicas de limpieza de datos. Como resultado de ese análisis crítico, se emite una valoración donde se identifican las mejores experiencias obtenidas, resultado de la revisión teórica realizada.

Análisis bibliométrico

En la presente investigación se utilizaron los buscadores especializados Scirus y Google Académico, se consultaron las bases de datos bibliográficas SCIELO, Latindex y Redalyc para realizar búsquedas asociadas a las temáticas definidas.

En la Tabla 1 se realiza el desglose de las fuentes bibliográficas consultadas y referenciadas en la investigación por fecha y por ciento de utilización.

Tabla 1: Desglose del análisis bibliométrico realizado en la investigación

Bibliografía	Últimos 5 años		Años anteriores	
Libros y monografías	6	33.3 %	12	66.7 %
Tesis de doctorado	3	100 %	0	0 %
Tesis de maestría	3	75 %	1	25 %
Tesis de grado	1	100 %	0	0 %
Artículos en Revistas referenciadas en Web of Science, SCOPUS	5	83.3 %	1	16.7 %
Memorias de eventos	9	81.8 %	2	18.2 %
Artículos publicados en la web	7	41.2 %	10	58.8 %
Páginas web	8	100 %	0	0 %
Reportes técnicos y conferencias	4	80 %	1	20 %
Entrevistas personales	3	100 %	0	0 %
Total	49	64.5 %	27	35.5 %

Rasgos de la personalidad de los RRHH

El interés por el estudio y medida de las diferencias individuales, así como la necesidad de evaluar a los individuos en determinadas situaciones han llevado a algunos autores a prestar mayor importancia a la evaluación psicológica de la personalidad de los RRHH.

Una aproximación al análisis de la personalidad debe considerar las cuestiones teóricas propias de esta área de estudio (Bermúdez et al. 2003). El amplio rango de fenómenos que se incluyen en el concepto de personalidad y la evolución que ha seguido este campo hace difícil la delimitación del tema; para aclararlo se conceptualizan los términos utilizados en la investigación.

Desde el punto de vista psicológico, (González 1987) define la personalidad como "(...) un constructo psicológico con el que se refiere a un conjunto dinámico de características de una persona. Pero nunca a un conjunto de características físicas o genéticas que determinan a un individuo, es su organización interior la que los hace actuar de manera diferente ante una o varias circunstancias".

Según el gran estudioso (Allport 1968), la personalidad es "(...) la organización dinámica de los sistemas psicofísicos que determina una forma de pensar y de actuar, única en cada sujeto en su proceso de adaptación al medio".

Tomando como base el concepto anterior, en la presente investigación la autora concluye que la personalidad puede sintetizarse, como el conjunto de características o patrón de sentimientos, emociones y pensamientos ligados al comportamiento. Es decir, los pensamientos, sentimientos, actitudes, hábitos y la conducta de cada individuo, que persiste a lo largo del tiempo frente a distintas situaciones, distinguiendo a un individuo de cualquier otro y haciéndolo diferente a los demás.

Los primeros teóricos de la personalidad, se refirieron a la personalidad en términos de tipos o tipologías, las cuales constan de categorías diferentes y aisladas en las que puede ser ubicada una persona (Jung 1971). El fomento de la psicología de la personalidad condujo a un cambio importante en el avance del estudio científico de la personalidad con la incorporación del concepto de rasgo (González 1987; Feldman 2009).

En los estudios de personalidad es común encontrar la caracterización del individuo a partir de la agrupación de los comportamientos afines en rasgos; y estos rasgos se agrupan en otros rasgos superiores y eventualmente en un número determinado de dimensiones (Pueyo 1996). En este enfoque los comportamientos de los rasgos se extraen a partir del análisis de los hábitos, y a su vez, estas respuestas habituales se extraen de respuestas específicas del individuo en diversas situaciones (Feist y Feist 2009).

En la Figura 1 se representa la jerarquía descrita anteriormente.

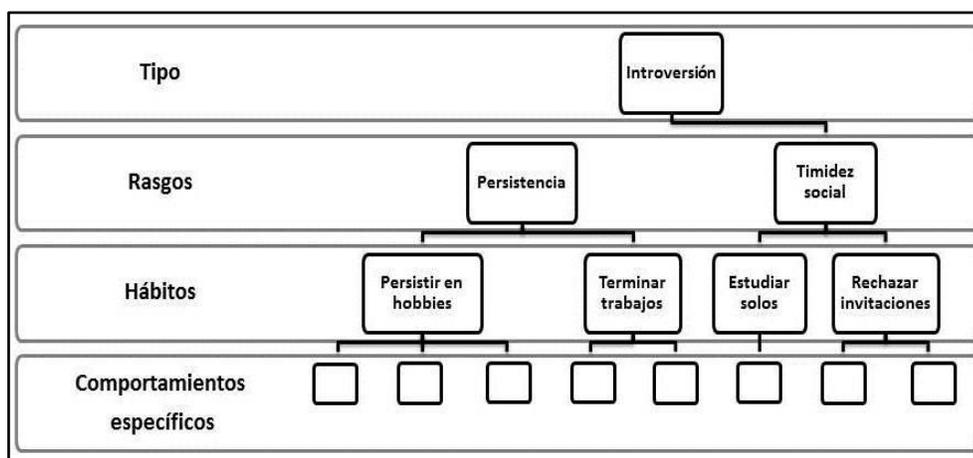


Figura 1: La definición del comportamiento en niveles. Adaptado de (Feist y Feist 2009)

Como parte de la investigación se asume la definición abordada por (Feldman 2009) por la sencillez y claridad del concepto que emplea para definir los rasgos. Este autor considera que "(...) los rasgos ofrecen una explicación clara y sencilla de las consistencias conductuales de las personas que permiten comparar fácilmente a una persona con otra. Los rasgos son característicos de la personalidad y comportamientos consistentes que se manifiestan en diferentes situaciones. La teoría de los rasgos busca explicar, de forma sencilla, las consistencias en el comportamiento de los individuos". Ese comportamiento tiene una tendencia a repetirse a través del tiempo de una forma determinada, sin que quiera decir que esa persona se comporte de modo igual en todos los casos.

En esta investigación se aborda la caracterización del individuo teniendo en cuenta los rasgos de su personalidad. Al caracterizar la personalidad se requiere segmentarla presuponiendo por ejemplo, que en un individuo la inteligencia está por un lado y el área afectiva por otra. Ahora bien, realizar un análisis riguroso de la personalidad de un individuo (integración del estudio) es una tarea laboriosa que demanda conocimientos previos, sustentados en el estudio de teorías de la personalidad y modelos que abordan el tema.

Análisis de modelos de la personalidad aplicados en proyectos informáticos

La comunidad científica se refiere a la personalidad mediante teorías. Una de las más extensamente estudiadas y aplicadas en el análisis de la personalidad es la teoría de (Jung 1971). A partir de las tres dimensiones especificadas en "Psychological types" por (Jung 1971), se han definido varios modelos e instrumentos enfocados en diversos aspectos de la personalidad tales como: el Keirsey Temperament Sorter (Keirsey 1996), Personal Style Inventory (PSI), Five Factor Model (FFM) y Myers-Briggs Type Indicator (MBTI) (Briggs y Myers 2004). Dentro de este grupo, el FFM y el MBTI son los más utilizados en el estudio de la estructuración cognitiva y las características de la personalidad en el ámbito profesional (Stiven 2012).

Modelo de Myers-Briggs

El modelo de Myers-Briggs es utilizado en el estudio de la estructuración cognitiva y las características de la personalidad en el ámbito profesional. Presenta cuatro dimensiones de las preferencias humanas que se miden en factores bipolares (Fernández 2008; Chapman 2011). A partir de los valores de cada dimensión se identifica el tipo psicológico de la persona entre los 16 tipos posibles, resultante de esta combinación de factores. Este modelo es un método para entender la personalidad y los modos de comportamiento preferidos de los individuos; es particularmente útil para entender y desarrollar la personalidad propia de un individuo y de otros compañeros, descubrir qué motiva a otros, cuáles son sus debilidades y fortalezas.

El modelo se ha utilizado fundamentalmente para identificar: el tipo o el patrón que predomina en el personal del software, los tipos más adecuados para desempeñar tareas y roles en el proceso de desarrollo, para evaluar el impacto de la diversidad de tipos en el equipo y de las distintas dimensiones en

los resultados del proyecto. Es un modelo relativamente fácil de aplicar en las instituciones, pero está sujeto a derechos de autor (Stiven 2012).

Modelo de Belbin

Otro de los modelos más utilizados mundialmente en proyectos informáticos es el modelo de Belbin caracterizado por evaluar la personalidad del equipo de trabajo (Gutiérrez 2003). Este modelo permite identificar los roles de equipo preferidos y evitados de cada persona, utilizados para identificar las fortalezas y debilidades del comportamiento de las personas en el entorno laboral. Además, permite seleccionar y desarrollar equipos de alto rendimiento; incrementar el autoconocimiento y la eficacia personal, generar confianza, comprensión mutua y ayudar en los procesos de reclutamiento (Belbin Associates 2010; Belbin 2010).

El modelo de Belbin ha sido utilizado en investigaciones vinculadas al software para evaluar el impacto de diferentes roles en equipos de proyectos (Stevens y et al 2002). Mientras (Rajendran 2005) publicó un estudio realizado sobre equipos de trabajo de hardware y software en el que concluye que los perfiles de Belbin bien balanceados contribuyen a la efectividad del equipo, que es su hipótesis fundamental. Por otro lado, (André 2009) utiliza los roles para definir patrones que tributan a la formación adecuada de equipos de proyectos de software.

Para el desarrollo de esta investigación no se asumirá debido a que no es objetivo identificar los roles de un equipo, sino los rasgos de las personas de manera individual.

Five Factor Model

Los defensores del modelo postulan que cualquier tipo de personalidad, normal o anormal, puede ser conceptualizada a partir de cinco amplias dimensiones ortogonales con las que se podría describir cómo es una persona: Extraversión, Responsabilidad, Amabilidad, Estabilidad emocional o Neurotismo, y Apertura a la experiencia (Feist y Feist 2009). Cada una de estas cinco dimensiones se categorizan a partir de la evaluación de un grupo de rasgos que explican la mayor parte de las diferencias entre las personalidades individuales, a pesar de las barreras del lenguaje o la cultura (McCrae y Costa 1985; Feldt et al. 2010).

Desde el momento en que empezó a utilizarse este modelo, se han realizado varias investigaciones que demuestran cómo el estudio de estas dimensiones puede utilizarse para predecir determinados comportamientos. Por ejemplo, se ha demostrado que el factor responsabilidad es el predictor general de la productividad en el trabajo, mientras que otras dimensiones se relacionan con aspectos específicos del rendimiento laboral. Por ejemplo, la amabilidad y estabilidad emocional predicen el rendimiento del trabajo en grupo, mientras que la extroversión predice el éxito en las ventas y puestos directivos (Muñoz 2014). Por otro lado, algunos hábitos y comportamientos descritos para los individuos con altos niveles de apertura son: la búsqueda de nuevos e interesantes restaurantes y cuestionar los valores tradicionales. En contraste el hábito de una persona con bajos niveles de apertura prefiere un lugar familiar, tienden a apoyar los valores tradicionales y a preservar un estilo de vida fijo (Feist y Feist 2009).

Los modelos de personalidad suelen utilizar test o instrumentos para medir el estudio de la personalidad. En el caso del modelo FFM no es propio de un instrumento en particular, en realidad es una descripción y explicación de la personalidad que posee diversos inventarios contruidos para su evaluación. Dentro de los instrumentos o test que existen para medir el modelo FFM se encuentran: Cuestionario Big Five de Caprara, Babaranelli y Borgogni, el Big Five Questionnaire (BFQ), The Personality Inventory NEO (NEO-PI) y el Sixteen Personality Factor Inventory (16PF) (Costa y McCrae 1992; Lingjaerde y O.Engvik 2001; Pérez et al. 2004).

En la Tabla 2 se realiza un análisis comparativo de los modelos psicológicos estudiados. Para la comparación fueron definidos un conjunto de atributos identificados en la literatura y que han sido parcialmente abordados: si son modelos libres; facilidad de aplicación teniendo en cuenta si requiere o no personal especializado para el análisis de los datos (ESPEC) y si está disponible en línea (Acceso); la existencia de instrumento de aplicación reconocido del modelo (Instrumento); el tiempo de ejecución del instrumento y la edad permitida para aplicar el instrumento; se consideró además el número de trabajos analizados que emplean algunos de estos modelos (Número Trabajos) y área de aplicación (Aplicación Informática).

Tabla 2: Análisis comparativo de modelos de la personalidad

Modelo/Test	Modelo Libre	Fácil Aplicación		Instrumento	Tiempo de Ejecución /min	Edad Aplicación	Número Trabajos	Aplicación Informática			
		ESPEC	Acceso					Definición de patrones de formación de equipos	Evaluación por competencias	Análisis diversidad/ efectividad	Evaluación de la personalidad
MBTI	No	Sí	Sí	MBTI	30-40	14	4	X		X	X
FFM	Sí	No	Sí	NEO-PI	45	16	5	X	X		X

Valorando esta comparación realizada en la Tabla 2, se concluye que el modelo FFM es el más factible para la investigación. En círculos científicos el FFM ha sido ampliamente utilizado en el ámbito del software según (Acuña 2002; Aluja y Blanch 2002; Acuña y Juristo 2004), no necesita de personal especializado para el análisis de sus datos. La información que proporciona su aplicación resulta valiosa para la evaluación de la personalidad a partir del análisis de los rasgos que tributan a su clasificación.

En la investigación se asumirá la utilización del instrumento NEO-PI, ya que además de medir las cinco dimensiones descritas en el modelo FFM, analiza también seis de los rasgos de la personalidad² específicos de cada dimensión. De manera general el instrumento analiza 30 rasgos que caracterizan a un individuo en una escala de: Bajo, Normal y Alto (CEDHAP 2014). El objetivo de este instrumento es identificar, a partir de auto-reporte reacciones fáciles de reconocer, las preferencias básicas de las personas en lo que se refiere a la percepción y el juicio (Costa y McCrae 1992).

² Tomado del modelo FFM seleccionado para la investigación.

NEO-PI entrega una descripción simple de la personalidad a través de los rasgos, destacando aspectos sobresalientes; y una extensión detallada y compleja que refiere a procesos cognitivos, bienestar psicológico, motivos, etc. (Herrerías 2005). Este test, presenta adecuados niveles de confiabilidad y posee también estudios que avalan su utilidad práctica y sus cimientos teóricos (Criado 2012).

El resultado de la aplicación del instrumento NEO-PI formará parte de la base de casos que se utilizará en el descubrimiento de conocimiento a partir de la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH.

Evaluación del rendimiento laboral de los RRHH en proyectos informáticos

El rendimiento de los empleados es la piedra angular para desarrollar la efectividad y el éxito de una institución, lo que motiva un constante interés de las empresas por mejorar el rendimiento de los empleados a través de continuos programas de capacitación y desarrollo.

Se define rendimiento según los autores (Moorhead y Griffin 1998; Pacheco 2012), como “(...) aquellas acciones o comportamientos observados en los empleados que son relevantes para los objetivos de la organización, que pueden ser medidos en términos de las competencias de cada individuo y su nivel de contribución a la empresa”. Por lo que la evaluación de este rendimiento sería un proceso para obtener un índice cuantitativo que muestre el nivel de desarrollo que tiene un individuo en uno o varios factores de rendimiento.

Autores como (Chiavenato 2001; Pacheco 2012) coinciden en que el rendimiento laboral es definido como el comportamiento de los empleados frente al trabajo, siendo imprescindible dentro de la gestión del personal por lo que es necesario medirlo, analizarlo, predecirlo y modificarlo. De esta manera se contribuye a la estrategia individual para lograr los objetivos de la organización. Es por esto que la manera más usada hoy día para estimar o apreciar el comportamiento de una persona en un centro laboral, es la evaluación del rendimiento (Pacheco 2012).

Como parte de la decisión de estandarizar los sistemas de control y seguimiento de la Red de Centros de desarrollo de software de la UCI, se decide desarrollar la herramienta GESPRO como solución integral para la gestión de sus proyectos (Piñero y et.al 2010). Basado en Redmine³ y otras herramientas libres construidas en la propia organización, el sistema está alineado con los estándares propuestos por (PMI 2009; SEI 2010), mostrando las interfaces adecuadas a los usuarios para que introduzcan los datos al sistema.

En (Lugo 2012) se utiliza esta herramienta en su versión 12.05, como fuente de datos de los procesos de planificación y ejecución de proyectos. En ella se propone la definición de un conjunto de indicadores asociados para la evaluación de proyectos informáticos que pueden ser calculados de modo automático. Estos indicadores brindan la información requerida para el control y seguimiento de los proyectos que se

³ Redmine: es una aplicación web flexible de gestión de proyectos, utiliza el framework Ruby on Rails y es multiplataforma. Disponible en <http://www.redmine.org>

gestionan con la herramienta. GESPRO posibilita la automatización del cálculo de los indicadores antes mencionados, dentro de estos se encuentran los indicadores propuestos para el área de RRHH.

Para el análisis y estudio de esta investigación se utiliza la propuesta de (Lugo 2012) referente al Índice de Rendimiento de los Recursos Humanos (IRRH), proponiendo el análisis del rendimiento del individuo en torno al desarrollo, impacto y prioridad de las tareas asignadas. Esta selección está fundamentada principalmente por la aplicabilidad que presentan estos indicadores en el entorno al que se refiere la investigación: los proyectos informáticos de la UCI.

De acuerdo a los planteado por (Lugo 2012), el indicador IRRH se calcula a partir de cuatro sub-indicadores: Índice de Correlación del Recurso Humano con respecto al Trabajo (IRHT), Índice de Rendimiento del Recurso Humano con respecto a la Eficacia (IRHE), Índice de Rendimiento del Recurso Humano con respecto a su Aprovechamiento (IRHA) y el Índice de Rendimiento del Recurso Humano con respecto a la Eficiencia (IRHF).

El cálculo de los cuatro sub-indicadores que permiten obtener el IRRH se realiza de la siguiente manera:

- 1) IRHT: correlación entre la estandarización de los tiempos estimados (TE_Tarea) y los tiempos dedicados (TD_Tarea) de las tareas involucradas con el recurso humano. Adquiere valores entre -1 y 1. Entre más cercano a 1, mejor.
- 2) IRHE: relación entre la cantidad de tareas cerradas evaluadas de Bien o Excelente (CTBERH) y el total de tareas cerradas del proyecto (CTC) involucradas con el recurso humano hasta la fecha de corte dd/mm/aaaa. Adquiere valores entre 0 y 1. Entre más cercano a 1, mejor.
- 3) IRHA: relación entre los totales de tiempo planificado (TTP) y disponible (TTD) asignados al recurso humano hasta la fecha de corte dd/mm/aaaa. Cuando los resultados toman valores: $IRHA < 0.75$, es sub-utilizado; $0.75 \leq IRHA \leq 1$, es aprovechado; $IRHA > 1$, es sobrecargado.
- 4) IRHF: muestra el estado de avance o progreso del recurso humano en la realización de las tareas que tiene asignadas. Relación entre la sumatoria de los por cientos de ejecución real y la sumatoria de los por cientos de ejecución planificado para cada una de sus tareas según la fecha de corte dd/mm/aaaa. Cuando los resultados toman valores: $IRHF < 1$, atraso; $IRHF = 0$, atraso por no haber comenzado; $IRHF = 1$, en tiempo; $IRHF > 1$, adelantado.

Los índices de rendimiento antes mencionados formarán parte del descubrimiento de conocimiento y complementarán la base de casos que se utilizará en la investigación.

Considerando que es necesario encontrar la relación entre los rasgos de la personalidad y el rendimiento laboral, se decide utilizar el descubrimiento de conocimiento y las técnicas que lo conforman como vía para obtener modelos predictivos que describan esta relación.

Descubrimiento de conocimiento

A medida que van creciendo las organizaciones y con ellas el volumen de datos e información que generan, se va haciendo más complejo el manejo de los mismos y más notable la incapacidad de los seres humanos de analizar y transformar la información en conocimiento útil para la toma de decisiones. En la actualidad, se hace cada vez más necesaria la utilización de técnicas y herramientas que permitan la extracción del conocimiento de los datos. Por tal motivo, desde comienzos de la década de los 90, las organizaciones que manejan grandes volúmenes de datos aplican lo que se denomina “descubrimiento de conocimiento en los datos” o “Knowledge Discovery in Databases” (KDD por sus siglas en inglés).

KDD es el proceso de extracción de información que se encarga de interpretar grandes cantidades de datos y encontrar relaciones o patrones. Las tareas comunes en KDD son la inducción de reglas, los problemas de clasificación y agrupamiento, el reconocimiento de patrones, el modelado predictivo, la detección de dependencias, entre otras (García y Molina 2006; Gervilla et al. 2009).

Otra definición de KDD según (Fayyad et al. 1996), es “(...) la extracción automatizada de conocimientos o patrones interesantes, no triviales, implícitos, previamente desconocidos, potencialmente útiles y predictivos de la información de grandes bases de datos”. Este proceso está compuesto por las fases de selección, preprocesamiento, transformación, minería de datos, y la evaluación e interpretación de los resultados (Fayyad, Shapiro et al. 1996).

En la Figura 2 se muestra la secuencia iterativa de etapas que componen el proceso de extracción del conocimiento.

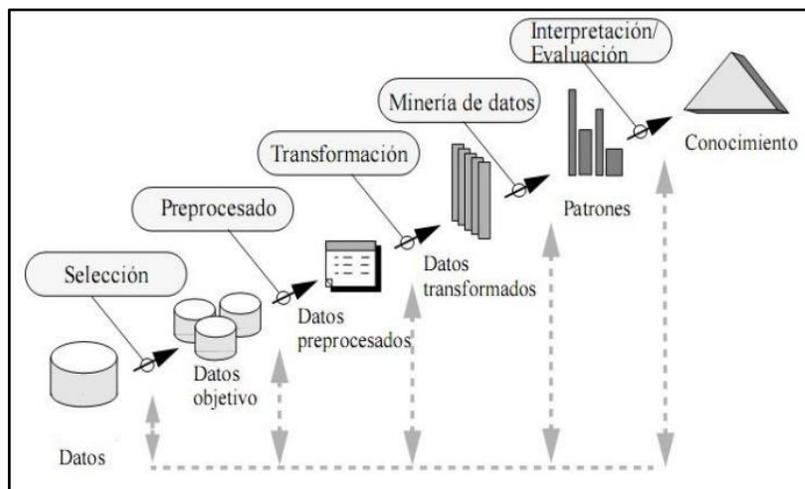


Figura 2: Etapas para el proceso de extracción del conocimiento. Tomado de (Fayyad, Shapiro et al. 1996)

A continuación se describen brevemente estas etapas teniendo en cuenta lo expresado por (Fayyad, Shapiro et al. 1996; Ril 2012):

- Selección: tiene como objetivo la preparación de las fuentes de datos y la selección de las mismas.

- Preprocesado y transformación: se aplican técnicas como limpieza de datos, la integración y transformación de los mismos, la reducción de ellos y la selección de atributos.
- Técnicas de Minería de Datos: se perfeccionan constantemente las técnicas y algoritmos que se encargan de extraer y representar el conocimiento de forma adecuada para la toma de decisiones. Se combinan técnicas potenciando las ventajas de cada una y atenuando sus debilidades.
- Interpretación y evaluación: se procede al análisis de los resultados descubiertos. Incluye a su vez la resolución de posibles inconsistencias con otros conocimientos anteriores a la investigación.

En la actualidad el descubrimiento de conocimiento se aplica en las organizaciones para convertir los datos en conocimiento utilizando técnicas de Minería de Datos. Al mismo tiempo se intenta derivar del conocimiento, resultados positivos para la organización.

Análisis de técnicas de Minería de Datos

La Minería de Datos, en el proceso de KDD desempeña un rol fundamental ya que, según (Fayyad, Shapiro et al. 1996) "(...) es la fase que consiste en aplicar análisis de los datos y algoritmos de descubrimiento para encontrar patrones o modelos sobre los datos". Las técnicas de la Minería de Datos provienen tanto de la Inteligencia Artificial como de la Estadística, siendo estas últimas las analizadas en la presente investigación.

En la Figura 3 se muestra un gráfico con técnicas de Minería de Datos clasificadas en dos categorías: supervisadas o predictivas y no supervisadas o descriptivas (Pacheco y Betancourt 2014).

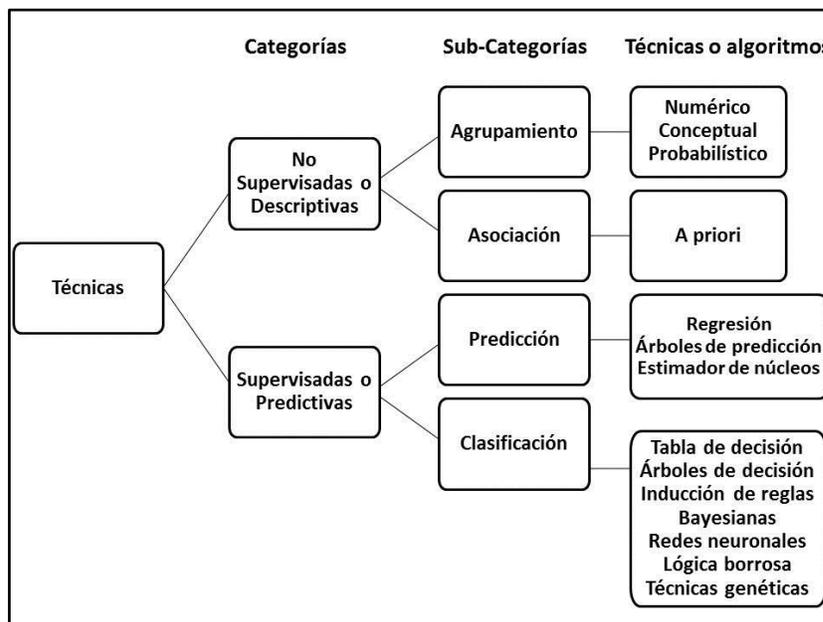


Figura 3: Técnicas de la Minería de Datos. Adaptado de (García y Molina 2006)

Como se muestra en la Figura 3, las técnicas de clasificación pertenecen a la primera categoría. Estas técnicas presentan relevante importancia para esta investigación y son abordadas más adelante.

Del conjunto de técnicas de Minería de Datos analizadas se determina la utilización de las técnicas de clasificación para identificar grupos mutuamente excluyentes a partir de la división de un conjunto de datos. De tal forma que cada miembro de un grupo esté lo más cerca posible de otros y grupos diferentes estén lo más lejos posible de otros, donde la distancia se mide con respecto a las variables especificadas que se quieren predecir. Esta técnica de Minería de Datos analiza un conjunto de datos cuya clasificación de clase se conoce y construye un modelo para cada clase que suele representarse con un árbol de decisión o reglas de clasificación que muestran las características de los datos (García y Molina 2006).

Técnicas de clasificación

A continuación se describen algunas de las técnicas de clasificación que a consideración de la autora se aplican al objetivo de la investigación. Estas técnicas permiten analizar un conjunto de datos y ayudar a la toma de decisiones.

- **Tabla de decisión:** esta técnica consiste en seleccionar subconjuntos de atributos y calcular su precisión para predecir o clasificar los ejemplos. Una vez seleccionado el mejor de los subconjuntos, la tabla de decisión estará formada por los atributos seleccionados (más la clase), en la que se insertarán todos los ejemplos de entrenamiento únicamente con el subconjunto de atributos elegido. Si hay dos ejemplos con exactamente los mismos pares atributo-valor para todos los atributos del subconjunto, la clase que se elija será la media de los ejemplos (en el caso de una clase numérica) o la que mayor probabilidad de aparición tenga (en el caso de una clase simbólica)(Martineaux 2008).
- **Árboles de decisión:** es un modelo de predicción utilizado en el ámbito de la IA. Es muy similar a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema. El aprendizaje de árboles de decisión suele ser más robusto frente al ruido y conceptualmente sencillo, pudiéndose interpretarse esencialmente como una serie de reglas de clasificación compactadas para su representación en forma de árbol (García y Molina 2006). Dado un conjunto de ejemplos, estructurados como vectores de pares ordenados atributo-valor, de acuerdo con el formato general en el aprendizaje inductivo a partir de ejemplos, el concepto que estos sistemas adquieren durante el proceso de aprendizaje consiste en un árbol. Una de las ventajas de los árboles de decisión es que representan las relaciones mediante reglas que son fáciles de interpretar, permitiendo representar los resultados de los árboles de decisión en lenguajes naturales (Jacas 2009). Se encuentran implementados en multitud de sistemas, muchos de ellos gratuitos, como: PSPP⁴, WEKA⁵ o RapidMiner⁶, entre otros.

⁴ Alternativa gratuita de código abierto para SPSS (Paquete estadístico para las ciencias sociales).

⁵ Waikato Environment for Knowledge Analysis. Herramienta de minería de datos de libre distribución.

⁶ Anteriormente conocido como YALE, Yet Another Learning Environment.

- Inducción de reglas: esta técnica permite la generación y entendimiento de árboles de decisión, o reglas y patrones a partir de los datos de entrada. La información de entrada será un conjunto de casos donde se ha asociado una clasificación o evaluación a un conjunto de variables o atributos. Con esa información estas técnicas obtienen reglas que soportan la evaluación o clasificación. En los casos en que la información de entrada posee algún tipo de "ruido" o defecto (insuficientes atributos o datos, atributos irrelevantes o errores u omisiones en los datos) estas técnicas pueden habilitar métodos estadísticos de tipo probabilístico para generar árboles de decisión recortados o podados. También en estos casos pueden identificar los atributos irrelevantes, la falta de atributos discriminantes o detectar brechas de conocimiento. Esta técnica suele llevar asociada una alta interacción con el analista de forma que este pueda intervenir en cada paso de la construcción de las reglas, para aceptarlas o para modificarlas (García y Molina 2006).

Realizando un análisis de estas técnicas de clasificación de acuerdo a sus características, se decide utilizar los árboles de decisión debido a que apoyan la elección entre varios cursos de acción, representando la dependencia lógica entre la decisión a tomar y los atributos considerados. Los resultados a través de la representación del conocimiento en árboles, a diferencia de las tablas de decisión, son fáciles de interpretar en posteriores análisis, de forma intuitiva y de fácil entendimiento para los seres humanos. La relación entre los árboles y las reglas de decisión, radica en elaborar reglas fácilmente interpretables, que definen las características que más diferencian a las distintas clases establecidas inicialmente (Colectivo de autores 2014). Igualmente predicen el valor de la variable llamada clase, en función de las variables de entradas diferentes y así apoyan la toma de decisiones.

En general, los clasificadores de árbol de decisión tienen buena precisión y han sido utilizados para la clasificación en áreas de aplicación como la Gestión de RRHH y la ayuda en la toma de decisiones en la Gestión de proyectos (Ril 2012; Silvente 2013; Colectivo de autores 2014). Además, los árboles de decisión son un método efectivo para la toma de decisiones por las siguientes razones: en el problema que se desea clasificar mediante esta técnica son analizadas todas las posibilidades, permite analizar totalmente el resultado para cada una de las posibles opciones y ayuda a tomar las mejores decisiones con base a la información existente y a las mejores suposiciones (Fayyad, Shapiro et al. 1996).

Análisis de algoritmos de clasificación de árboles de decisión

Para comprender el ámbito de los árboles de decisión se analizaron los algoritmos: ID3, C4.5, CHAID y J4.8. Estos algoritmos de forma automática y siguiendo el enfoque de aprendizaje supervisado, identifican patrones en datos para crear un modelo, en este caso un árbol. Seguidamente se realiza una breve descripción de ellos:

- ID3: es un algoritmo cuyo objetivo es la elaboración de un árbol de decisión como un método para aproximar una función objetivo de valores discretos. El algoritmo es resistente al ruido en los datos

eligiendo el mejor atributo a través del concepto de entropía⁷, cuanto menor sea el valor de la entropía, menor será la incertidumbre y más útil será el atributo para la clasificación (García y Molina 2006).

- C4.5: de acuerdo a la investigación realizada por (Martineaux 2008), es uno de los algoritmos más utilizados en el ámbito de los árboles de clasificación teniendo en cuenta que en cada nodo del árbol, elige el atributo que más eficazmente lo divide (Colectivo de autores 2014). Este algoritmo es una extensión del algoritmo ID3 mejorado, ya que trabaja con datos numéricos y simbólicos y manipula los casos con valores de atributo que faltan. El C4.5 construye árboles de clasificación de un conjunto de datos de la misma manera que el algoritmo ID3, usando el concepto de entropía (García y Molina 2006).
- CHAID: (en inglés, Chi-square automatic interaction detector) es un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, creando segmentos y perfiles con respecto al resultado deseado. Permite la detección automática de interacciones mediante Chi-cuadrado, el cual refleja cuan similares o asociadas están las variables (Silvente 2013).
- J4.8: es una implementación libre del C4.5 y está integrada en la plataforma WEKA. El parámetro más importante a tener en cuenta es el factor de confianza para la poda (confidence level), que influye en el tamaño y capacidad de predicción del árbol construido. Para cada operación de poda, define la probabilidad de error. A probabilidad menor, se exigirá que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar. El valor por defecto es de 25%, según baje este valor se permiten más operaciones de poda (García y Molina 2006; Betancourt y Pacheco 2014). Este algoritmo permite trabajar con valores continuos para los atributos, separando los posibles resultados en dos ramas $A_i \leq Z$ y $A_i > Z$; siendo Z un umbral escogido anteriormente. Cada caso $S_i = x_1, x_2, \dots$ es un vector, donde x_1, x_2, \dots representan los atributos o características de la muestra. Los datos de entrenamiento se ven aumentados con un vector $C = C_1, C_2, \dots$ donde C_1, C_2, \dots representan la clase a que pertenece. En cada nodo del árbol, por su semejanza con el algoritmo C4.5, elige el atributo que más eficazmente lo divide a partir de un conjunto de datos de entrenamiento. Su criterio es el de proporción o ratio de ganancia (gain ratio) que dependiendo de los resultados tomará la elección de un atributo para dividir los datos. De esta manera se consigue evitar que las variables con mayor número de categorías salgan beneficiadas en la selección. En el caso de la ganancia como en el de la proporción de ganancia, el mejor atributo para la división es aquel que la maximiza. El algoritmo luego seguirá recursivamente en los siguientes subnodos (Betancourt y Pacheco 2014).

⁷ Se trata de una medida que permite calcular el grado de incertidumbre de una muestra. Una muestra completamente homogénea tiene entropía 0.

Luego de presentar las características principales de los algoritmos anteriores, se decide compararlos teniendo en cuenta aspectos tales como: la variable predictora, el tipo y criterio de división, método de poda y código de implementación. Los resultados de esta comparación se muestran en la Tabla 3.

Tabla 3: Características de los algoritmos de clasificación de árboles de decisión estudiados. Adaptado de (et al, 2014)

Algoritmo	Variabes predictoras	Tipo de división	Criterio de división	Casos perdidos	Método de poda	Implementación
CHAID (1975)	Discretas	<i>n</i> -aria	χ^2	SI	Pre (nivel de significancia)	Comercial
ID3 (1979)	Discretas	<i>n</i> -aria	Entropía	NO	NO	Comercial
C4.5 (1993)	Continuas/ Discretas	Binaria/ <i>n</i> -aria	Entropía	SI	Pre-/ Post	Libre Comercial
J4.8 (1997)	Continuas/ Discretas	Binaria/ <i>n</i> -aria	Entropía	SI	Pre-/ Post	Libre (WEKA)

Analizando ventajas y deficiencias de cada uno respecto al tipo y el criterio de división, casos perdidos y disponibilidad del código, se destaca como el más factible para la presente investigación el algoritmo J4.8. La selección está sustentada principalmente en que este algoritmo se enfoca explícitamente en los atributos relevantes e ignora los irrelevantes. Además, permite generar árboles de decisión y a partir de ellos obtener reglas de decisión para la ayuda a la toma de decisiones. J4.8 permite el tratamiento de casos con información incompleta o perdida, es relativamente rápido mostrando fiabilidad en los resultados. Representa uno de los algoritmos de aprendizaje más comúnmente usados en aplicaciones de Minería de Datos (García y Molina 2006; Jacas 2009).

Análisis de herramientas de Minería de Datos

Un sistema o herramienta de Minería de Datos es definido como una aplicación integrada diseñada para ejecutar varios algoritmos y otras operaciones del proceso de KDD sobre un conjunto de datos bien gestionados por el propio sistema o accesibles de diversas fuentes de datos externas (Sánchez 2001).

Se estudiaron varios software comerciales y académicos que permiten realizar tareas de Minería de Datos e implementan el algoritmo J4.8 para descubrir el conocimiento en la investigación. Las herramientas estudiadas son: KNIME⁸, RapidMiner y WEKA. A continuación se realiza una breve descripción de cada una de ellas:

- KNIME: es una plataforma de fácil uso y comprensible para la integración de datos, procesamiento, análisis y exploración. Es un sistema totalmente gratuito para el desarrollo y ejecución de técnicas de Minería de Datos que permite la creación de modelos estadísticos y de Minería de Datos, como árboles de decisión y regresiones. La API⁹ del software permite el montaje fácil y rápido de nodos para procesamiento de datos (ETL: extracción, transformación, carga), para el análisis de datos, modelado y visualización (Berzal 2013; KNIME 2013).

⁸ Konstanz Information Miner.

⁹ Interfaz de programación para añadir nuevos algoritmos al sistema.

- **RapidMiner:** es una herramienta de código abierto implementada en Java, basada en el aprendizaje automático para el descubrimiento de conocimiento y la Minería de Datos. RapidMiner tiene implementado un amplio rango de algoritmos y técnicas de Minería de Datos. Tiene soporte API completo, lo que hace que sea posible acceder a una amplia variedad de funcionalidades y soporte. Es un software de tipo Open-Source¹⁰ con licencia GNU¹¹, trabaja bajo las plataformas Windows y Linux, usa el lenguaje de scripting XML para describir los operadores y su configuración (RapidMiner 2014).
- **WEKA:** es una plataforma de software libre desarrollada en Java, centrada en el aprendizaje automático y en la Minería de Datos. La funcionalidad de la API de WEKA proporciona a los usuarios la capacidad de los diversos componentes del software para comunicarse entre sí. El software contiene la capacidad de realizar más de 100 tipos de métodos de Minería de Datos, incluyendo los métodos de agrupamiento o generación de grupos de datos, clasificación, regresiones, visualización, selección de propiedades y análisis estadísticos. Incluye una interfaz gráfica de usuario para acceder y configurar las diferentes herramientas integradas. Su licencia GNU permite que cualquiera pueda modificar o acceder al código de WEKA. Una de las debilidades del software es su apoyo a la visualización. Aunque proporciona visualización de los datos, resultados y procesos, el soporte que se proporciona es limitado. El uso de WEKA a nivel de usuario no es muy complicado pero se requiere un conocimiento básico de aprendizaje automático y Minería de Datos. Además, como WEKA está programado en Java, es independiente de la arquitectura, por lo que funciona en cualquier plataforma con una máquina virtual Java disponible (Morate 2013; WEKA 2013).

A partir del estudio de los sistemas usados en el proceso de Minería de Datos, en la Tabla 4 se realizó un análisis comparativo teniendo en cuenta los siguientes aspectos: si la licencia es libre o no, si requiere de conocimientos sobre Minería de Datos para interpretar los resultados, si es multiplataforma, si contiene modelo de clasificación, si implementa árboles de decisión, específicamente el algoritmo J4.8 (implementación del C4.5), si presenta una interfaz amigable y si dispone o no de un API.

Tabla 4: Análisis comparativo de sistemas de Minería de Datos

Sistemas	Licencia libre	Requiere de conocimientos sobre Minería de Datos	Multiplataforma	Modelo de Clasificación	Implementa árboles de decisión	J4.8	Interfaz amigable	API
KNIME	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
RapidMiner	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
WEKA	Sí	Sí	Sí	Sí	Sí	Si	Sí	Sí

¹⁰ Software de código abierto.

¹¹ General Public License.

Al comparar las herramientas de Minería de Datos antes mencionadas, se evidencia que tanto RapidMiner, KNIME como WEKA son aplicaciones de código abierto y de libre distribución, implementan el algoritmo J4.8 y son multiplataforma. Uno de los problemas que los usuarios podrían encontrar en estas herramientas radica en las funciones de visualización, las mismas no son muy avanzadas. Esto significa que los usuarios simplemente quieren una salida de aspecto agradable resuelta fácilmente en una interfaz entendible. Además, los gráficos que se producen en dichas herramientas probablemente necesitarán ser cambiados y hechos visualmente más atractivos, sobre todo si el público objetivo son personas que no presentan conocimientos sobre la Minería de Datos o el análisis estadístico.

De acuerdo a las características de los sistemas antes descritos, se decide utilizar WEKA como la herramienta que regirá el proceso de KDD que se realizará en la investigación, ya que a pesar de los inconvenientes anteriormente descritos, presenta a su favor ventajas como (Broitman y Escobar 2007):

- Constituido por una serie de paquetes de código abierto con diferentes técnicas de preprocesado, clasificación, agrupamiento, asociación y visualización.
- Facilidad de aplicación y análisis de prestaciones cuando son aplicadas a los datos de entrada seleccionados.
- Es una aplicación de código abierto, lo que significa que no solo se puede hacer uso de los algoritmos implementados, sino también puede analizarse la implementación realizada de cada uno de ellos.
- Presenta una interfaz para la importación de datos desde archivos con formato ARFF (*.arff, del inglés Attribute Relation File Format), de instancias binarias serializadas, de archivos C4.5, de archivos separados por comas (*.csv), de archivos separados por tabulaciones utilizando el conversor que ofrece, de una URL o de una base de datos PostgreSQL, MySQL u Oracle mediante su JDBC.

WEKA implementa además un conjunto de técnicas de selección de variables mediante evaluadores con el objetivo de reducir la dimensionalidad¹² de los datos a procesar.

Selección de las variables

Los problemas de selección de variables (SV) han adquirido especial interés en diversas áreas de aplicación donde el número de variables para un problema de clasificación es elevado y muchas de ellas son redundantes o irrelevantes. La meta de la SV es encontrar un subconjunto óptimo de variables que minimice la pérdida de información, maximice la reducción de ruido y que debe ser más significativa para la clasificación, para dar así una mejor interpretación a los resultados. La estrategia para la SV más común consiste en evaluar cada variable individualmente y seleccionar aquellas variables que aportan información de mayor calidad. El paradigma de la selección de variables trata de cubrir tres aspectos: mejorar el error de predicción de los clasificadores, proporcionar clasificadores más rápidos y eficientes y facilitar la comprensión del proceso subyacente Generador de los Datos (Luján 2009).

¹² Tamaño del conjunto de datos.

Un procedimiento para la SV basa su funcionamiento en dos aspectos fundamentales: un criterio de selección y un procedimiento de búsqueda que proporcione buenos resultados. Existen diversos criterios de selección y estos se clasifican en: Filtros, que utilizan características generales de los datos para evaluar los atributos y funcionan de forma independiente de cualquier algoritmo de aprendizaje, dentro de estos se encuentran la selección de atributos teniendo en cuenta el ranking de importancia de estos; y Envolturas las cuales usan el desempeño de algún clasificador (algoritmo de aprendizaje) para determinar lo esperado de un subconjunto. De acuerdo a (Pérez 2007) estos criterios de selección son:

- Selección de atributos basado en la correlación (CFS por sus siglas en inglés): evalúa un subconjunto de atributos considerando la habilidad predictiva individual de cada variable, así como el grado de redundancia entre ellas. Se prefieren los subconjuntos de atributos que estén altamente correlacionados con la clase y tengan baja correlación entre ellos (Hall 2003).
- Evaluación de conjunto basada en la consistencia (CSE por sus siglas en inglés): estos métodos buscan combinaciones de atributos cuyos valores dividen los datos en subconjuntos que contienen una gran mayoría de clase única. Por lo general, la búsqueda está sesgada hacia los pequeños subconjuntos de características de clase alta consistencia (Hall 2003).
- Evaluación de conjunto en envoltura (WSE por sus siglas en inglés): evalúa los subconjuntos de atributos utilizando un clasificador. Emplea la validación cruzada para estimar la exactitud del esquema de aprendizaje en cada conjunto (Kohavi R y John 1997).

El criterio de selección a utilizar en esta investigación es CFS, de tal forma que se obtenga un subconjunto de rasgos de personalidad altamente correlacionados con los índices de rendimiento que en este caso son los atributos clases, intentando mantener un bajo nivel de correlación lineal entre ellos. Desde el punto de vista del diseño del clasificador, un algoritmo de selección de variables forma parte del proceso de clasificación, donde una vez aplicada la selección de variables continúa la aplicación de un clasificador estándar sobre el subconjunto escogido.

Análisis de sistemas basados en el conocimiento

Los métodos de aprendizaje y clasificación pueden ser organizados entre otros en Sistemas Basados en el Conocimiento (SBC). Existe variedad de SBC, diferenciados por la forma de representar el conocimiento y por la naturaleza de la máquina de inferencia asociada, ejemplo de ellos son: sistemas basados en probabilidades (SBP), en reglas (SBR), en marcos (SBF) y sistemas basados en casos (RBC) (Gutiérrez 2003). Por su importancia para este trabajo se abordan brevemente los sistemas basados en casos.

Sistemas Basados en Casos (RBC)

Estos sistemas basan sus potencialidades para el aprendizaje en los mecanismos de detección de similitudes entre casos y también en la propiedad de incorporar de forma natural un nuevo conocimiento, mejorando gradualmente su funcionamiento. Estos sistemas se utilizan en el proceso de resolución de

nuevos problemas sobre la base de las soluciones de problemas similares del pasado. Su principal limitación radica en la capacidad de representación de los valores de los rasgos, aspecto que dificulta su aplicación en situaciones que requieren representaciones de casos complejas como las que se presentan cuando existe un alto grado de interrelación entre los rasgos (Gutiérrez 2003; Jiménez 2012).

Sus componentes fundamentales son la base de casos, el módulo de recuperación de casos y el módulo de adaptación de las soluciones. La forma de representar y almacenar la información se realiza a través de casos.

Un caso es constituido por dos elementos generales: los rasgos predictores y los rasgos objetivos. La especificación de los primeros dará lugar a la salida o solución del problema, representado por el segundo. Los casos pueden representar distintos tipos de conocimiento y pueden ser almacenados en diferentes formatos. La información almacenada en un caso, debe ser importante, tanto para el propósito del sistema, como para asegurar que siempre será elegido el caso más apropiado para solucionar un nuevo problema (Sierra 2013). El caso representa situaciones experimentadas previamente y se distingue de otras formas de conocimiento o experiencia porque son ejemplos específicos de hechos. El caso debe contener la descripción del problema o situación; una solución, es decir, información que no pueda ser derivada a partir del nivel actual de conocimiento y por último, información acerca de si la solución fue correcta o incorrecta (Jiménez 2012).

La base de casos es el módulo encargado de almacenar y organizar todos los casos disponibles y su estructura es crucial para la fase de recuperación de casos similares. En la creación de la estructura que formará la base de casos hay que tener en cuenta aspectos fundamentales: el tamaño que se aspira que tenga la base de casos, los requerimientos que se imponen en el dominio específico en el que se trabaja y la inserción eficiente de nuevos casos (Bello et al. 2002; Pérez 2002).

Tradicionalmente se han propuesto dos estructuras principales para el almacenamiento de casos (Sierra 2013):

- Memoria plana: en esta estructura se presentan los casos completos de forma secuencial utilizando algún tipo de indexación ya sea manual o automática. Los casos se almacenan secuencialmente en una lista simple, un arreglo o un fichero. Para lograr una recuperación eficiente, se indexan los casos de la base. Los índices se eligen para representar los aspectos importantes del caso y la recuperación involucra comparar las características consultadas con cada caso de la base de casos. Esta estructura tiene el principal inconveniente de que la búsqueda de casos es menos eficiente, por lo que no es recomendable utilizarla en sistemas que trabajan con una base de casos de gran tamaño y necesitan respuesta en tiempo real. Por otra parte la inserción de nuevos casos en este tipo de estructuras es muy sencilla ya que basta con incluir un nuevo registro con el nuevo caso respetando el método de indexación que se esté utilizando.
- Memoria jerárquica: en una estructura con memoria jerárquica se utilizan representaciones en forma de árbol, en los que cada nodo interior representa un atributo del caso y en las hojas se almacenan las

soluciones a los mismos. Cada recorrido desde la raíz hasta una de las hojas del árbol representa un caso completo. La ventaja de este tipo de almacenamiento es la eficiencia en la búsqueda pero a cambio se sacrifica la sencillez de inserción de nuevos casos.

Para realizar la selección de una de las estructuras anteriores para crear la base de casos, se debe realizar un consenso entre la eficiencia de recuperación y eficiencia de inserción. De la correcta elección de la estructura dependerá en gran medida la eficiencia global del funcionamiento del sistema. Por tal motivo, en la presente investigación se decide realizar el almacenamiento de casos mediante una memoria plana. La selección de esta estructura facilitará la inserción de nuevos casos a la base siempre que sea necesario aumentar el conjunto de datos de entrenamiento a utilizar para la clasificación.

Los casos que se almacenarán en la base de casos de la presente investigación serán clasificados aplicando el algoritmo de árboles de decisión J4.8 mediante la herramienta WEKA para descubrir conocimiento a partir de la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH en proyectos informáticos.

Conclusiones del capítulo

Como resultado del estudio realizado en el presente capítulo referente a los rasgos de la personalidad, la evaluación del rendimiento de los RRHH en proyectos informáticos, el descubrimiento del conocimiento y las técnicas de Minería de Datos se concluye lo siguiente:

- Para descubrir adecuadamente el conocimiento a partir de la relación existente entre los rasgos de la personalidad y el rendimiento laboral de los RRHH, se decide emplear técnicas de Minería de Datos.
- Del análisis de la concepción teórica para el descubrimiento de conocimiento, debe considerarse:
 - La identificación de los rasgos de la personalidad utilizando el instrumento de medición de la personalidad NEO-PI propuesto por el modelo FFM.
 - La evaluación de los índices de rendimiento laboral de los RRHH a través de los indicadores respecto a la eficiencia, eficacia, el trabajo y aprovechamiento implementados en la herramienta GESPRO.
 - La confección de una base de casos con estructura de memoria plana y que contenga los rasgos de la personalidad y los índices de rendimiento laboral de los RRHH que formarán el conjunto de datos de entrenamiento.
 - La clasificación del conjunto de datos de entrenamiento a partir de la aplicación del algoritmo árboles de decisión J4.8 y la selección de variables mediante el evaluador CFS.

CAPÍTULO 2: DESCUBRIMIENTO DE CONOCIMIENTO A PARTIR DE LA RELACIÓN RASGOS-RENDIMIENTO EN PROYECTOS INFORMÁTICOS

En el presente capítulo se define un proceso para descubrir conocimiento a partir de la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH en proyectos informáticos. Para su construcción se tomaron en cuenta las etapas de extracción del conocimiento propuestas por (Fayyad, Shapiro et al. 1996) descritas en el Capítulo 1.

Se definen como etapas del proceso para el descubrimiento de conocimiento en la investigación: Selección, Preprocesado, Transformación, Minería de los datos, Interpretación y Evaluación de los resultados obtenidos, agrupando estas etapas en 3 subprocesos. Se analiza el conocimiento que se descubre mediante la identificación de los rasgos relevantes de la personalidad y se generan un conjunto de reglas de decisión. En la definición de cada uno de los subprocesos y etapas se materializa su aplicación y se ofrecen los resultados.

Proceso de descubrimiento de conocimiento a partir de la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH

Existe una amplia variedad de definiciones referidas a un proceso, a pesar de sus diferencias, en su generalidad tienen en común que un proceso está compuesto por fases y que se relacionan entre ellas con el fin de cumplir un objetivo (RAE 2001).

En el marco de esta investigación se utilizará la definición de proceso que propone (PMI 2009), el cual plantea que un proceso: “(...) es un conjunto de acciones y actividades interrelacionadas que se llevan a cabo para alcanzar un conjunto previamente especificado de productos, resultados o servicios”.

La Figura 4 ofrece un resumen general de las interacciones y del flujo básico entre los subprocesos definidos en la investigación. Un subproceso, incluye las etapas que están vinculadas por las respectivas entradas y salidas. Por lo general, el resultado o salida de una etapa se convierte en la entrada de otra.

El proceso propuesto organiza la extracción de conocimiento en tres subprocesos, representados en la Figura 4:

- 1) Subproceso Base de Casos: en este subproceso se seleccionan las fuentes de datos a utilizar y se confecciona la base de casos.
- 2) Subproceso Minería de Datos: en este subproceso se realiza la transformación de los datos que fueron seleccionados y se aplican técnicas de Minería de Datos.
- 3) Subproceso Descubrimiento: es este subproceso se interpretan y evalúan los resultados obtenidos en el proceso de descubrimiento de conocimiento.

Para especificar las actividades de cada subproceso se referenciaron las etapas del proceso de extracción del conocimiento en bases de datos descritas por (Fayyad, Shapiro et al. 1996).

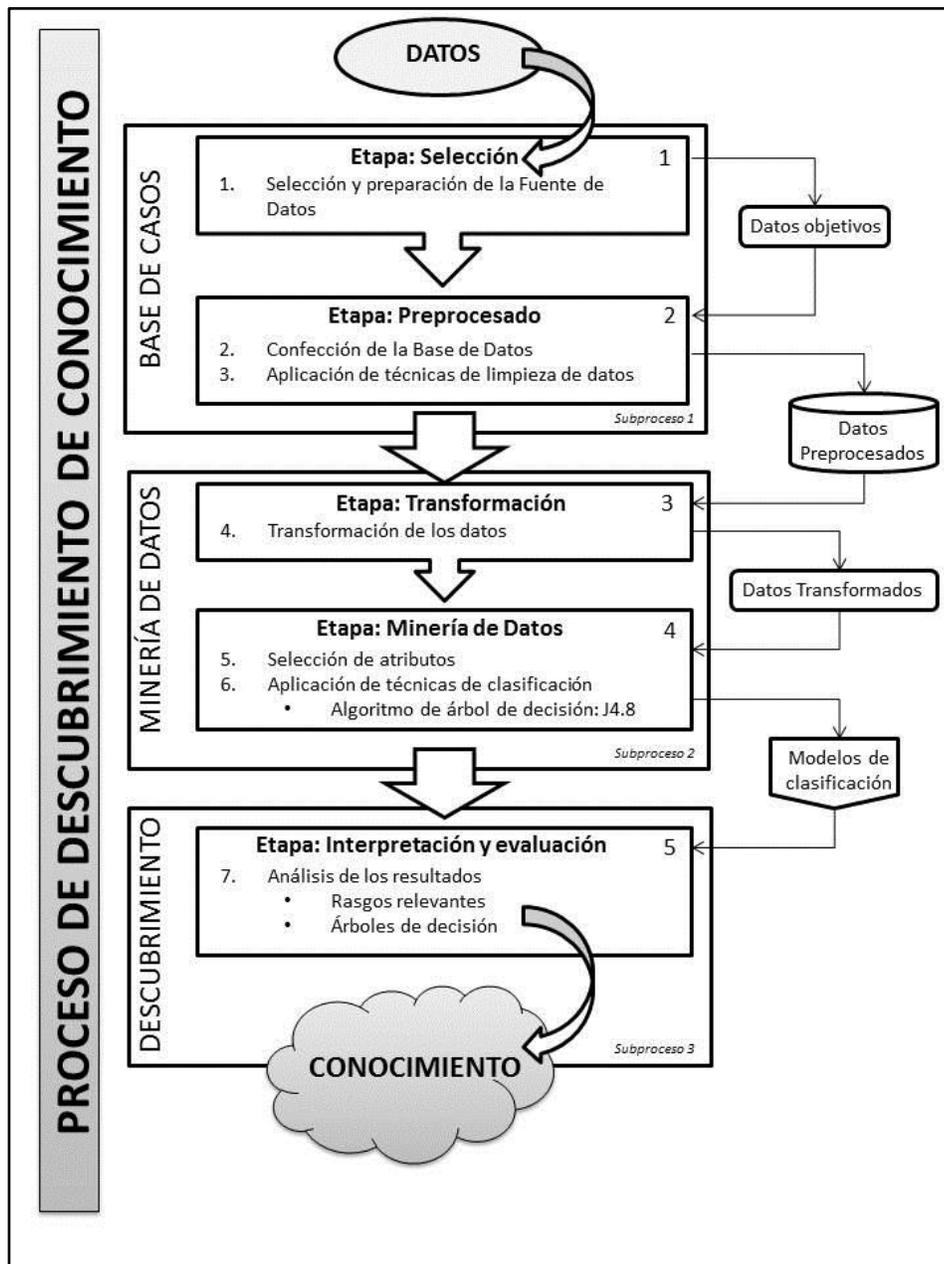


Figura 4: Representación del proceso de descubrimiento de conocimiento propuesto

Esta propuesta puede ser aplicable a la mayoría de los proyectos informáticos. Su objetivo es apoyar la toma de decisiones en relación a los procesos de la Gestión de RRHH a partir de evidencias individuales, con la finalidad de aumentar las posibilidades de que los individuos que componen equipos alcancen altos niveles de rendimiento laboral en su entorno productivo.

Subproceso Base de Casos

El subproceso Base de Casos define la selección de los datos y su preprocesamiento como parte de la extracción del conocimiento. Las actividades llevadas a cabo en este subproceso se muestran en la Figura 5 y se describen a continuación.

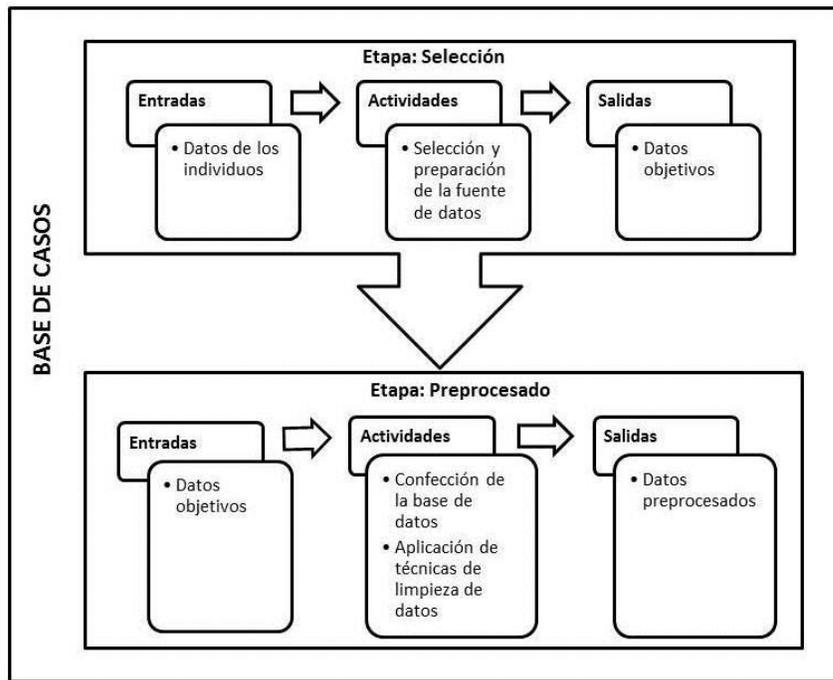


Figura 5: Descripción del subproceso Base de Casos

Etapa 1: Selección de los datos

Como paso inicial para la extracción del conocimiento se deben seleccionar los datos. Por esta razón, estos serán la entrada base al proceso de descubrimiento y por ende a esta primera etapa de Selección. Esta etapa tiene como objetivo la preparación de las fuentes de datos y la selección de las mismas.

Actividad 1: Selección y preparación de la Fuente de Datos

Esta actividad parte del análisis del conjunto de entradas, en este caso los datos de los individuos que formarán parte del descubrimiento. Cada persona va a constituir un caso y cada caso va a estar formado por los elementos: identificador según GESPRO, nombre de la persona y género.

Mediante el proceso que se propone en este capítulo se realizó el análisis de los datos provenientes de dos fuentes fundamentalmente: resultados de la autoevaluación de la personalidad caracterizada a través de rasgos mediante el instrumento NEO-PI y evidencias recopiladas en función de los resultados del rendimiento laboral de los RRHH, obtenidos de la herramienta GESPRO en el período 2011-2013.

Las evidencias mencionadas son variables en el tiempo, dependen del rol ejercido, el tipo de actividad desarrollada, el proyecto en que se encuentre, el equipo al cual pertenezca y otros factores como el nivel de experticia en la actividad laboral y el desarrollo de las habilidades. Por tanto, en cada actualización de los datos objetivos se podrá almacenar el comportamiento del rendimiento laboral de una persona en diferentes momentos. Esta representación permite realizar inferencias sobre los casos almacenados en función de identificar conocimiento como apoyo a la toma de decisiones en el proceso de Gestión de RRHH en proyectos informáticos de la UCI.

Para obtener como salida los datos objetivos inicialmente se realizaron varias consultas SQL a la base de datos del GESPRO. La búsqueda se centró en los proyectos del CEIGE para obtener la información referente al desempeño de las personas que componen la muestra en torno al desarrollo, impacto y prioridad de las tareas asignadas. Estos datos permitieron calcular numéricamente a través de la propuesta de (Lugo 2012), los cuatro índices de rendimiento utilizados en la investigación: IRHT, IRHE, IRHA, IRHF.

En seis de los proyectos del CEIGE fueron seleccionadas 200 personas, indistintamente al sexo que presentan y al rol desempeñado. Para complementar los datos objetivos se aplicó a ese conjunto de personas seleccionadas el instrumento NEO-PI definido en el modelo FFM para evaluar los rasgos de la personalidad que presentaban. De acuerdo al análisis de las fuentes de datos utilizadas y a la definición abordada según el modelo FFM, el instrumento NEO-PI clasifica los rasgos de la personalidad evaluados en 5 factores. Esta clasificación puede observarse en la Figura 6.

Neuroticismo	Extraversión	Apertura a la experiencia
Ansiedad	Amigabilidad	Imaginación
Ira	Gregarismo	Intereses artísticos
Depresión	Asertividad	Sensibilidad emocional
Timidez	Actividad	Aventurero
Impulsividad	Búsqueda de emociones	Curiosidad intelectual
Vulnerabilidad	Emoción positiva	Liberalismo
Amigabilidad		Responsabilidad
Confianza		Autoeficacia
Valores éticos		Orden
Altruismo		Obediencia
Cooperación		Capacidad de logro
Modestia		Autodisciplina
Simpatía		Prudencia

Figura 6: Rasgos de personalidad definidos por el FFM con el instrumento NEO-PI

Para lograr un resultado lo más completo posible y dejar a un lado las diferencias de género, el test fue respondido por 129 mujeres y 121 hombres, agrupados en los siguientes roles:

- Programador
- Analista
- Jefe de equipo
- Probador
- Director de centro
- Jefe de proyecto
- Administrador de BD
- Administrador de la calidad
- Diseñador de sistema
- Implantador de soluciones
- Planificador

Estos datos objetivos constituyen las salidas de esta actividad y las entradas de la etapa de Preprocesado.

Etapa 2: Preprocesado

La etapa Preprocesado contribuye a garantizar la calidad de la información sobre la que se pretende extraer conocimiento antes de llegar a aplicar las técnicas de Minería de Datos. Esto se debe a que entre mayor sea esta calidad, mayor será la calidad de los modelos de clasificación generados a partir de dicha información y que pueden utilizarse en la toma de decisiones. En este sentido, la obtención de información útil para ser posteriormente procesada es un factor clave.

En esta etapa se estudian los datos seleccionados y aplican técnicas de limpieza para detectar errores de integración, estandarizar datos, hacer agrupaciones, etc. (García y Molina 2006; González et al. 2007). Se filtran los datos de forma que se eliminen valores incorrectos, no válidos o desconocidos, según las necesidades y las técnicas estadísticas a utilizar y se obtienen muestras de los mismos en busca de una mayor velocidad de respuesta del proceso. Dentro de esta etapa se deben seleccionar las variables más influyentes en el problema de la presente investigación, así como los atributos relevantes. Además se deben eliminar o corregir los datos incorrectos y a su vez seleccionar una estrategia para los datos incompletos.

Actividad 2: Confección de la Base de Casos

Una vez identificados los datos objetivos (rasgos de la personalidad y rendimiento laboral, que constituyen las entradas a esta actividad) se construye la base de casos designando un vector de persona por cada caso. Esta base de casos forma parte de los resultados de la investigación.

La representación de un elemento en un vector se hará mediante el identificador del vector seguido de los índices entre corchetes, paréntesis o llaves (García y Molina 2006).

La confección del vector utilizado en esta propuesta presenta la siguiente estructura: *Persona [rasgo1, rasgo2, rasgo3... rasgoN, IRHT, IRHE, IRHA, IRHF]*, donde *IRHT, IRHE, IRHA e IRHF* son los indicadores de rendimiento utilizados en esta investigación. Un ejemplo de este vector serían los valores de: *Persona1 [Timidez Social, Apertura Experiencia, Confianza, Responsabilidad, Valores éticos....etc., IRHT, IRHE, IRHA, IRHF]*.

La base de casos quedó estructurada como se muestra en la Figura 7 y se decide almacenarla en un archivo con formato CSV, específicamente con extensión *.xls o *.csv.

ID	Rasgo 1	Rasgo 2	...	Rasgo n	IRHT	IRHF	IRHE	IRHA
1	Alto	Normal	Bajo	Normal	Medio	En_tiempo	Alto	Aprovechado

Figura 7: Estructura de la base de casos en hoja de cálculo Excel

En la elaboración de la base de casos se debe revisar el nivel de calidad de los datos que la conforman. El trabajo inadecuado con los sistemas de gestión de información y la unión de datos de distintas fuentes pueden generar anomalías en los datos. Estas anomalías suelen provocar inferencias erróneas sobre los datos almacenados y por tanto que se tomen decisiones incorrectas. Para evitar este tipo de problemas la

autora de esta investigación realiza técnicas de limpieza de datos como parte de la etapa de Preprocesado (Pacheco y Betancourt 2014).

Actividad 3: Aplicación de técnicas de limpieza de datos

Aunque idealmente, los datos almacenados no deberían contener errores, es casi inevitable que existan y merecen toda la atención para poder hacer inferencias válidas y tomar decisiones acertadas (Amón 2009).

En este sentido, la presente investigación adopta la taxonomía que propone Kim basada en la premisa de que los datos sucios se manifiestan en tres formas diferentes: los datos faltantes; los datos que no faltan pero su valor está mal; y los datos que no faltan y no están mal, pero que no se pueden utilizar (Kim et al. 2003). En correspondencia con la taxonomía adoptada y para limpiar la base de casos creada en la investigación se utilizaron métodos estadísticos con el objetivo de eliminar datos duplicados. A su vez, se utilizaron métodos de imputación para los valores ausentes, evitando que cambien las características más importantes de los valores presentes, cuando son incluidos los valores imputados (Amón 2010; López 2011).

El proceso de limpieza de datos se realizó dividiendo la base de casos en subconjuntos de 50 personas, con el fin de analizar con mayor detenimiento la información referente a cada caso. Tras realizar un proceso de limpieza de los datos no se detectó duplicación de los mismos ni pérdida de información. Por tal motivo no decrecieron los casos manteniéndose el tamaño original de la muestra. Como salida de esta actividad se obtiene la base de casos con los datos preprocesados.

Subproceso Minería de datos

El subproceso Minería de datos propuesto abarca las etapas de Transformación y Minería de Datos en un proceso KDD. Las actividades realizadas en este subproceso se representan en la Figura 8 y se describen a continuación.

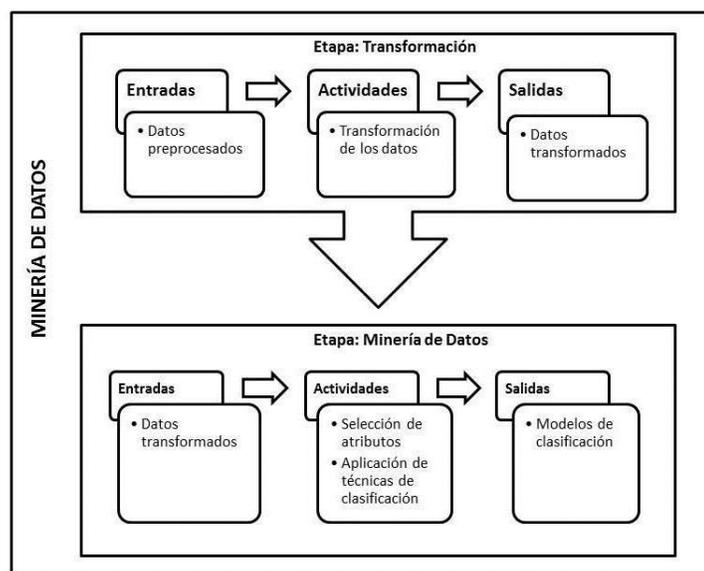


Figura 8: Descripción del subproceso Minería de Datos

Etapa 3: Transformación

La etapa de Transformación en la extracción del conocimiento se encuentra estrechamente relacionada con la etapa de Preprocesado. Esto se debe a que para transformar los datos y aplicar técnicas en el subproceso Minería de Datos, deben haberse preprocesado los datos anteriormente. La actividad que se realiza en esta etapa es la transformación de los datos.

Actividad 4: Transformación de los datos

Existen diversas formas de transformación de los datos (Broitman y Escobar 2007). Las asumidas en esta propuesta consisten en transformar los atributos de las variables en formatos específicos y la transformación de la base de casos en formatos soportados por la herramienta a utilizar en la próxima etapa de Minería de Datos.

En esta actividad se transforman los datos a un formato común, utilizando un libro de cálculo o haciendo uso de un programa estadístico. Además, se transforman los atributos de las variables que así lo requieran para facilitar el uso de técnicas que necesiten tipos de datos específicos.

En esta investigación se aplican las técnicas del aprendizaje automatizado en la generación de reglas de decisión a partir de los árboles, utilizando una base de casos con datos numéricos y simbólicos.

Los índices de rendimiento definidos en (Lugo 2012) pueden ser expresados cualitativamente como se muestra en la Tabla 5.

Tabla 5: Transformación de valores numéricos por simbólicos en los índices IRHE, IRHF, IRHA e IRHT

Transformación de los datos IRHE		Transformación de los datos IRHT	
Escala Cuantitativa	Escala Cualitativa	Escala Cuantitativa	Escala Cualitativa
0.75 a 1	Alto	0.34 a 1	Alto
0.50 a 0.74	Medio	-0.33 a 0.33	Medio
0 a 0.49	Bajo	-1 a -0.34	Bajo
Transformación de los datos IRHA		Transformación de los datos IRHF	
$IRHA < 0.75$	Sub-utilizado	$IRHF > 1$	Adelantado
$0.75 \leq IRHA \leq 1$	Aprovechado	$IRHF = 1$	En tiempo
$IRHA > 1$	Sobrecargado	$IRHF = 0$	Atraso por no haber comenzado
		$IRHF < 1$	Atraso

Otro tipo de transformación realizada a todos los datos preprocesados fue convertir la base de casos a la estructura idónea y requerida por la herramienta WEKA, donde los datos de entrada deberán estar codificados en formato ARFF (extensión "arff"). Este requisito no funcional es indispensable para trabajar con esta herramienta.

El fichero de datos, objeto de análisis en este descubrimiento, contiene información correspondiente a 200 personas seleccionadas aleatoriamente de la población. El fichero está estructurado por el nombre del mismo, seguidamente los atributos con los posibles valores que puede tomar (en este caso los atributos son nominales), generalmente el último atributo es el que posee la clase en la que se desea

clasificar. En lo adelante, entiéndase por clase los índices de rendimiento utilizados. Por último, se detallan los datos obtenidos para entrenar o probar el clasificador.

Por tanto, puede comprobarse en la Figura 9 que la cabecera del fichero de datos, "persona.arff", sigue el formato mencionado anteriormente.

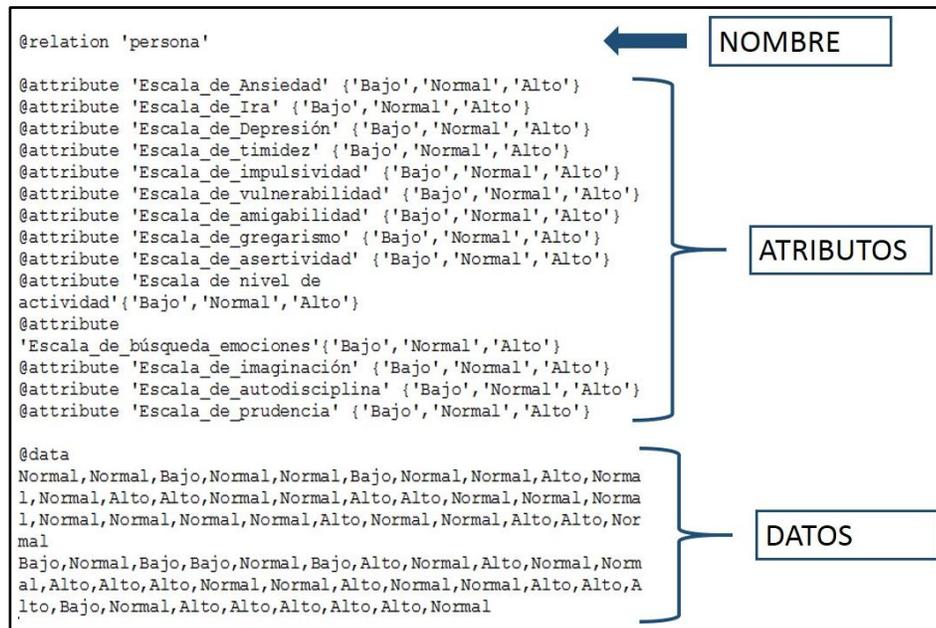


Figura 9: Estructura del fichero de datos persona.arff

Etapa 4: Minería de Datos

El minado de datos, en el proceso de KDD desempeña un rol fundamental ya que, según (Fayyad, Shapiro et al. 1996) "(...) es la fase que consiste en aplicar análisis de los datos y algoritmos de descubrimiento para encontrar patrones o modelos sobre los datos". Durante esta etapa, se aplican las diferentes técnicas de extracción de conocimiento, a los datos que ya han sido preprocesados y transformados en etapas anteriores (Cano de Amo 2004). En esta etapa se deben seleccionar los atributos relevantes para el problema y aplicar la técnica de clasificación seleccionada.

Actividad 5: Selección de atributos

Esta actividad se realiza aplicando técnicas de reducción de la dimensionalidad de los datos a procesar. En WEKA el criterio de selección utilizado es CFS, de tal forma que se obtiene, mediante la reducción, un subconjunto de rasgos de personalidad (rasgos relevantes). Estos rasgos están altamente correlacionados con los índices de rendimiento que en este caso son los atributos clases, intentando mantener un bajo nivel de correlación lineal entre ellos.

La selección de atributos relevantes realizada para las cuatro clases arrojó un subconjunto de atributos considerando la habilidad predictiva individual de cada variable, así como el grado de redundancia entre ellas. Se prefiere en la selección los subconjuntos de atributos (rasgos de la personalidad) que estén altamente correlacionados con la clase (índice de rendimiento) y tengan baja correlación entre ellos. En la

clase IRHE fueron seleccionados como atributos relevantes 7 rasgos representando el 23 % del total. En el caso de las clases IRHA, IRHF e IRHT fueron seleccionados 11, 8 y 8 rasgos relevantes respectivamente, representando el 37 % y el 27 % del total de rasgos evaluados.

Los resultados de la selección para estas clases expresan una baja correspondencia entre los rasgos seleccionados para cada una de ellas. Por otro lado se obtuvo un nivel de coincidencia de un 45 %. A pesar de que este dato no es alto, se concluye que no influye en la investigación debido a que para cada una de las clases existen rasgos que tienen mayor correlación que otros y no es necesario que coincidan entre ellos.

En la Tabla 6 se muestran los resultados generales tras el análisis de la selección para las cuatro clases. Los resultados demuestran que los rasgos relevantes para las cuatro clases son 22, representando un 73 % del total de rasgos del subconjunto evaluado. También se aprecia que para la clase IRHE los rasgos seleccionados pertenecen mayormente al factor de la personalidad Apertura a la Experiencia. De este análisis se interpreta que los rasgos que más influyen en el rendimiento laboral con respecto a la eficiencia son los que se agrupan en el factor antes mencionado.

En el caso de la clase IRHA la selección expresa que los rasgos que mayor correlación presentan con el índice de rendimiento con respecto al aprovechamiento son los que pertenecen al factor Responsabilidad. Sin embargo, con respecto a la eficacia en el estado de avance o progreso de un individuo en la realización de las tareas que tiene asignadas, se evidencia que influyen los rasgos pertenecientes a los factores Apertura a la Experiencia y a la Amigabilidad. Los factores que mayor correlación presentan en sus rasgos con respecto al índice de rendimiento referente al trabajo son la Responsabilidad y la Extraversión.

Tabla 6: Selección de los atributos relevantes para cada clase a través de la técnica CFS

Rasgos más relevantes	IRHA	IRHE	IRHF	IRHT
Escala de Ansiedad	x			
Escala de Depresión	x			
Escala de Amigabilidad	x			
Escala de Nivel de actividad	x		x	x
Escala de Valores éticos	x			
Escala de Imaginación		x		x
Escala de Aventuras				x
Escala de Liberalismo	x		x	
Escala de Confianza	x		x	
Escala de Simpatía	x		x	
Escala de Orden	x			x
Escala de Prudencia				x
Escala de Impulsividad		x		
Escala de Asertividad		x	x	
Escala de Obediencia			x	
Escala de Autodisciplina	x		x	x
Escala de Búsqueda de emociones		x		x
Escala de Cooperación				x
Escala de Intereses artísticos	x		x	
Escala de Sensibilidad emocional		x		
Escala de Curiosidad intelectual		x		
Escala de Modestia		x		
Total	11	7	8	8

En la Figura 10 se muestran los rasgos relevantes y la cantidad de índices que evalúan estos rasgos.

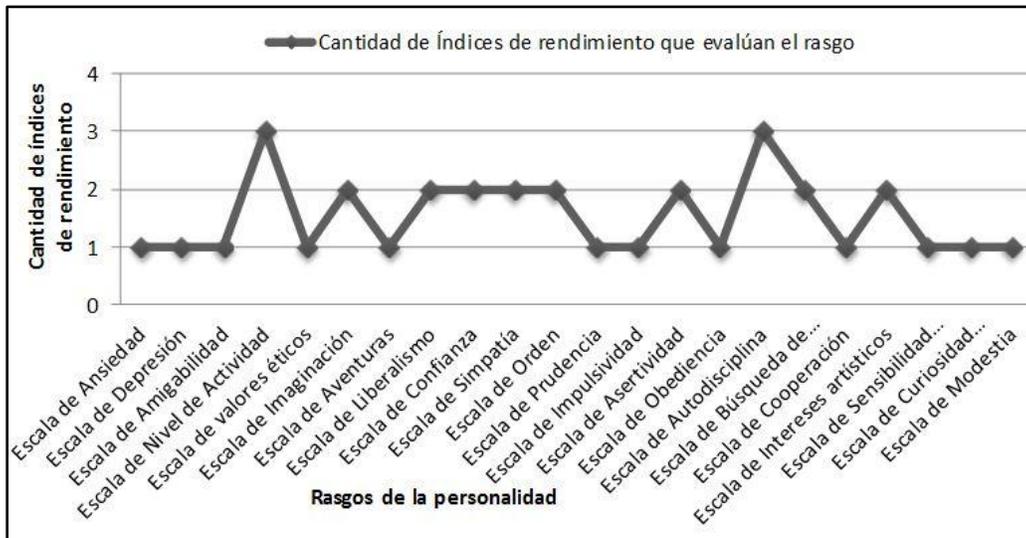


Figura 10: Gráfica de rasgos relevantes a partir de la selección de atributos

Los rasgos que más prevalecen en al menos dos de los cuatro índices de rendimiento son: escala de nivel de actividad, de imaginación, de liberalismo, de confianza, de simpatía, de orden, de asertividad, de autodisciplina, de búsqueda de emociones y escala de intereses artísticos.

La selección de los rasgos relevantes constituye la salida de esta actividad.

Actividad 6: Aplicación de técnicas de clasificación

Desde el punto de vista del diseño del clasificador, un algoritmo de selección de variables forma parte del proceso de clasificación, donde una vez aplicada la selección de variables continúa la aplicación de un clasificador estándar sobre el subconjunto escogido.

En la etapa Minería de Datos, tras emplear la técnica CFS para la selección de atributos, se procede a aplicar el clasificador sobre los datos seleccionados. En este sentido, la herramienta WEKA presenta una amigable e intuitiva interfaz visual, lo que permite que el usuario cuente con diferentes facilidades en su trabajo con la misma, independientemente del problema de clasificación que se presente (clasificación no supervisada o la clasificación supervisada).

La investigación se encuentra ante un problema de clasificación supervisada, es decir, según (Martineaux 2008; Corso 2009; Pacheco y Betancourt 2014) se parte de un conjunto de elementos descritos por un conjunto de características y se conoce la clase a la cual pertenecen estos elementos. A este concepto se le suele denominar conjunto de datos de entrenamiento o conjunto de aprendizaje.

Para aplicar la clasificación mediante el algoritmo de árboles de decisión (J4.8), se debe elegir primeramente el test a utilizar. En este caso se empleó el test Use training set. Este test evalúa el clasificador sobre el mismo conjunto del que se construye el modelo predictivo para determinar el error, que en este caso se denomina "error de resustitución" (Corso 2010).

El algoritmo J4.8, según (Betancourt y Pacheco 2014) emplea el concepto de entropía de la información para la construcción de árboles de clasificación de un conjunto de entrenamiento y la selección de variables que mejor clasifiquen a la variable (clase) estudiada.

Árboles de decisión

Una vez utilizado el clasificador los resultados se muestran en un modelo de clasificación. Este modelo muestra el árbol de decisión generado para la clase elegida. Estos árboles sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, o sea, generar las reglas de decisión para cada clase.

Las reglas de decisión se construyen de arriba hacia abajo y de izquierda a derecha de manera escalonada desde el nodo raíz hasta los nodos hojas. El nodo ubicado en el extremo izquierdo en la representación constituye la raíz del árbol. Los nodos hojas por su parte son aquellos a los que le sigue el valor alcanzado por el tiempo de vida (variable a predecir). Por ejemplo, al clasificar para la clase "IRHE", con un árbol de decisión de tipo J4.8, el modelo de clasificación permite generar un total de 13 reglas. Para el resto de las clases IRHA, IRHF e IRHT se construyeron igualmente árboles de decisión y para cada uno de ellos se generaron 50, 44 y 50 reglas de decisión respectivamente.

En la Tabla 7 se presentan los resultados de las clasificaciones realizadas, acompañados de diferentes valores estadísticos que se calculan asociados al proceso de validación del modelo obtenido. Se observa el porcentaje de instancias correctamente e incorrectamente clasificadas para cada una de las clases, así como el Coeficiente de Kappa y el Error absoluto para cada una de las clases. WEKA calcula el Coeficiente de Kappa para mostrar la concordancia entre los datos de prueba y la clasificación realizada por el modelo. Cuando todas las instancias son clasificadas correctamente se obtiene la máxima concordancia, esto significa un Coeficiente de Kappa = 1. La última columna de la tabla arroja el resultado del nivel de error generado en el modelo al haber aplicado el algoritmo.

Tabla 7: Resultados de la ejecución del algoritmo J4.8 (200 instancias)

Clases	Instancias Clasificadas Correctamente		Instancias Clasificadas Incorrectamente		Coeficiente de Kappa	Error absoluto
	Instancias	%	Instancias	%		
IRHA	194	97 %	6	3 %	0.9514	0.0316
IRHE	192	96 %	8	4 %	0.9395	0.0463
IRHF	182	91 %	18	9 %	0.8562	0.0641
IRHT	187	93.5 %	13	6.5 %	0.8959	0.0552

Como se puede apreciar en la Tabla 7 el por ciento de instancias clasificadas correctamente en todas las clases es superior en número a la cantidad de instancias clasificadas incorrectamente. Estos resultados expresan que existe una menor proporción de instancias mal clasificadas en las cuatro clases. El promedio de la clasificación correcta entre las 4 clases es de 94.5 %, indicando un alto nivel de clasificación por el algoritmo J4.8 en el conjunto de datos de manera correcta. En los resultados anteriores se evidencia que la mejor clasificación corresponde a la clase IRHA con solo un 3 % de instancias clasificadas incorrectamente.

Para el caso de los cuatro árboles de decisión obtenidos, el Coeficiente de Kappa midió la coincidencia de la predicción con la clase real (1.0 significa que ha habido coincidencia absoluta). El Coeficiente de Kappa promediado resultó ser igual a 0.9108, lo cual indica un alto nivel de concordancia entre los datos de prueba y la clasificación realizada por el modelo para las cuatro clases.

Otro de los datos estadísticos más relevantes en la clasificación mostrada en la Tabla 7, lo constituye el Error absoluto. El valor alcanzado como promedio en las cuatro clases es de 0.0493, considerándose un adecuado resultado que complementa la confiabilidad en la clasificación.

WEKA proporciona la información referente a los parámetros de exactitud del modelo de clasificación obtenido. Estos parámetros son:

- Tasa de Verdaderos Positivos (VP): es la proporción de ejemplos que fueron clasificados como clase x , de entre todos los ejemplos que exactamente tienen clase x , es decir, qué cantidad de la clase ha sido capturada. En la matriz de confusión, es el valor del elemento de la diagonal dividido por la suma de la fila relevante.
- Tasa de Falsos Positivos (FP): es la proporción de ejemplos que fueron clasificados como clase x , pero en realidad pertenecen a otra clase, de entre todos los ejemplos que no tienen clase x . En la matriz de confusión, es la suma de la columna menos el valor del elemento de la diagonal dividido por la suma de las filas de las otras clases.
- Precisión: en cambio, mide el número de términos correctamente reconocidos respecto al total de términos predichos, sean estos verdaderos o falsos términos.

$$- \text{Precisión} = VP / VP + FP$$

La herramienta arrojó los resultados descritos en la Tabla 8.

Tabla 8: Resultados de la clasificación por clase

Clases	Criterios	Tasa de Verdaderos Positivos	Tasa de Falsos Positivos	Precisión
IRHA	Aprovechado	0.988	0.035	0.955
	Sobrecargado	0.875	0.006	0.966
	Sub_utilizado	0.988	0.008	0.988
IRHE	Alto	0.967	0.022	0.952
	Medio	0.919	0.007	0.983
	Bajo	0.987	0.033	0.950
IRHF	En_tiempo	0.954	0.053	0.933
	Adelantado	0.897	0.029	0.839
	Atraso	0.000	0.000	0.000
	Atraso_no_comenzado	0.924	0.058	0.913
IRHT	Alto	0.920	0.027	0.964
	Bajo	1.000	0.024	0.886
	Medio	0.927	0.051	0.927

Con el objetivo de lograr un mejor entendimiento de la Tabla 8 se explica lo que sucede con la clase IRHE. Se observa que el criterio “Alto” presenta una Tasa de Verdaderos Positivos del 96.7%, el criterio “Medio”

del 91.9% y “Bajo” del 98.7 %, por lo cual se interpreta que: dada una instancia el árbol de decisión la clasifica como “Alto” el 96.7 % de las veces, “Medio” el 91.9 % y “Bajo” el 98.7 %. En el mismo sentido, se tiene que la Tasa de Falsos Positivos para las instancias clasificadas por el modelo es del 2.2 % para el criterio “Alto”, del 0.7 % para “Medio” y del 3.3 % para “Bajo”. Con las medidas anteriores, se obtiene la precisión del modelo para la clase IRHE en los criterios “Alto”, “Medio” y “Bajo”, la cual es del 95.2 %, 98.3 % y 95 % respectivamente. Este porcentaje indica la proporción de aciertos del modelo de clasificación obtenido.

Las clases IRHA, IRHE, IRHF e IRHT presentaron una precisión promedio de 0.97, 0.96, 0.89 y 0.94 respectivamente. El conjunto de patrones que se presentan en el modelo fueron obtenidos con una precisión entre las cuatro clases de 0.939, lo que equivale decir que se clasificaron correctamente el 94 % del total de casos.

Los resultados anteriores evidencian que los índices de rendimiento que mejor clasificaron fueron el IRHA e IRHE, con el 97 % y 96 % de las instancias respectivamente. Además se muestra la tasa de valores positivos, que detalla por clase el porcentaje de instancias con ese valor que son correctamente predichas y la tasa de falsos positivos que representa el porcentaje de instancias con otros valores que son incorrectamente clasificadas a ese valor aunque tenían otro. Igualmente con las medidas anteriores, se obtiene la precisión del modelo para cada criterio de las clases.

Matriz de confusión

Existen diversos criterios para evaluar la bondad de los clasificadores. Este proceso de validación permite efectuar una medición sobre la capacidad de predicción del modelo generado a partir de un clasificador. La matriz de confusión es una alternativa para verificar o medir la bondad del clasificador (Porta 2006).

La matriz de confusión permite visualizar mediante una tabla de contingencia la distribución de errores cometidos por un clasificador. Esta matriz de confusión para el caso de dos clases tiene la apariencia que se muestra en la Tabla 9.

Tabla 9: Matriz de Confusión

Clase real	Clase Predicha	
	Play	No Play
Play	Verdaderos positivos (VP)	Falsos negativos (FN)
No Play	Falsos positivos (FP)	Verdaderos negativos (VN)

Suponiendo que N es el número del conjunto de datos de entrenamiento:

$$- N = VP + FN + FP + VN.$$

Donde:

- VP: instancias correctamente reconocidas por el sistema (Verdaderos positivos).
- FN: instancias que son positivas y que el sistema dice que no lo son (Falsos negativos).
- FP: instancias que son negativas pero el sistema dice que no lo es (Falsos positivos).

- VN: instancias que son negativas y correctamente reconocidas como tales (Verdaderos negativos).

En la ejecución del algoritmo la herramienta arroja una matriz de confusión que es de tamaño $n * n$, siendo n el número de clases.

En la Figura 11 se muestra la matriz de confusión generada para la clase IRHE. Teniendo en cuenta que el número de instancias clasificadas correctamente es la suma de la diagonal de la matriz y el resto están clasificadas de forma incorrecta, el resultado para esta clase es de 192 instancias clasificadas correctamente y 8 clasificadas incorrectamente.

```

=== Confusion Matrix ===
  a  b  c  <-- classified as
59  1  1 | a = Alto
 2 57  3 | b = Medio
 1  0 76 | c = Bajo

```

Figura 11: Matriz de confusión generada para la clase IRHE

Por tanto, en la matriz representada en la Figura 11, de los 59 individuos que presentaron un índice de eficiencia alto, solo 2 fueron clasificados incorrectamente; de los que presentaron un índice Medio, hay 57 correctamente clasificados y 5 incorrectos; mientras que de los 76 con índice Bajo solo 1 presentó error.

Los resultados de las matrices generadas para las clases IRHA, IRHF e IRHT reflejaron que para las clases anteriores las instancias clasificadas incorrectamente fueron 6, 18 y 13 respectivamente. En la diagonal de la Matriz de Confusión resultante de la clase IRHF se obtuvieron criterios (Atraso) en las que no se agrupó elemento alguno. De manera general, en las cuatro matrices de confusión generadas en los modelos de clasificación casi todos los elementos se encuentran en la diagonal, o sea, casi todos los elementos fueron clasificados correctamente en la clase a la que pertenecían.

Subproceso Descubrimiento

En este subproceso se determina si existe asociación entre las variables definidas en la investigación: rasgos de la personalidad-rendimiento laboral, mediante la interpretación y evaluación de los resultados. Esta relación se expresará a través del análisis y la interpretación de las reglas derivadas de los modelos de clasificación. En la Figura 12 se representan los elementos que componen esta etapa.

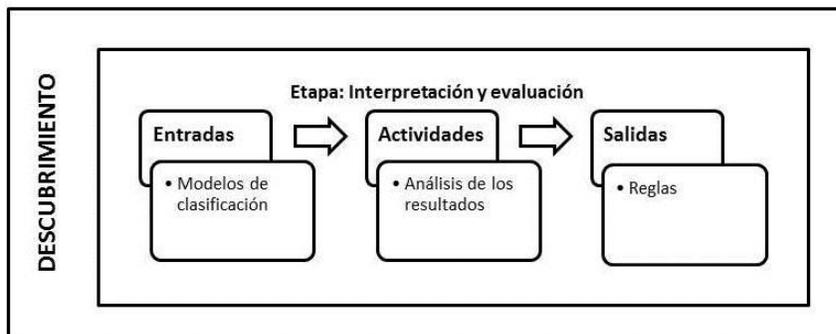


Figura 12: Descripción del subproceso Descubrimiento

Etapa 5: Interpretación y Evaluación

En esta etapa se procede al análisis de los resultados obtenidos en los modelos de clasificación generados. Incluye a su vez la resolución de posibles inconsistencias con otros conocimientos anteriores al presente estudio.

Actividad 7: Análisis de los resultados

Las entradas de esta actividad son los modelos de clasificación generados para las cuatro clases seleccionadas por el clasificador. En esta etapa se analizaron los rasgos relevantes resultantes de la selección de atributos y los árboles de decisión de los modelos de clasificación. A partir de estos resultados se descubre conocimiento a partir de la relación rasgos de la personalidad y el rendimiento laboral de los RRHH. Este conocimiento es expresado mediante reglas de decisión para cada una de las clases en todos sus criterios.

La construcción de reglas en base a clasificaciones se puede realizar mediante métodos directos, donde se extraen reglas desde el conjunto de datos de entrenamiento, o por medio de métodos indirectos como los árboles de decisión. Este último es el usado en la presente investigación para generar las reglas de cada clase.

Los 22 rasgos relevantes seleccionados en la etapa Minería de Datos, apoyaron la construcción de las reglas de decisión a partir de la interpretación de los árboles. Esta selección permitió identificar aquellos rasgos que no fueron seleccionados por el algoritmo porque no presentaron relación con ninguna de las cuatro clases analizadas.

Para extraer las reglas de decisión en los árboles generados se analizaron los nodos, las ramas y los nodos finales. A partir de los nodos se representaron los atributos existentes en el modelo de clasificación, de las ramas se representaron los valores de los atributos identificados y de los nodos finales se obtuvo los valores de la clase. Cada camino del árbol de decisión analizado representó una regla.

En el caso de la clase IRHE se describieron un total de 13 reglas a partir de la interpretación del árbol de decisión correspondiente. De acuerdo con los valores que puede tomar esta clase, las reglas se dividen en: 4 reglas para cuando el IRHE es Alto, 4 para cuando el índice es Medio y 5 cuando el valor es Bajo.

A continuación se interpretan las reglas obtenidas a partir del árbol de decisión para la clase IRHE, expresadas en lenguaje natural. La precisión de estas reglas se puede consultar en la Tabla 8 para cada uno de los valores que toma la clase.

Reglas de decisión generadas para cuando el IRHE es Alto, Medio o Bajo:

- 1) Si el individuo es evaluado con los rasgos modestia normal y asertividad bajo, entonces la predicción es que el índice "IRHE es Alto".
- 2) Si el individuo es evaluado con los rasgos modestia normal, asertividad normal y curiosidad intelectual alto, entonces la predicción es que el índice "IRHE es Alto".

- 3) Si el individuo es evaluado con los rasgos modestia normal, asertividad normal y además curiosidad intelectual bajo e imaginación normal, entonces la predicción es que el índice "IRHE es Medio".
- 4) Si el individuo es evaluado con los rasgos modestia normal, asertividad normal y además curiosidad intelectual bajo e imaginación alto, entonces la predicción es que el índice "IRHE es Bajo".
- 5) Si el individuo es evaluado con los rasgos modestia normal, asertividad alto y búsqueda de emociones normal, entonces la predicción es que el índice "IRHE es Alto".
- 6) Si el individuo es evaluado con los rasgos modestia normal, asertividad alto y búsqueda de emociones alto, entonces la predicción es que el índice "IRHE es Alto".
- 7) Si el individuo es evaluado con los rasgos modestia bajo e impulsividad normal, entonces la predicción es que el índice "IRHE es Medio".
- 8) Si el individuo es evaluado con los rasgos modestia normal, asertividad normal y curiosidad intelectual normal, entonces la predicción es que el índice "IRHE es Bajo".
- 9) Si el individuo es evaluado con los rasgos modestia normal, asertividad alto y búsqueda de emociones bajo, entonces la predicción es que el índice "IRHE es Bajo".
- 10) Si el individuo es evaluado con los rasgos modestia alto e imaginación normal, entonces la predicción es que el índice "IRHE es Bajo".
- 11) Si el individuo es evaluado con los rasgos modestia alto e imaginación alto, entonces la predicción es que el índice "IRHE es Medio".
- 12) Si el individuo es evaluado con los rasgos modestia bajo e impulsividad bajo, entonces la predicción es que el índice "IRHE es Medio".
- 13) Si el individuo es evaluado con los rasgos modestia bajo e impulsividad alto, entonces la predicción es que el índice "IRHE es Bajo".

WEKA muestra además en la clasificación dos datos estadísticos que tienen amplia relación con las reglas. Estos datos son el F-Measure (confiabilidad) y el Recall (cobertura). El primero representa la media armónica entre la precisión y el Recall. Entre más cercana a 1, mayor es la confiabilidad del modelo de clasificación en la clase analizada. Por ejemplo, en la clase IRHE para el criterio "Alto" se tiene una confiabilidad de 0.959 representando un 95.9 %, para el criterio "Medio" de 0.950 equivalente a un 95 % y por último para el criterio "Bajo" un 0.968, o sea un 96.8 %, promediando un F-Measure de 0.960. De esta forma, el árbol de decisión obtenido clasifica de manera aceptable las instancias pertenecientes a la clase IRHE en los criterios "Alto", "Medio" y "Bajo". Estos datos demuestran que el conjunto de reglas de decisión clasifican con un buen grado de confiabilidad a las instancias en esa clase.

Las reglas de decisión interpretadas a partir de los árboles generados para el resto de las clases pueden examinarse en el Anexo 1. El número de reglas, por cada clase y por cada criterio, se expresa en la Tabla 10.

Incluso, cuando los datos de la precisión alcanzada en las cuatro clases nunca estuvo por debajo de 0.70, sino por encima del 0.85, fue necesario seleccionar las reglas de decisión que mayor precisión alcanzaron (entre más cercana a 1 mejor). Con este objetivo, se desearon las reglas que presentaron precisión por

debajo del 93%, debido a que los antecedentes de estas reglas contenían condiciones irrelevantes para la investigación.

Tabla 10: Reglas de decisión generadas por clase y por criterios.

Clases	Criterios	Cantidad de reglas	Total de reglas
IRHA	Sub_utilizado	17	50
	Aprovechado	23	
	Sobrecargado	10	
IRHE	Bajo	5	13
	Medio	4	
	Alto	4	
IRHF	Atraso	0	44
	Atraso_no_comenzado	15	
	En_tiempo	21	
	Adelantado	8	
IRHT	Bajo	9	50
	Medio	15	
	Alto	26	

En este sentido, según (Servente 2002) las reglas pueden ser generalizadas eliminando dichas condiciones, pero en lugar de analizar todos los subconjuntos posibles de condiciones que podrían eliminarse, el algoritmo J4.8 realizó una eliminación directa golosa (Straightforward Greedy Elimination). De todas las reglas que podían eliminarse por este método, se eliminaron aquellas que producen la menor proporción de error en la regla generalizada. Por lo tanto, se puede afirmar que en el proceso de generalización estas reglas se descartaron ya que dejaron de ser mutuamente excluyentes y exhaustivas.

En la Tabla 11 se muestra el listado de las reglas de decisión seleccionadas para la clase IRHE de acuerdo al valor de precisión alcanzado. Todas las reglas generadas para esta clase fueron seleccionadas debido al alto nivel de precisión obtenido.

Tabla 11: Reglas de decisión generadas a partir del árbol de decisión de la clase IRHE.

Reglas	Precisión	Reglas	Precisión
Regla 1	0.952	Regla 8	0.950
Regla 2	0.952	Regla 9	0.950
Regla 3	0.983	Regla 10	0.950
Regla 4	0.950	Regla 11	0.983
Regla 5	0.952	Regla 12	0.983
Regla 6	0.952	Regla 13	0.950
Regla 7	0.983		

Este proceso de evaluación de las reglas de decisión se realizó para el resto de las clases con el objetivo de determinar las que son correctas. La selección permitió determinar que el total de reglas correctas son 76. En la clase IRHA se seleccionaron 25 reglas, en la clase IRHF 17 reglas y en la IRHT 21. Estas reglas constituyen el conocimiento referente a la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH.

Patrón ideal de la personalidad

Tomando como base las reglas de decisión generadas y tras un análisis de los resultados obtenidos en la clasificación sobre el conjunto de datos de entrenamiento seleccionado, se propone un patrón ideal de la personalidad. Este patrón se construyó a partir de la unificación de las reglas generadas para las cuatro clases en su mejor criterio, o sea, el máximo valor que puede alcanzar una clase.

En la Tabla 12 se muestra el conjunto de rasgos de la personalidad que integran las reglas generadas y los valores que deben tomar estos rasgos para cada una de las clases en su mejor criterio.

Tabla 12: Rasgos de la personalidad y sus valores para cada clase en su mejor criterio

Rasgos de la personalidad y valores a cumplir		Mejor criterio para cada clase
Amigabilidad=Normal, Alto. Confianza = Alto, Normal, Bajo. Autodisciplina = Normal, Bajo, Alto. Simpatía = Normal, Alto, Bajo. Intereses artísticos=Normal, Alto, Bajo.	Orden = Normal, Alto, Bajo. Liberalismo= Alto, Bajo, Normal. Valores éticos= Normal, Ninguno. Ansiedad=Alto, Bajo. Depresión=Bajo, Normal, Ninguno.	IRHA=Aprovechado
Modestia= Normal. Responsabilidad = Alto. Asertividad= Bajo, Normal, Alto.	Curiosidad intelectual= Alto. Búsqueda emociones=Normal, Alto.	IRHE= Alto
Obediencia = Normal, Alto. Simpatía = Normal, Alto, Bajo. Asertividad= Alto, Normal, Bajo. Nivel de Actividad=Bajo, Normal, Alto, Ninguno.	Liberalismo= Alto, Bajo, Normal. Confianza = Alto, Normal, Bajo. Autodisciplina = Normal. Intereses artísticos=Normal, Alto, Bajo.	IRHF=En tiempo
Prudencia=Normal, Alto. Aventuras=Normal, Bajo, Alto, Ninguno. Autodisciplina = Alto, Normal, Bajo. Nivel de Actividad=Bajo, Normal, Alto, Ninguno.	Orden = Normal, Alto, Bajo. Imaginación=Bajo, Normal, Alto. Cooperación = Normal, Alto. Búsqueda emociones=Normal, Alto, Bajo.	IRHT= Alto

Teniendo en cuenta lo expresado en la Tabla 12, en la investigación se considera que una personalidad ideal es aquella que obtenga a la vez en los cuatro índices de rendimiento su mejor criterio, o lo que es lo mismo, que cumpla con las reglas de decisión propuestas en la Tabla 13.

Tabla 13: Propuesta de patrón ideal de la personalidad

Índices de rendimiento	Reglas a partir de los rasgos relevantes
IRHA=Aprovechado IRHE=Alto IRHF=En tiempo IRHT=Alto	Un individuo presenta una personalidad ideal si cumple con la siguiente evaluación para los rasgos: Responsabilidad = Alta Liberalismo= Alto Amigabilidad = Alto Valores éticos= Normal Intereses artísticos = Alto Ansiedad= Bajo Confianza = Alto Modestia= Normal Autodisciplina = Alto Asertividad= Alto Simpatía = Alto Obediencia = Alto Depresión=Bajo o Ninguno Nivel de Actividad=Alto o Normal Orden = Alto Imaginación= Alto Prudencia= Alto Cooperación = Alto Aventuras=Normal o Alto Curiosidad intelectual=Alto Búsqueda emociones=Alto

De acuerdo a lo planteado por (Corso 2009), estas reglas deben cumplir las siguientes propiedades:

- Mutuamente Exclusivas (ME): es decir las “ r ” de R son (ME) si no se dispara más de una r para un determinado registro.
- Exhaustivas: R tiene una cobertura exclusiva si posee una “ r ” para cada combinación de valores de los atributos. Esto asegura que toda instancia es cubierta por al menos una r de R .

El patrón no persigue identificar a los buenos ingenieros, ni sus resultados predicen de manera absoluta la calidad de la ejecución de un trabajo. Se requiere un balance entre las características o rasgos de la personalidad y el rendimiento.

El descubrimiento del conocimiento obtenido en el presente capítulo puede ser empleado en los procesos de selección y evaluación de los RRHH en proyectos informáticos, con la finalidad de utilizar los rasgos relevantes obtenidos, las reglas de decisión generadas y la propuesta de patrón ideal de la personalidad en la evaluación individual. Esta propuesta permite predecir a partir de los rasgos de la personalidad relevantes y la confección de las reglas de decisión, los posibles valores que un individuo puede alcanzar en su rendimiento de acuerdo a los cuatro índices: IRHA, IRHE, IRHF e IRHT, independientemente del entorno productivo donde se desarrolle.

Conclusiones del capítulo

De la aplicación de las etapas para la extracción de conocimiento en el presente capítulo, se arribó a las siguientes conclusiones:

- La aplicación de la técnica CFS permitió seleccionar 22 rasgos relevantes, representando un 73% de un total de 30 rasgos evaluados, empleados posteriormente en la clasificación del conjunto de datos de entrenamiento utilizado.
- Los rasgos que más afectan positiva o negativamente a los cuatro índices de rendimiento laboral: son Nivel de actividad, Imaginación, Liberalismo, Confianza, Capacidad de logro, Orden, Asertividad, Autodisciplina, Búsqueda de emociones y Curiosidad intelectual. Estos rasgos están asociados en su mayoría a los tipos de personalidad: Extraversión, Apertura a la experiencia y Responsabilidad.
- La aplicación del algoritmo de árboles de decisión J4.8 permitió realizar la clasificación del conjunto de datos de entrenamiento mediante la construcción de cuatro árboles de decisión. El análisis e interpretación de estos árboles permitió generar un conjunto de 76 reglas que apoyarán la toma de decisiones en proyectos informáticos.
- Se construye una propuesta de patrón ideal de la personalidad a partir de las reglas de decisión generadas para las clases en su mejor criterio, que permite a los directivos identificar los individuos que mejores índices de rendimiento laboral presentan.

CAPÍTULO 3: VALIDACIÓN DE LOS RESULTADOS

En el presente capítulo se describe la validación realizada al conocimiento descubierto: rasgos relevantes y reglas de decisión generadas, al aplicar el proceso de descubrimiento de conocimiento a partir de la relación entre los rasgos de la personalidad y el rendimiento laboral. Se valida la aplicación del algoritmo árboles de decisión J4.8, teniendo en cuenta el nivel de confiabilidad alcanzado en la clasificación al conjunto de datos de entrenamiento. Por último, se realiza la validación del patrón ideal de la personalidad a partir del criterio de expertos y estudio de casos.

Estrategia de validación

La estrategia de validación definida en la investigación agrupa los resultados alcanzados en el proceso de descubrimiento de conocimiento propuesto en el capítulo anterior, así como la validación de la aplicación del algoritmo J4.8 y el patrón ideal de la personalidad.

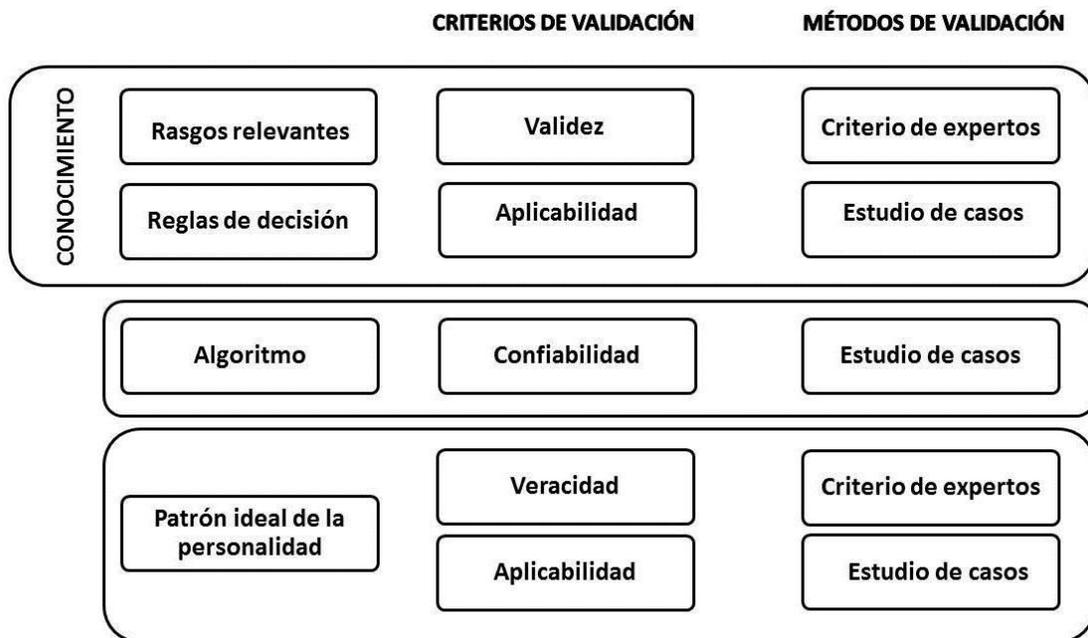


Figura 13: Estrategia de validación propuesta en la investigación.

La validación se realizó de acuerdo al conocimiento descubierto referente a la selección de los rasgos relevantes, la generación e interpretación de las reglas de decisión y finalmente la propuesta del patrón ideal de la personalidad como valor agregado en la investigación.

Como se muestra en la Figura 13, en la validación del primer resultado se aplicó una encuesta para evaluar la validez de los rasgos relevantes utilizando el criterio de expertos. Posteriormente, se validaron las reglas de decisión teniendo en cuenta su aplicabilidad mediante un estudio de casos en el CEIGE. Como segundo resultado, se validó la confiabilidad del algoritmo J4.8 utilizando también un estudio de casos. Por último, la veracidad del patrón ideal de la personalidad a través de criterio de expertos y su aplicabilidad empleando el método de validación estudio de casos.

Validación de los rasgos relevantes

El primer resultado obtenido en el subproceso Minería de Datos fue la identificación de los rasgos más relevantes para cada una de las clases. Este resultado se obtuvo mediante la selección de atributos como parte del clasificador y utilizando el método de búsqueda Greedy Stepwise y el evaluador CFS Subset Evaluator.

Para asegurar la validez de los rasgos relevantes se construyó una propuesta a partir del consenso de un grupo de expertos con el fin de obtener opiniones acerca de cuáles rasgos consideraban ellos que más influyen en los índices de rendimiento utilizados. De esta forma se pudo comparar los resultados expresados por los expertos con los obtenidos por el algoritmo J4.8. La finalidad de la utilización de este método es encontrar criterios que estuvieran más condicionados por la experiencia de los expertos y no desde el punto de vista estadístico con el objetivo de que existiera una complementariedad de conocimientos.

Para obtener el criterio de los expertos fue necesario primeramente la selección de los mismos. Luego se aplicó una encuesta mediante la cual cada experto emitió su criterio respecto al nivel de importancia que le concedían a cada rasgo de la personalidad en función de los índices de rendimiento. Por último se analizaron los resultados y se confeccionó una propuesta de los rasgos relevantes de la personalidad que más influyen en el rendimiento laboral.

Selección de los expertos

La selección de los expertos se realizó con el objetivo de validar los datos obtenidos por el algoritmo, teniendo en cuenta la utilización sistemática de las valoraciones intuitivas del grupo de expertos para obtener un consenso de opiniones, que de acuerdo entre las partes, refuerzan la validez de la propuesta. Para la selección de los expertos se decidió utilizar el análisis de síntesis curricular, a partir de tres criterios:

- Más de cinco años de experiencia como psicólogos (válido solo para los expertos psicólogos).
- Más de cinco años de experiencia en el proceso de desarrollo de software (válido solo para los expertos que no son psicólogos).
- Experiencia en las siguientes temáticas: trabajo en equipo, rendimiento y competencias específicas y genéricas. Se tuvieron en cuenta las publicaciones y la participación en eventos científicos.

Fueron seleccionados 8 expertos, pertenecientes a las siguientes áreas e instituciones: Facultad 3 UCI, CEIGE Facultad 3, CICE UCI y la Facultad de Psicología de la Universidad de Oriente. Los cargos desempeñados por los expertos encuestados son listados a continuación:

- 1) Psicóloga del CICE.
- 2) Directora del CEIGE, Facultad 3.

- 3) Jefe del equipo de desarrollo Departamento Banco, CEIGE, Facultad 3.
- 4) Analista principal de CEDRUX, CEIGE, Facultad 3.
- 5) Graduada en Psicología en la Universidad de Oriente
- 6) Jefe de Departamento Desarrollo de Productos, CEIGE, Facultad 3.
- 7) Subdirector de Formación CEIGE, Facultad 3.
- 8) Jefe de Departamento Aduana, CEIGE, Facultad 3.

Aplicación de la encuesta y análisis de los resultados

Cada experto recibió la encuesta (Ver Anexo 2) mediante la cual emitió su criterio siguiendo la escala numérica mostrada en la Tabla 14. Esta escala corresponde a la asignación del peso para cada uno de los rasgos de la personalidad en función de su importancia para los índices de rendimiento IRHA, IRHE, IRHF e IRHT. Las respuestas enviadas por los expertos se recogieron en un documento diseñado en Microsoft Excel 2010.

Tabla 14: Escala de peso utilizada en la encuesta

Criterios de acuerdo a la importancia	Pesos
Si es muy importante para el índice de rendimiento.	5 puntos
Si es importante para el índice de rendimiento.	4 puntos
Si la importancia para el índice de rendimiento es media.	3 puntos
Si no es tan importante para el índice de rendimiento.	2 puntos
Si no tiene importancia alguna para el índice de rendimiento.	1 punto

Para determinar el consenso de los expertos respecto al nivel de cada competencia se empleó como medida de tendencia la moda. En la Figura 14 se muestra el criterio de cada experto y el nivel de relevancia de cada rasgo a partir de la convergencia de los criterios emitidos.

Este método permitió identificar los rasgos relevantes de acuerdo al criterio de los expertos, basados en las experiencias con los RRHH en el área en la que se desempeñan. Para determinar cuáles son los rasgos relevantes se seleccionaron los que mayor peso obtuvieron en el análisis de los resultados, tomándose los que presentaron un peso de 3, 4 y 5, con evaluación de “Tiene importancia media”, “Importante” y “Muy importante” respectivamente.

En los resultados obtenidos prevaleció el criterio de que los rasgos más relevantes son los que pertenecen al tipo de personalidad Extraversión y Responsabilidad. El criterio asignado a los rasgos “Muy importantes” con peso 5 fluctúa entre 10 y 12 rasgos agrupados en el resto de los tipos de personalidad, restándole importancia al Neuroticismo como se refleja en la Figura 14.

Tipo de Personalidad	Rasgos	CRITERIO DE LOS EXPERTOS								Concenso	Nivel
		E1	E2	E3	E4	E5	E6	E7	E8		
Neuroticismo	Ansiedad	5	5	4	5	5	5	5	3	5	Es muy importante
	Ira	4	2	4	1	3	2	1	1	1	No tiene importancia alguna
	Depresión	5	4	2	5	5	1	4	5	5	Es muy importante
	Timidez	2	2	3	1	4	5	2	3	2	No es tan importante
	Impulsividad	2	3	4	4	3	2	5	3	3	Tiene importancia media
	Vulnerabilidad	1	2	1	2	4	5	3	2	2	No es tan importante
Extraversión	Amigabilidad	2	1	3	4	5	4	3	4	4	Es importante
	Gregarismo	4	1	2	4	3	4	5	2	4	Es importante
	Asertividad	5	4	4	5	3	1	5	5	5	Es muy importante
	Nivel de actividad	4	5	4	5	5	5	5	5	5	Es muy importante
	Búsqueda de emociones	4	3	2	4	5	4	3	5	4	Es importante
	Emoción positiva	3	2	2	4	3	4	5	4	4	Es importante
Apertura a la experiencia	Imaginación	2	1	4	3	5	4	5	3	4	Es importante
	Intereses artísticos	1	2	3	4	2	4	3	1	1	No tiene importancia alguna
	Sensibilidad emocional	3	2	1	3	4	4	2	2	2	No es tan importante
	Aventura	4	1	4	5	4	5	3	2	4	Es importante
	Curiosidad intelectual	5	5	5	5	5	5	5	5	5	Es muy importante
	Liberalismo	4	3	4	5	3	2	1	3	3	Tiene importancia media
Amigabilidad	Confianza	5	5	5	4	5	3	5	4	5	Es muy importante
	Valores éticos	4	3	3	4	4	5	4	5	4	Es importante
	Altruismo	2	3	4	1	2	3	2	2	2	No es tan importante
	Cooperación	5	3	4	5	3	4	5	5	5	Es muy importante
	Modestia	4	5	5	5	5	5	5	5	5	Es muy importante
	Simpatía	3	5	3	4	4	4	3	4	4	Es importante
Responsabilidad	Autoeficacia	5	5	5	5	5	4	5	4	5	Es muy importante
	Orden	1	3	4	5	4	4	5	4	4	Es importante
	Obediencia	5	4	5	5	5	4	5	4	5	Es muy importante
	Capacidad de logro	5	5	3	4	2	4	4	3	4	Es importante
	Autodisciplina	5	5	5	5	5	5	4	5	5	Es muy importante
	Prudencia	5	4	5	5	4	2	5	4	5	Es muy importante

Figura 14: Resultado de la encuesta realizada a los expertos para determinar los rasgos relevantes

La comparación entre los datos arrojados por este método y los resultantes tras aplicar el algoritmo de Minería de Datos J4.8, mostró que el algoritmo identificó un conjunto de 22 rasgos relevantes, mientras que los expertos consideraron que son 24 rasgos los que presentan mayor nivel de importancia. En la Tabla 15 se muestran los criterios expresados por ambas partes.

Tabla 15: Rasgos relevantes identificados por los expertos y por la clasificación del algoritmo J4.8

Rasgos relevantes Criterio de expertos		Rasgos relevantes Algoritmo J4.8	
Nivel de actividad	Curiosidad intelectual	Nivel de Actividad	Curiosidad intelectual
Ansiedad	Búsqueda de Emociones	Ansiedad	Búsqueda de emociones
Depresión	Amigabilidad	Depresión	Amigabilidad
Aventura	Valores éticos	Aventura	Valores éticos
Liberalismo	Impulsividad	Liberalismo	Impulsividad
Imaginación	Obediencia	Imaginación	Obediencia
Confianza	Prudencia	Confianza	Prudencia
Simpatía	Cooperación	Simpatía	Cooperación
Orden	Modestia	Orden	Modestia
Autodisciplina	Asertividad	Autodisciplina	Asertividad
Autoeficacia	Gregarismo	Intereses artísticos	Sensibilidad emocional
Emoción positiva	Capacidad de logro		

Al comparar los resultados de la Tabla 15 se evidencia que, 20 de los 22 rasgos relevantes seleccionados por el algoritmo son un subconjunto de los 24 propuestos por los expertos. Los encuestados consideraron

que de los rasgos expresados por el algoritmo, 20 cobran relevante importancia también para ellos y los siguientes dos carecen de ella: Intereses Artísticos y Sensibilidad Emocional.

Como resultado de la comparación entre los rasgos relevantes seleccionados por los expertos y los rasgos relevantes seleccionados por el algoritmo (Ver Tabla 15), se evaluó el nivel de coincidencia entre ellos. A pesar del alto grado de similitud en la selección de los rasgos relevantes por ambas partes, los expertos expresaron que deberían tenerse en cuenta además los siguientes rasgos: Autoeficacia, Emoción Positiva, Capacidad de Logro y Gregarismo. Como se puede apreciar en la Figura 15, se obtuvo un 91 % de coincidencia lo cual es considerado como un por ciento aceptable según el criterio expresado por (Grau 2012).

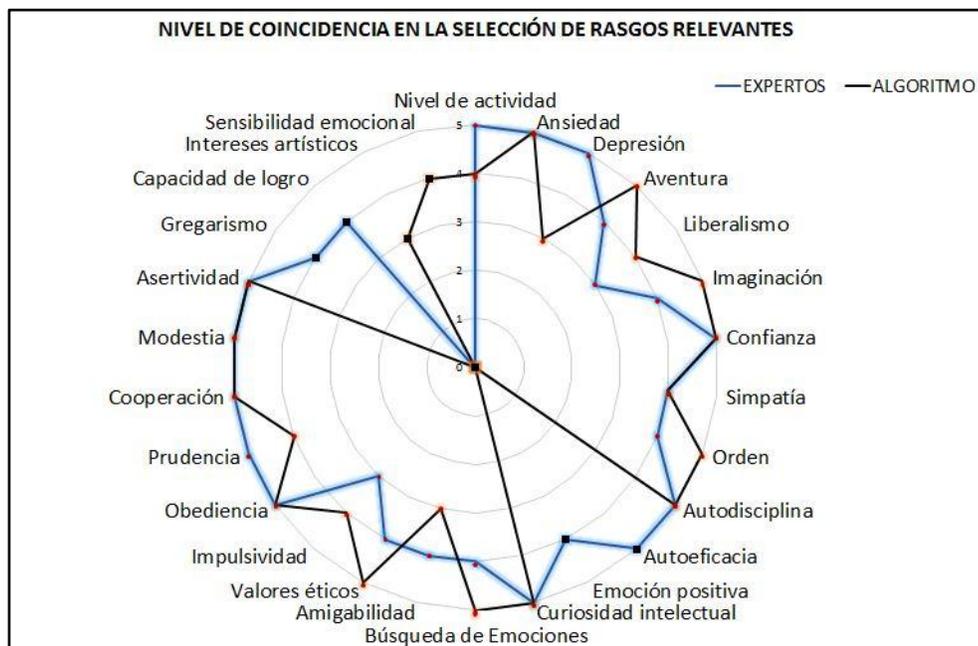


Figura 15: Nivel de coincidencia en la selección de rasgos relevantes

Considerando lo expresado anteriormente se concluye que los rasgos relevantes son aquellos que coincidentemente fueron seleccionados por el algoritmo y por los expertos, arrojando un total de 20 rasgos. Sin embargo, se desechan por no existir coincidencia en su selección los siguientes rasgos: Autoeficacia, Emoción positiva, Gregarismo, Capacidad de logro, Intereses Artísticos y Sensibilidad emocional.

Validación de las reglas de decisión generadas

Los modelos descriptivos en general, son complejos de evaluar debido a la ausencia de una clase determinada donde medir el grado de acierto del modelo. La evaluación más adecuada es saber si el modelo resultante de la fase de aprendizaje tiene un comportamiento útil cuando se utilice en su área de aplicación.

Preparación de los datos

En un primer nivel de validación se realizó un estudio de casos basado en el análisis del comportamiento de los rasgos de la personalidad en función del rendimiento de los mismos. Los datos utilizados en el estudio de casos correspondieron a una muestra de 50 individuos pertenecientes a cuatro proyectos del CEIGE.

El tamaño de la muestra fue tomado a conveniencia pero la composición de la misma es heterogénea en cuanto a nivel escolar, edad, sexo, proyecto donde labora y rol que desempeña, por lo que se considera una muestra representativa de la amplia gama de individuos que componen el CEIGE.

Se tomaron datos primarios que permitieron calcular el valor de los índices del rendimiento de las 50 personas que integraron la muestra con respecto al aprovechamiento, la eficiencia, la eficacia y el trabajo en el período 2011-2013. La selección se sustentó fundamentalmente en la necesidad de alcanzar una correspondencia en este estudio de casos con los datos utilizados en la investigación. El cálculo de estos índices se realizó teniendo en cuenta la evaluación del rendimiento mediante los indicadores propuestos por (Lugo 2012).

Los datos para calcular el rendimiento de los RRHH fueron extraídos de la herramienta GESPRO. De cada individuo se obtuvo un conjunto de información adicional: nombres y apellidos, sexo, proyecto al que pertenece y el rol que desempeña.

Los individuos fueron encuestados mediante el instrumento NEO-PI. Este test permitió que cada individuo se autoevaluara. Como datos a utilizar se tomó la calificación obtenida en los 30 rasgos de la personalidad que en este instrumento se evaluaron. Se incluyeron los individuos con todos los rasgos evaluados y los valores correspondientes a los cuatro índices de rendimiento. De este modo se garantizó el número inicial de casos establecidos para el estudio.

El propósito de este estudio de casos fue identificar los individuos de la muestra que cumplieron con las reglas de decisión generadas en el descubrimiento de conocimiento. Es por esto que fue necesario, una vez generadas las reglas, medir la calidad de mismas. Esta medición de acuerdo a (Corso 2010) en su estudio sobre la aplicación de algoritmos en WEKA, se debe realizar mediante las siguientes mediciones: la cobertura y la exactitud.

La cobertura mide la proporción de términos correctamente reconocidos respecto al total de términos reales. La cobertura y la precisión mantienen una relación inversa, cuando aumenta la cobertura del modelo generado disminuye la precisión y viceversa, cuando disminuye la cobertura aumenta la precisión. De manera tal que se obtiene una cobertura total, a costo de una precisión nula (Corso 2009).

Los parámetros de exactitud por su parte son: los verdaderos positivos, los falsos positivos, la cobertura, la precisión y la confiabilidad (Corso 2009).

Las mediciones cobertura y exactitud se calculan:

$$1) \text{cobertura}(r) = |A| / |D|$$

$$2) \text{ exactitud}(r) = |A \cap y| / |A|$$

Donde:

- $|A|$: es la cantidad de instancias de un conjunto de entrenamiento D , que satisface la precondition.
- $|A \cap y|$: cantidad de instancias del conjunto de entrenamiento que satisface la precondition y el consecuente.
- $|D|$: representa el número total del conjunto de entrenamiento.

Aplicación de las reglas de decisión

Para realizar el proceso de validación de las reglas generadas mediante los valores anteriores fue necesario calcularlos en función de la muestra seleccionada.

Tomando como referencia la clase IRHE con criterio Alto se calculó la exactitud y la cobertura de acuerdo a las fórmulas planteadas anteriormente:

Siendo:

- $|D|$ = conjunto de entrenamiento de 50 personas.
- $|A|$ = 16 instancias de un conjunto de entrenamiento D que satisface la precondition.
- $|A \cap y|$ = 15 instancias del conjunto de entrenamiento que satisface la precondition y el consecuente.

$$1) \text{ cobertura}(r) = |A| / |D|$$

$$= 16/50$$

$$= 0.32$$

$$2) \text{ exactitud}(r) = |A \cap y| / |A|$$

$$= 15/16$$

$$= 0.9375$$

Teniendo en cuenta que uno de los criterios de éxito de la Minería de Datos es obtener las predicciones con un valor de certeza igual o superior al 80% (Corso 2010; Morate 2013), se concluyó que la clase IRHE con criterio Alto presenta un por ciento de exactitud con valor aproximado al 94%, por encima del expresado como límite por los autores. Este cálculo se realizó para las cuatro clases en todos sus criterios. Fue necesario ejecutar una búsqueda en el conjunto de 50 personas para identificar cuántas cumplían con la precondition de las reglas de decisión y cuántas cumplían además con el consecuente. Los resultados de esta búsqueda se expresan en la Tabla 16, así como los valores resultantes del cálculo de la cobertura y la exactitud para cada una de las clases.

Los resultados de la Tabla 16 muestran una alta exactitud en todas las clases. Debe destacarse el criterio Adelantando en la clase IRHF, el Alto en IRHT y el Medio en IRHE debido a que alcanzaron un 100% de exactitud. Sin embargo, se obtuvo un 80% en el criterio Atraso_no_comenzado en la clase IRHF.

Los estimados de exactitud obtenidos en las cuatro clases fueron promediados para realizar una estimación completa de la clasificación, obteniéndose como resultado un 93.1% de exactitud. Este valor aporta a la investigación un alto grado de precisión en las reglas de decisión generadas en el descubrimiento de conocimiento y garantiza un nivel de confianza alto en el resultado obtenido.

Tabla 16: Resultados del cálculo de la cobertura y la exactitud

Clases	No. Criterio	Criterios	A ∩ y	A	Cobertura	Exactitud (%)
IRHA	1.	Sub_utilizado	9	10	0.20	90.0%
	2.	Aprovechado	21	23	0.46	91.3%
	3.	Sobrecargado	13	15	0.3	86.6%
IRHE	4.	Bajo	15	17	0.34	88.2%
	5.	Medio	8	8	0.16	100%
	6.	Alto	15	16	0.32	93. %
IRHF	7.	Atraso	No se calcula			
	8.	Atraso_no_comenzado	4	5	0.10	80.0%
	9.	En_tiempo	32	34	0.68	94.1%
	10.	Adelantado	10	10	0.20	100%
IRHT	11.	Bajo	21	22	0.44	95.4%
	12.	Medio	18	19	0.38	94.7%
	13.	Alto	9	9	0.18	100%

En el caso de la clase IRHF en la Tabla 16, se observa que para el criterio Atraso no existe ningún cálculo debido a que para este valor del índice no se generaron reglas. Como se observa muchas veces el resultado alcanzado puede detectar relaciones triviales o conocidas previamente, o puede ocurrir el hecho de no encontrar relaciones significativas. Después del análisis se concluyó que la clase IRHF con criterio Atraso no es condicionada significativamente por ningún rasgo de la personalidad de acuerdo al modelo la clasificación para esa clase.

En el gráfico de la Figura 16 se representan los individuos que cumplieron con la precondición y el consecuente de las reglas de decisión generadas por cada clase. Se sometieron al análisis las cuatro clases utilizadas en la investigación con el fin de determinar cuáles presentaron mejores resultados de acuerdo al total de instancias que cumplieron con la precondición y el consecuente de las reglas de decisión.

Los resultados obtenidos representados en la Figura 16 reflejan que las dos clases que mayor cantidad de reglas cumplieron fueron la IRHT y la IRHF, para un total de 48 y 46 instancias respectivamente, coincidiendo con las mayores precisiones alcanzadas. Al comparar los resultados con la Tabla 16, el promedio de exactitud de la clase IRHA es de un 90 % y el de la clase IRHE es de 92.7 %, por debajo ambos del promediado por las clases IRHT (96%) e IRHF (94 %).

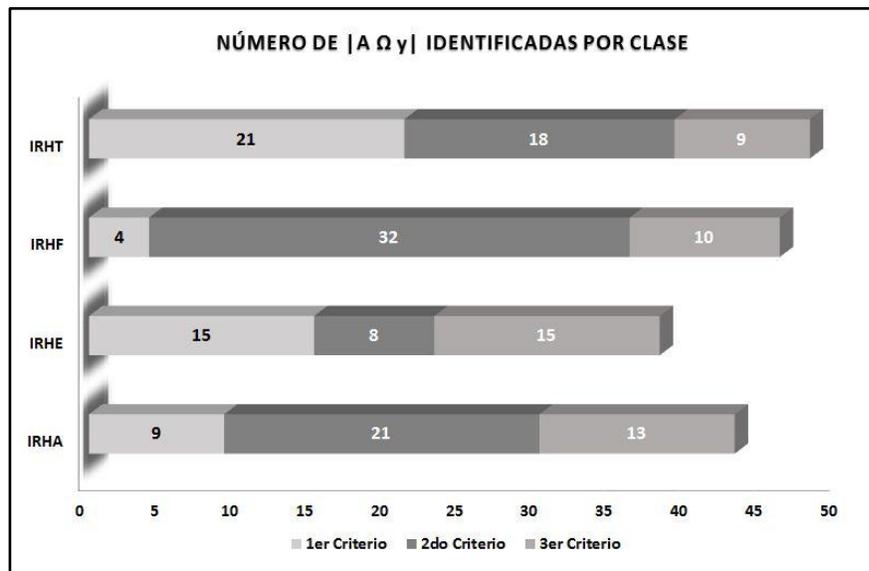


Figura 16: Número de instancias que cumplieron con la precondición y el consecuente

Al aplicar las reglas de decisión para cada una de las clases se identificaron 13 individuos que no formaron parte de este proceso, debido a que no cumplieron en ningún momento con la precondición de las reglas generadas, exceptuando la clase IRHT para la cual todos los individuos satisfacen la precondición.

En las Figura 17, 18, 19 y 20 se representan los resultados para cada una de las clases referentes a la cantidad de instancias del conjunto de datos de entrenamiento que satisfacen la precondición y las que satisfacen la precondición y el consecuente.

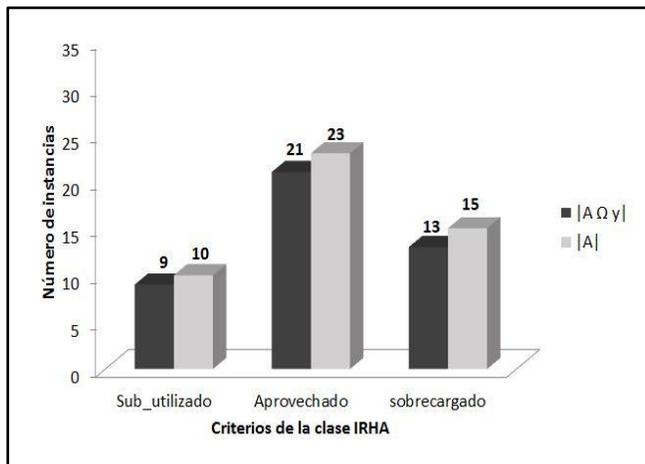


Figura 17: Resultados de la aplicación de las reglas de decisión para la clase IRHA en 50 instancias

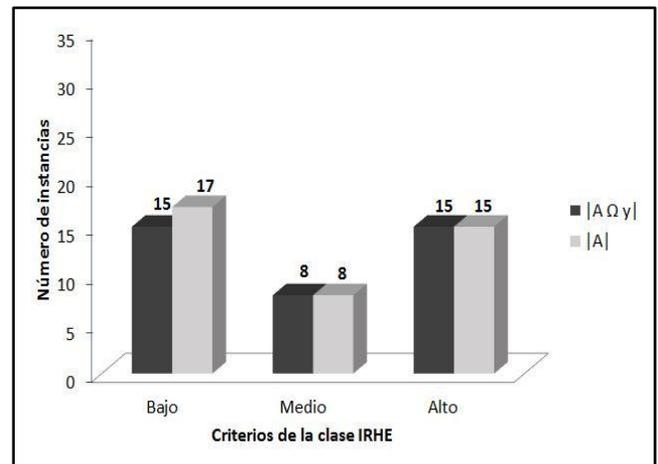


Figura 18: Resultados de la aplicación de las reglas de decisión para la clase IRHE en 50 instancias

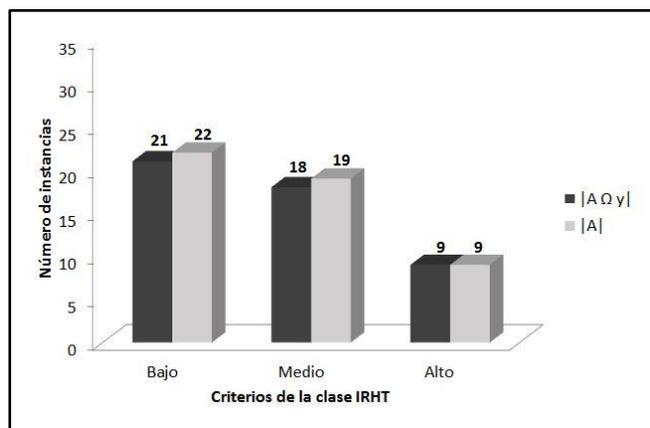


Figura 19: Resultados de la aplicación de las reglas de decisión para la clase IRHT en 50 instancias

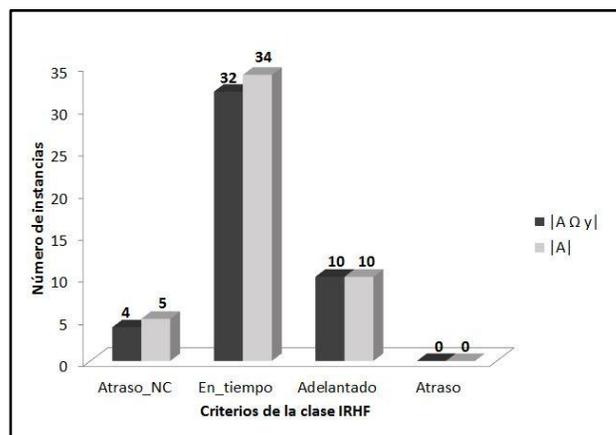


Figura 20: Resultados de la aplicación de las reglas de decisión para la clase IRHF en 50 instancias

Las cifras expresadas en los gráficos anteriores muestran que a pesar de haber clasificado mayor número de reglas, la clase IRHE no fue la que mayor promedio de exactitud obtuvo. Este análisis reflejó que las dos mejores clases en el cumplimiento de las reglas fueron la IRHE en primer lugar y la IRHT en segundo, teniendo en cuenta el número de instancias que cumplieron con las reglas de decisión generadas y el promedio de exactitud alcanzado.

Los análisis realizados sobre la validación del conocimiento obtenido, demuestran la validez de las reglas de decisión generadas en la presente investigación. Se estableció una relación no conocida, la cual ofrece una visión global de cuánto influyen los rasgos de la personalidad en el rendimiento laboral de un individuo en proyectos informáticos.

Validación de la aplicación del algoritmo J4.8

Se realizó un estudio del nivel de confianza que presenta el algoritmo J4.8 en la clasificación con respecto a las cuatro clases del modelo. Este análisis comenzó con la obtención de los valores de las instancias clasificadas correctamente en cada una de las clases. El objetivo de este análisis fue identificar cuál es el nivel de confianza promedio arrojado por el algoritmo en el conjunto de datos de entrenamiento.

Nivel de confianza

Considerando lo planteado en (Curso 2010; WEKA 2013), respecto a que la eficiencia de los algoritmos se mide de acuerdo al tiempo de ejecución empleado en la construcción del modelo y al número de instancias clasificadas correctamente sobre el conjunto de entrenamiento, se considera correcta la elección del algoritmo J4.8. Esta afirmación se basa en que los resultados reflejaron que el tiempo de ejecución promedio en la clasificación realizada para construir el modelo en las cuatro clases es de 0.012 segundos y el promedio de instancias clasificadas correctamente es de un 94.35 %. Estos números reflejan una alta confiabilidad en los resultados alcanzados con la aplicación del algoritmo.

En la Figura 21 se refleja el nivel de variación relacionado con la confiabilidad de la clasificación realizada por el algoritmo J4.8 en cada una de las cuatro clases del modelo. Al analizar el gráfico se evidencia que

con 200 instancias, las cuatro clases obtuvieron un nivel de confianza por encima del 90 % en la clasificación utilizando el algoritmo. Este resultado expresa que existe una alta confiabilidad del modelo de clasificación en cada una de las clases.

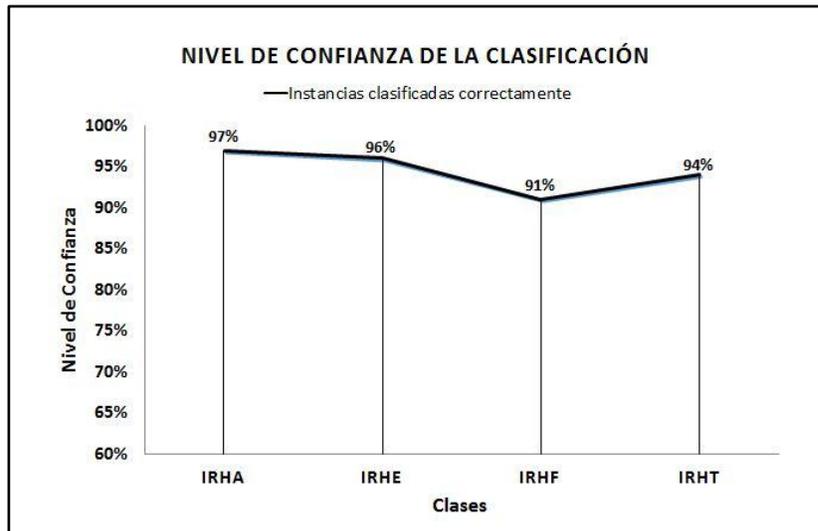


Figura 21: Representación del nivel de confianza de las clases del modelo con 200 instancias

Validación del patrón ideal de la personalidad

Para validar la calidad del patrón ideal de la personalidad propuesto a partir de su aplicabilidad se consultaron expertos con el objetivo de comprobar el nivel de veracidad de la propuesta. Los expertos fueron seleccionados a partir de los siguientes aspectos:

- Experiencia de más de cinco años en proyectos de desarrollo del CEIGE.
- Cargos o puestos directivos desempeñados.
- Prestigio en el colectivo de trabajo.

Fueron seleccionados un total de 12 expertos para lograr una representación de cada uno de los departamentos del CEIGE y direcciones que lo componen, con el objetivo de garantizar que al menos 3 de ellos pudieran expresar y reconocer las características de sus subordinados. Esta consulta permitió aprovechar la sinergia del debate en el grupo de expertos seleccionados, mediante el uso de flujos de retroalimentación. El criterio de estos expertos resultó de relevante importancia para validar la propuesta de patrón ideal de la personalidad

Esta forma de validación resultó especialmente significativa ya que permitió identificar los individuos que a consideración de los expertos son los que más se destacan en las labores productivas y los que más resultados han alcanzado en los proyectos del centro. La propuesta de la selección de estos individuos quedó abierta para profesionales y estudiantes, pero se enmarcó en el período analizado en la investigación (2011-2013). A los expertos se les entregó un listado con los nombres y apellidos de los 50 individuos que componen la muestra. Los expertos estudiaron las personas de la muestra e incluso conferenciaron entre ellos para emitir criterios.

Con un nivel de coincidencia de no menos de 3 expertos se obtuvo un total de 8 individuos considerados los de más alto rendimiento entre la muestra. La representación de estos resultados se observa en la Figura 22, representando esta selección un 16 % de la cantidad total de individuos analizados.



Figura 22: Representación de la opinión de los expertos con respecto a la muestra

En función de validar el patrón fue necesario verificar en el mismo conjunto de entrenamiento, cuántos individuos cumplían con las reglas definidas en el patrón ideal (Ver Tabla 13) para establecer las relaciones entre las clases.

Se sometieron al examen los 26 individuos que resultaron ser los que presentaron alto rendimiento en al menos dos de los cuatro índices de un total de 50 personas. Este análisis permitió identificar que de los 26 seleccionados inicialmente, solo 9 cumplieron con las reglas de decisión generadas para al menos tres clases en su mejor criterio. De ellos solo 7 cumplieron con las reglas propuestas en el patrón ideal de la personalidad para las clases IRHA, IRHE, IRHF e IRHT con los criterios Aprovechado, Alto, En_tiempo y Alto respectivamente.

En la Figura 23 se representan los resultados de esta búsqueda en función de la correspondencia con las reglas de decisión identificadas para el patrón ideal de la personalidad.

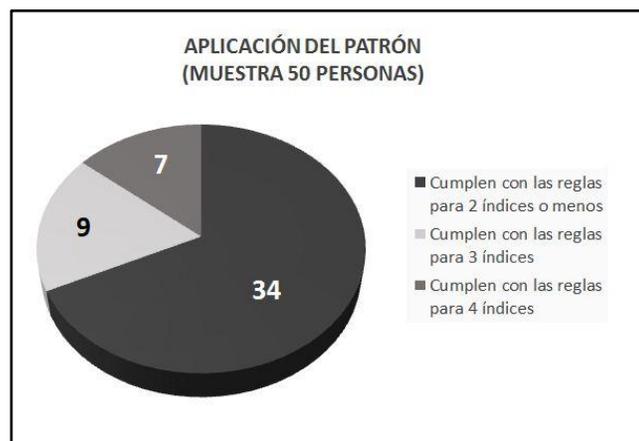


Figura 23: Representación de la aplicación del patrón ideal de la personalidad en la muestra

Al revelar los nombres de los 7 individuos que cumplieron con las reglas propuestas por el patrón, se determinó que coincidieron 6 de ellos con los nombres seleccionados por los expertos, representando un

86 % de coincidencia, considerado este por ciento como aceptable según el criterio expresado por (Grau 2012).

Los resultados anteriormente expuestos son alentadores, ya que permiten identificar un patrón ideal de la personalidad con un mayor nivel de precisión.

El 100% de los individuos que cumplieron con las reglas generadas para el patrón presentaron una escala de responsabilidad evaluada de alta, a consideración de la autora de esta investigación uno de los rasgos predictores de la productividad en el trabajo. Sin embargo, se identificó que otros rasgos se relacionan con aspectos específicos del rendimiento laboral como el aprovechamiento: Confianza, Autodisciplina, Obediencia, Curiosidad Intelectual y Asertividad. Por lo que se concluye que, para que exista un alto aprovechamiento en el trabajo es necesario ser evaluado en los rasgos anteriores con una escala alta, de lo contrario hay grandes probabilidades que el aprovechamiento no sea el más idóneo.

Por otro lado, algunos hábitos y comportamientos descritos para los individuos con altos niveles de ansiedad y depresión no aseguran el cumplimiento del patrón. En contraste, el hábito de una persona con bajos niveles de ansiedad y depresión, tienden a apoyar los valores definidos para el resto de los rasgos del patrón propuesto.

El escoger un conjunto de entrenamiento para la validación totalmente diferente al conjunto al que se le aplicó el algoritmo de Minería de Datos seleccionado resulta provechoso, ya que no condiciona los resultados alcanzados. De lo contrario, esta opción puede proporcionar una estimación demasiado optimista del comportamiento del clasificador, al evaluarlo sobre el mismo conjunto sobre el que se construyó el modelo.

Análisis del impacto económico y social de la propuesta

El análisis del impacto económico y social se encuentra relacionado al escenario real de aplicación de la investigación.

Desde el punto de vista individual el resultado de la relación entre los rasgos de la personalidad y el rendimiento laboral de los RRHH permitirá al individuo realizar un proceso de autoevaluación con respecto a su desempeño y conocer en qué medida influyen estos rasgos en sus resultados.

En equipos de proyectos informáticos, contribuirá a predecir los valores de los índices de rendimiento laboral de sus integrantes, facilitará a los directivos conocer las características de sus subordinados, apoyará la creación de estrategias para aumentar la productividad individual y por equipos, facilitará la asignación adecuada de tareas y responsabilidades y favorecerá el alcance de evaluaciones justas en la medición del desempeño laboral de los RRHH.

Por otra parte, económicamente la aplicación de la propuesta determinará una predicción temprana de comportamientos que puedan afectar el rendimiento en los subordinados con el fin de atenuar pérdidas materiales, atrasos o incumplimientos en los objetivos de los proyectos informáticos.

La aplicación del conocimiento descubierto en esta investigación, contribuirá a identificar patrones ideales de la personalidad entre los RRHH de proyectos informáticos a partir de la relación existente entre los rasgos de la personalidad y el rendimiento laboral. A nivel organizacional, la utilización de los resultados obtenidos apoyará la toma de decisiones en los procesos de Gestión de RRHH. Además, favorecerá la satisfacción de los individuos para el alcance de los objetivos organizacionales y en función de estos, la planificación de sus recursos.

Conclusiones del capítulo

De la validación del conocimiento descubierto en el presente capítulo, se arribó a las siguientes conclusiones:

- Los rasgos relevantes fueron evaluados por los expertos en relación a su importancia con respecto a los índices de rendimiento laboral con un alto grado de aceptación, alcanzándose un 91 % de coincidencia con los arrojados por el algoritmo J4.8.
- El por ciento de exactitud promedio obtenido (94.35 %) en las reglas de decisión después de su aplicación en un conjunto de entrenamiento de 50 personas certificó el nivel de precisión alcanzado en la clasificación de cada una de las clases.
- Teniendo en cuenta que la eficiencia de los resultados de un algoritmo se mide de acuerdo al tiempo de ejecución empleado en la construcción del modelo y al número de instancias clasificadas correctamente sobre el conjunto de entrenamiento, se concluye que existe una alta confiabilidad en los resultados. Sustentado por el tiempo de ejecución alcanzado que es de 0.012 segundos y el promedio de instancias clasificadas correctamente que es superior al 94 %.
- La aplicación del patrón arrojó como resultado un 86 % de veracidad, sustentado por la selección por parte de los expertos de 6 de los 7 individuos que cumplieron con las reglas.

CONCLUSIONES

Luego de realizada la presente investigación se arribó a las siguientes conclusiones:

- Se construyó una base de casos que contiene la información referente a la evaluación de los 30 rasgos de la personalidad, mediante la aplicación del instrumento NEO-PI de los individuos del CEIGE que fueron seleccionados en el estudio. La utilización de esta base de casos permitirá el acceso a los datos y su integración con la herramienta GESPRO o diferentes aplicaciones de la universidad para fomentar una mejor selección del personal para integrar proyectos informáticos.
- Se definió un proceso para descubrir conocimiento a partir de la relación existente entre los rasgos de la personalidad y el rendimiento laboral de los RRHH, obteniéndose como resultado 22 rasgos relevantes y 76 reglas de decisión. Este resultado contribuirá a la toma de decisiones en la selección de individuos para integrar proyectos y conformar equipos eficientes, a partir de la caracterización de su comportamiento independientemente del entorno productivo donde se desarrolle.
- El conocimiento obtenido, expresado mediante las reglas generadas mostró un 98 % de exactitud al clasificar un conjunto de entrenamiento de 200 personas y un 94 % al ser aplicada en una muestra de 50 personas. Estos resultados demuestran que el descubrimiento de conocimiento propuesto a partir de la relación rasgos de la personalidad y el rendimiento laboral presenta una validez aceptable.
- Se corrobora que los rasgos que más afectan a los cuatro índices de rendimiento están asociados en su mayoría a los tipos de personalidad: Extraversión, Apertura a la experiencia y Responsabilidad.
- Se conformó el patrón ideal de la personalidad a través de los resultados arrojados por el algoritmo de clasificación. Este patrón permitirá agregar valor a los elementos utilizados en los procesos de la Gestión de RRHH de la UCI.

RECOMENDACIONES

Se recomienda para futuras investigaciones:

- Continuar recolectando datos referentes a los individuos seleccionados como el rol, competencias genéricas y específicas en función de seguir fortaleciendo el conocimiento para la Selección del Personal Idóneo para la conformación de equipos.
- Construir un Sistema de aprendizaje automatizado a partir del descubrimiento del conocimiento propuesto en la investigación.
- Integrar el patrón ideal de la personalidad a la herramienta GESPRO para apoyar la toma de decisiones en proyectos informáticos.

REFERENCIAS BIBLIOGRÁFICAS

- Acuña, S. (2002). Capabilities-Oriented Integral Software Process Model. Unpublished. España, Universidad Politécnica de Madrid. Tesis presentada en opción al grado científico de Doctor en Ciencias Técnicas.
- Acuña, S. y N. Juristo (2004). Assigning people to roles in software projects Software: Practice and Experience. Madrid, España Vol. 34 (7), ISSN:0038-0644: 675-696.
- Aluja, A. y A. Blanch (2002). Análisis descriptivo y comparativo del 16PF-5 en muestras americanas y española. Boletín de Psicología. España., ISSN: 0212-8179. Vol. 74: 27-38.
- Allport, G. (1968). The Person in Psychology.
- Amón, I. (2009). Hacia una Metodología para la Selección de Técnicas de Depuración de Datos. Revista Avances en Sistemas e Informática. Vol 6(1): 6-9.
- Amón, I. (2010). Guía Metodológica para la selección de técnicas de depuración de datos. Facultad de Minas, Escuela de Sistemas Medellín. Medellín, Colombia, Universidad Nacional de Colombia: 120.
- André, A. M. (2009). Modelo para la asignación de recursos humanos a equipos de proyectos de software. Facultad de Ingeniería Informática. La Habana, Cuba, Instituto Superior Politécnico "José Antonio Echeverría". Tesis presentada en opción al grado científico de Doctor en Ciencias Técnicas.
- Belbin Associates (2010). Home to belbin® team roles Belbin Copyright information: 25 de abril del 2011. De <http://www.belbin.com/rte.asp?id=2017>.
- Belbin, M. (2010). Team Roles at Work Butterworth Heinemann. 2da Edición: Consultado el 18 de enero del 2014. De <http://www.belbin.com/rte.asp?id=2029>.
- Bello, R., Z. García, et al. (2002). Aplicaciones de la inteligencia artificial. Guadalajara, Jalisco, México, Ediciones de la Noche: 495.
- Bermúdez, M., G. Pérez, et al. (2003). Psicología de la personalidad: Teoría e investigación. Personalidad. 2: 16.
- Berzal, J. (2013). Herramientas de Minería de Datos: Introducción a KNIME. Granadas Departamento de Ciencias de la Computación e IA, Universidad de Granadas: Consultado el 5 de junio del 2013. De <http://elvex.ugr.es/decsai/intelligent/workbook/D2011%2020KNIME.pdf>.
- Betancourt, P. y R. Pacheco (2014). Aplicación de un algoritmo de clasificación supervisada mediante una herramienta para el análisis de la influencia de los rasgos de personalidad en el rendimiento laboral. La Habana, Universidad de las Ciencias Informáticas. Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas: 80.
- Briggs, K. y k. Myers (2004). Introducción al Type (MBTI): Una guía para entender los resultados de su evaluación Myers-Briggs Type Indicador. Sexta ed. Consulting Psychologists Press. California, USA.
- Broitman, E. y I. Escobar (2007). Minería de Datos. México, Ciencias Computacionales ITESM.
- Cano de Amo, J. (2004). Reducción de Datos basada en Selección Evolutiva de Instancias para Minería de Datos. Departamento de Ciencias de la Computación e Inteligencia Artificial. Granada, Universidad de Granada. Tesis presentada en opción al grado científico de Doctor en Informática. Tesis Doctoral parcialmente subvencionada por la Comisión Interministerial de Ciencia y Tecnología con el proyecto TIC 2002-04036-C05-0.

- CEDHAP (2014). *Terapia Psicologica.com.mx*[Online]. Todos los derechos reservados. Centro de Desarrollo de Habilidades Psicolingüísticas. México D.F.: Consultado el 25 de febrero del 2014. De <http://TerapiaPsicologica.com.mx>.
- Colectivo de autores (2014). La técnica del árbol para la toma de decisiones. *Magazine Estrategia*, ISSN 978-1-85617-8006. Vol. 5: Consultado el 19 de abril del 2014. De <http://www.estrategiamagazine.com/administracion/la-tecnica-del-arbol-para-la-toma-de-decisiones-efectivas-herramientas/>.
- Colectivo de autores Universia (2012). Los 10 factores que afectan la productividad empresarial. *Revista Española UNI>ERSIA* Vol 1(2): Consultado el 27 de mayo del 2014. De <http://noticias.universia.es/en-portada/noticia/2012/2012/2019/989711/989710-factores-afectan-productividad-empresarial.html>.
- Corso, C. (2009). Aplicación de algoritmos de clasificación supervisada usando Weka. Universidad Tecnológica Nacional, Facultad Regional Córdoba: Consultado el 8 de septiembre del 2014. De http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos_labsis/cynthia/CNIT_2009_Aplicacion_Algoritmos_Weka.pdf.
- Corso, C. (2010) Técnica de Análisis de Datos en Weka. Vol. 1, 33-51. Consultado el 33 de marzo del 2014. De <http://www.giaa.inf.uc2013m.es/miembros/lmarti/media/master-intel%2013Btutorial-selectividad.pdf>.
- Costa, P. y R. McCrae (1992). Normal Personality Assesment in Clinical Practice: The NEO Personality Inventory. *Psychological Assessment* Vol. 4(1): 5 - 13.
- Criado, L. (2012). Psicología de la personalidad y diferencial. Psicóloga Especialista en Psicología Clínica. Centro Documentación de Estudios y Oposiciones (CEDE). 2da Edición: Consultado el 5 de junio del 2014. De http://www.pir.es/temas_muestra_2011/2007.PERSONALIDADYDIFERENCIAL.pdf.
- Chapman, A. (2011). Personality theories, types and tests: Consultado el 27 de abril del 2013. De <http://www.businessballs.com/personalitystylesmodels.htm>.
- Chiavenato, I. (1994). Administración de Recursos Humanos. Bogotá, Colombia, Editorial Presencia. Vol. 1.
- Chiavenato, I. (2001). Administración de Recursos Humanos. McGraw Hill/Interamericana Editores, S.A. L. S. Arévalo. Santafé de Bogotá, Colombia. 5ta Edición: 362.
- Eyzaguirre, N. (2003). Seminario “Responsabilidad Social Empresarial”. Intervención del Ministro de Hacienda en seminario organizado por la Fundación Pro Humana y PNUD. Chile.
- Fayyad, U., Shapiro, et al. (1996). From data mining to knowledge discovery in databases. AAAI Press/The MIT Press (Artificial Intelligence Magazine). American Association for Artificial Intelligence, Journal Data Mining and Knowledge Discovery. Vol. 17: 37-54.
- Feist, J. y G. Feist (2009). Theories of Personality. M. H. Primis. New York. 7ma Edición.
- Feldman, R. (2009). Psicología con aplicaciones en países de habla hispana. Mc Graw Hill. ISBN: 9701053486. 6ta Edición.
- Feldt, R., R. Torkar, et al. (2010). Links Between the Personalities, Views and Attitudes of Software Engineers. *Information & Software Technology* Vol. 56: 611-624.
- Fernández, S. (2008). Utilidad del Test de personalidad MBTI en el campo de la Educación. *Psicología de la Personalidad* No. 13: 4-8.
- García, J. y L. Molina (2006). Técnicas de análisis de datos: 12. Consultado el 18 de junio del 2014. De <http://www.giaa.inf.uc2013m.es/docencia/II/ADatos/apuntesAD.pdf>.

- García, M. (2008). La evaluación por competencias en la educación superior. Universidad de Barcelona. Competences assessment 2: 2-16.
- Gervilla, G., L. Jiménez, et al. (2009). La metodología del Data Mining. Una aplicación al consumo de alcohol en adolescentes Adicciones. Revista de sociodrogalcohol. Vol. 21: 65-80.
- Goleman, D. (1999). La inteligencia emocional en la empresa. Vergara Editores. Buenos Aires, Argentina. Vol. 1. ISBN: 9501519708.
- González, E., Z. Pérez, et al. (2007). Obtención de patrones y reglas en el proceso académico de la Universidad de las Ciencias Informáticas utilizando técnicas de minería de datos. La Habana, Cuba, Instituto Superior Politécnico José Antonio Echeverría (CUJAE).
- González, J. (1987). Psicología de la personalidad. Madrid, España. Editorial Biblioteca Nueva. ISBN: 84-7030-306-6.
- Grau, R. (2012). La eficiencia en la graduación universitaria analizada con descubrimiento de conocimientos en la base de estudiantes de la Universidad Central de Las Villas.
- Guillén, C. y R. Bozal (2000). Psicología del Trabajo para las Relaciones Laborales. Madrid, España, McGraw Hill/Interamericana de España, S.A.U.
- Gutiérrez, M. (2003). Un Modelo para la Toma de Decisiones usando Razonamiento Basado en Casos en condiciones de Incertidumbre. Revista Investigación Operacional. Vol 2. No. 1.
- Hall, M. (2003). Benchmarking Attribute Selection Techniques for Data Mining. Journal IEEE Transactions on Knowledge and Data Engineering Vol. 15. No. 3: Consultado el 20 de agosto del 2014. De <http://www.cs.waikato.ac.nz/~mhall/HallHolmesTKDE.pdf>.
- Herreras, E. (2005). Modelo alternativo de evaluación de la personalidad: Modelo de los Cinco Factores, Modelo 16 PF y otros. Revista Internacional On-line, Órgano Oficial de expresión de la Fundación OMIE, Universidad de León. Instituto de Psicoterapia, Manuel Allende, Bilbao, España. Vol.4, ISSN: 1579-3516.
- Jacas, D. (2009). Aplicación de la minería de datos para la exploración y detección de patrones delictivos. La Habana, Universidad de las Ciencias Informáticas, Facultad 8. Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas.
- Jano, D. y S. Ortiz (2005) Determinación de los factores que afectan al rendimiento académico en la educación superior. XII Jornadas de la Asociación de Economía de la Educación, Consultado el 3 de enero del 2014. De <http://www.economicsofeducation.com/wp-content/uploads/oviedo2005/P2014.pdf>.
- Jiménez, P. (2012). Base de Casos para estimar la duración de un proyecto de desarrollo de software. La Habana, Cuba, Universidad de las Ciencias Informáticas. Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas.
- Jung, C. (1971). Psychological types. In R. F. C. Hull, The collected works of Carl Gustav Jung Princeton. NJ: Princeton University Press. Vol 6, 2da Edición, 608: Consultado el 5 de junio del 2014. De <http://sbpa.org.br/portal/wp-content/uploads/2012/2008/volume-2016.pdf>.
- Jung, C. G. (1971). Psychological types. In R. F. C. Hull, The collected works of Carl Gustav Jung Princeton Vol 6.(NJ: Princeton University Press).
- Keirse, D. (1996). Keirse Temperament Sorter: Consultado el 6 de mayo del 2014. De <http://www.keirse.com/sorter/register.aspx>.

- Kim, W., B. Choi, et al. (2003). A Taxonomy of Dirty Data. Data mining and knowledge discovery 7(1): 81-99.
- KNIME (2013). KNIME: Consultado el 5 de marzo del 2014. De <http://www.knime.org/>.
- Kohavi R y G. John (1997). Wrappers for feature subset selection. Journal of Educational Psychology Vol. 97: 273-324.
- Landeta, J., C. Ynzunza, et al. (2011). Factores que afectan el desempeño académico de los estudiantes de nivel superior en Rioverde. Revista de Investigación Educativa. Vol. 12.
- Lingjaerde, R. y H. O. Engvik (2001). Personality structure in patients with winter depression, assessed in a depression - free state according to the five - factor model of personality.
- López, B. (2011). Limpieza de Datos: Reemplazo de valores ausentes y Estandarización. Facultad de Matemática y Computación. Santa Clara, Cuba, Universidad Central "Marta Abreu" de Las Villas. Tesis presentada en opción al grado científico de Doctor en Ciencias Técnicas: 45.
- Lugo, G. (2012). Modelo para el control de la ejecución de proyectos basado en indicadores y lógica borrosa. Facultad 5, Laboratorio de Gestión de Proyectos. La Habana, Cuba, Universidad de las Ciencias Informáticas. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos: 79.
- Luján, I. (2009). Selección de variables mediante programación cuadrática. Departamento de Ingeniería Informática. Madrid Universidad Autónoma de Madrid. Trabajo final presentado para la obtención del título Máster en Ingeniería Informática y de Telecomunicaciones (Inteligencia Computacional): Consultado el 20 de mayo del 2014. De https://repositorio.uam.es/xmlui/bitstream/handle/10486/10022/51408_Rodriguez_Lujan_Irene.pdf?sequence=10481.
- Martineaux, K. (2008). Comparación de algoritmos de clasificación y agrupamiento aplicando técnicas de Minería de datos. La Habana, Cuba, Universidad de las Ciencias Informáticas. Trabajo de diploma para optar por el título de Ingeniero en Ciencias Informáticas.
- McCrae, R. y P. Costa (1985). Openness to experience. En R.Hogan y W.H. Jones (Eds), *Perspectives in Personality*. Greenwich, CT: JAI Press Vol 1.
- Moorhead, G. y R. Griffin (1998). *Organizational behavior: managing people and organizations*. Boston, Houghton Mifflin. 5ta Edición, ISBN:0395841968: 662.
- Morate, G. (2013). Manual de WEKA, Universidad de Waikato, Hamilton, Nueva Zelanda: Consultado el 20 de junio del 2014. De <http://www.metaemotion.com/diego.garcia.morate/download/weka.pdf>.
- Muñoz, A. (2014). Las cinco grandes dimensiones de la personalidad. Consultado el 17 de abril del 2014. Desde <http://motivacion.about.com/od/psicologia/a/Las-Cinco-Grandes-Dimensiones-De-La-Personalidad.htm>.
- Pacheco, E. (2012). Modelo de evaluación del rendimiento en los proyectos de desarrollo de software. 3er Congreso Iberoamericano de Proyectos. CIIPRO 2012. Mar del Plata, Argentina. Vol.4: 1-13.
- Pacheco, E. y Y. Betancourt (2014). Propuesta de aplicación del algoritmo GENRUL bajo el enfoque de aprendizaje supervisado en el análisis de la influencia de los rasgos de la personalidad en el rendimiento. 9na Peña Tecnológica. La Habana: 12.
- Pascarella, E. y P. Terenzini (1983). Predicting voluntary freshman year persistence/withdrawal behavior in a residential university: a path analytic validation of Tinto's model. Journal of Educational Psychology Vol 75(2): 215-226.

- Pedraza, E., G. Amaya, et al. (2010). Desempeño laboral y estabilidad del personal administrativo contratado de la Facultad de Medicina de la Universidad del Zulia. Revista de Ciencias Sociales. Vol.16 No.3.
- Pérez, A. (2002). Memoria Organizacional Basada en Casos. Revista de Ciencia e Tecnología Política e Gestao para a Periferia (RECITEC) Vol. 6 (No.1): 22-39, ISSN 1415-3262. Consultado el 1417 de junio del 2014. De <http://www.aperez.mx/recitec2002.pdf>.
- Pérez, E., M. Cupani, et al. (2004). Adaptación del Inventario de Personalidad 16PF-IPIP a un Contexto de Orientación. Estudio Preliminar. Córdoba, Argentina, Laboratorio de Evaluación Psicológica y Educativa, Facultad de Psicología, Universidad Nacional de Córdoba. ISSN N ° 1515 - 1867.
- Pérez, H. (2007). Aplicación de métodos de selección de atributos para determinar factores relevantes en la evaluación nutricional de los niños. Facultad de Ciencias Médicas, Gaceta Médica Espirituana Retrieved 10 de abril, 2014, from http://bvs.sld.cu/revistas/gme/pub/vol.9.%281%29_01/p1.html.
- Pinto, J. (2007). Sistemas de gestión de competencias basados en capacidades y recursos. Gestión de Competencias Vol.1: 3-27.
- Piñero, P. y et.al (2010). Paquete de Herramientas para la Gestión de Proyectos GESPRO.
- PMI (2009). Guía de los Fundamentos para la Dirección de Proyectos (Guía del PMBOK®). 4ta Edición [online]. ISBN:978-1-933890-72-2, EE.UU, 125-368.
- Porta, Z. (2006). Técnicas cuantitativas para la extracción de términos en un corpus, Consultado el 30 de mayo del 2014. De <http://arantxa.ii.uam.es/~jporta/uji06/uji.slides.pdf>.
- Pueyo, A. (1996). Manual de Psicología Diferencial. Mc Graw-Hill, Interamericana de España, S.A., LVEL S.A, Madrid. ISBN: 8448107926: 640.
- RAE (2001). Real Academia Española. Diccionario de la lengua española. 22ma Edición Consultado el 5 de junio del 2014. De <http://www.rae.es/recursos/diccionarios/drae>.
- Rajendran, M. (2005). Analysis of team effectiveness in software development teams working on hardware and software environments using Belbin Self-perception Inventory. Journal of Management Development Vol. 24(8): 738-753.
- RapidMiner (2014). RapidMiner, Consultado el 5 de mayo del 2014. De <http://rapidminer.com/>.
- Ril, V. (2012). Descubrimiento de conocimiento a partir de lecciones aprendidas documentadas en los procesos de cierre de proyectos informáticos. Facultad 5, Laboratorio de Gestión de Proyectos. La Habana, Universidad de las Ciencias Informáticas. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos: 80.
- Robbins, S. (2004). Comportamiento organizacional: Conceptos, controversias y aplicaciones. Editorial Prentice Hall Hispanoamericana, México. 10ma Edición.
- Sánchez, J. (2001). Arquitectura distribuida de control para sistemas con capacidades Data Mining, Consultado el 27 de marzo del 2014. De <https://serviciosgate.upm.es/tesis/tesis/3286>.
- SEI (2010). CMMI for Dev v1.3. Pittsburgh: Carnegie Mellon University.
- Servente, M. (2002). Algoritmos TDIDT aplicados a la minería de datos inteligente. Facultad de Ingeniería. Buenos Aires, Universidad de Buenos Aires. Ingeniería informática: 75.
- Sierra, F. (2013). Base de casos y motor de inferencia para apoyar el aprendizaje del diseño conceptual de bases de datos. La Habana, Cuba, Universidad de las Ciencias informáticas. Trabajo de diploma para optar por el título de Ingeniero en Ciencias informáticas.

- Silvente, V. (2013). Cómo aplicar árboles de decisión SPSS. Vol. 6 (1): 22. Consultado el 22 de junio del 2014. De <http://www.ub.edu/ice/reire.htm>.
- Standish Group, I. (2008). Trends report: Trends in optimization. Boston, Standish Group International Inc: De <http://www.standishgroup.com>.
- Stevens, K. y et al (2002) Roles de Belbin. Consultado el 27 de febrero del 2014. De <http://www.radford.edu/~kstevens2012/>.
- Stiven, R. (2012). Modelo para la evaluación de la composición de equipos de proyectos informáticos. Facultad 5, Laboratorio de Gestión de Proyectos. La Habana, Universidad de las Ciencias Informáticas. Trabajo final presentado en opción al título de Máster en Gestión de Proyectos Informáticos: 83.
- WEKA (2013). Weka, Herramienta del Grupo de Aprendizaje de la Universidad de Waikato, Consultado el 5 de mayo del 2014. De <http://www.cs.waikato.ac.nz/ml/weka/>.

GLOSARIO DE TÉRMINOS

Entropía: Es la cantidad de información que se espera observar cuando un evento ocurre según una distribución de probabilidades. Mide la incertidumbre dada una distribución de probabilidades. Si tomamos un conjunto con elementos positivos y negativos, la entropía variará entre 0 y 1.

Ganancia: Información del conjunto menos la que aporta el atributo X. Cuanto mayor sea, menor es la cantidad de información que aporta X, es decir, es un suceso muy probable, lo que implica que sea un buen candidato como atributo representativo del conjunto. Es la diferencia entre la entropía de un nodo y la de uno de sus descendientes.

Ganancia de información: propiedad estadística que mide que mide qué tan bien un atributo dado separa el conjunto de entrenamiento conforme a las clases.

Información de la división: representa la información potencial generada al dividir un atributo en un número n de subconjuntos.

Proporción de la ganancia (técnica del Gain Ratio): es una medida basada en información que considera diferentes números y diferentes probabilidades de los resultados de las pruebas.

ANEXOS

Anexo 1. Reglas de decisión generadas que fueron seleccionadas por el nivel de precisión para las clases IRHA, IRHF e IRHT.

Árbol 1- Se generaron para la clase IRHA un total de 25 reglas de decisión descritas a continuación:

Regla 1. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal y además con simpatía normal, autodisciplina alto y liberalismo alto, entonces la predicción es que el índice “IRHA es Sub_utilizado”.

Regla 2. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina alto, liberalismo normal, así como valores éticos normal y ansiedad alto, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 3. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina alto, liberalismo normal, así como valores éticos normal y ansiedad normal, entonces la predicción es que el índice “IRHA es Sub_utilizado”.

Regla 4. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal y además con simpatía normal, autodisciplina normal y orden alto, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 5. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina normal, orden normal y valores éticos normal, así como liberalismo alto y depresión bajo, entonces la predicción es que el índice “IRHA es Sub_utilizado”.

Regla 6. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina normal, orden normal y valores éticos normal, así como liberalismo alto, depresión normal e intereses artísticos normal, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 7. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina normal, orden normal y valores éticos normal, así como liberalismo normal y ansiedad bajo, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 8. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina normal, orden normal y valores éticos normal, así como liberalismo normal, ansiedad normal y nivel de actividad bajo, entonces la predicción es que el índice “ IRHA es Sub_utilizado”.

Regla 9. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina normal, orden normal y valores éticos normal, así como liberalismo normal, ansiedad normal y nivel de actividad alto, entonces la predicción es que el índice “IRHA es Sobrecargado”.

Regla 10. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina normal, orden normal, así como valores éticos normal y liberalismo bajo, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 11. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina normal, orden normal y valores éticos alto, así como depresión normal y liberalismo bajo, entonces la predicción es que el índice “IRHA es Sub_utilizado”.

Regla 12. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal, simpatía normal y además con autodisciplina normal, orden normal y valores éticos ninguno, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 13. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal y además con simpatía alto y ansiedad bajo, entonces la predicción es que el índice “IRHA es Sub_utilizado”.

Regla 14. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal y además con simpatía alto y ansiedad alto, entonces la predicción es que el índice “IRHA es Sobrecargado”.

Regla 15. Si el individuo es evaluado con los rasgos amigabilidad normal, confianza normal y simpatía bajo, entonces la predicción es que el índice “IRHA es Sub_utilizado”.

Regla 16. Si el individuo es evaluado con los rasgos amigabilidad normal y confianza bajo, entonces la predicción es que el índice “IRHA es Sobrecargado”.

Regla 17. Si el individuo es evaluado con los rasgos amigabilidad alto, intereses artísticos normal y liberalismo normal, entonces la predicción es que el índice “IRHA es Sub_utilizado”.

Regla 18. Si el individuo es evaluado con los rasgos amigabilidad alto, intereses artísticos alto y además con confianza alto y simpatía normal, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 19. Si el individuo es evaluado con los rasgos amigabilidad alto, intereses artísticos alto y además con confianza alto, simpatía alto, así como depresión bajo y orden alto, entonces la predicción es que el índice “IRHA es Sobrecargado”.

Regla 20. Si el individuo es evaluado con los rasgos amigabilidad alto, intereses artísticos alto y además con confianza alto, simpatía alto, así como depresión bajo y orden normal, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 21. Si el individuo es evaluado con los rasgos amigabilidad alto, intereses artísticos alto y además con confianza alto, simpatía alto y depresión normal, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 22. Si el individuo es evaluado con los rasgos amigabilidad alto, intereses artísticos alto y además con confianza alto, simpatía alto y depresión ninguno, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 23. Si el individuo es evaluado con los rasgos amigabilidad alto, intereses artísticos alto y además con confianza normal y autodisciplina bajo, entonces la predicción es que el índice “IRHA es Sub_utilizado”.

Regla 24. Si el individuo es evaluado con los rasgos amigabilidad alto, intereses artísticos alto y confianza bajo, entonces la predicción es que el índice “IRHA es Aprovechado”.

Regla 25. Si el individuo es evaluado con el rasgo amigabilidad bajo, entonces la predicción es que el índice “IRHA es Sobrecargado”.

Árbol 3- Se generaron para la clase IRHF un total de 17 reglas de decisión descritas a continuación:

Regla 1. Si el individuo es evaluado con los rasgos obediencia normal, simpatía normal, asertividad alto y además con nivel de actividad bajo y liberalismo alto, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 2. Si el individuo es evaluado con los rasgos obediencia normal, simpatía normal y además con asertividad alto y nivel de actividad normal, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 3. Si el individuo es evaluado con los rasgos obediencia normal, simpatía normal y además con asertividad alto y nivel de actividad alto, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 4. Si el individuo es evaluado con los rasgos obediencia normal, simpatía normal, asertividad normal y además con intereses artísticos normal, nivel de actividad bajo y liberalismo normal, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 5. Si el individuo es evaluado con los rasgos obediencia normal, simpatía normal, asertividad normal y además con intereses artísticos normal, nivel de actividad bajo y liberalismo bajo, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 6. Si el individuo es evaluado con los rasgos obediencia normal, simpatía normal, asertividad normal y además con intereses artísticos normal y nivel de actividad normal, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 7. Si el individuo es evaluado con los rasgos obediencia normal, simpatía normal, asertividad normal y además con intereses artísticos alto y liberalismo bajo, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 8. Si el individuo es evaluado con los rasgos obediencia normal, simpatía normal y además con asertividad normal e intereses artísticos bajo, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 9. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos normal y además con liberalismo bajo, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 10. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos alto y además con confianza alto y simpatía normal, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 11. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos alto, confianza alto y además con simpatía alto, nivel de actividad bajo y asertividad normal, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 12. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos alto y además con confianza alto, simpatía alto y nivel de actividad normal, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 13. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos alto y además con confianza alto, simpatía alto y nivel de actividad ninguno, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 14. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos alto y además con confianza alto, simpatía alto y nivel de actividad alto, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 15. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos alto y además con confianza alto y simpatía bajo, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 16. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos alto y además con confianza normal y autodisciplina normal, entonces la predicción es que el índice “IRHF es En_tiempo”.

Regla 17. Si el individuo es evaluado con los rasgos obediencia alto e intereses artísticos alto y confianza bajo, entonces la predicción es que el índice “IRHF es En_tiempo”.

Árbol 4- Se generaron para la clase IRHT un total de 21 reglas descritas de decisión a continuación:

Regla 1. Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal y autodisciplina alto y además con nivel de actividad bajo, orden alto e imaginación bajo, entonces la predicción es que el índice “IRHT es Alto”.

- Regla 2.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal y autodisciplina alto y además con nivel de actividad bajo, orden alto, imaginación normal y búsqueda de emociones alto, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 3.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal y autodisciplina alto y además con nivel de actividad bajo, orden normal e imaginación alto, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 4.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal y además con autodisciplina alto y nivel de actividad normal, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 5.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal y además con autodisciplina alto y nivel de actividad alto, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 6.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal y además con autodisciplina normal e imaginación bajo, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 7.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal y además con autodisciplina normal, imaginación alto, cooperación normal y búsqueda de emociones bajo, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 8.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal, autodisciplina normal y además con imaginación alto, cooperación normal y búsqueda de emociones normal, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 9.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal, autodisciplina normal y además con imaginación normal, nivel de actividad normal y cooperación normal, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 10.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal, autodisciplina normal y además con imaginación normal, nivel de actividad alto y orden alto, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 11.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras normal, autodisciplina normal y además con imaginación normal, nivel de actividad alto y orden bajo, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 12.** Si el individuo es evaluado con los rasgos prudencia normal, aventuras alto y orden alto, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 13.** Si el individuo es evaluado con los rasgos prudencia alto, cooperación normal y además con aventuras bajo y nivel de actividad normal, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 14.** Si el individuo es evaluado con los rasgos prudencia alto, cooperación normal y además con aventuras bajo y nivel de actividad ninguno, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 15.** Si el individuo es evaluado con los rasgos prudencia alto, cooperación normal y además con aventuras bajo y nivel de actividad alto, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 16.** Si el individuo es evaluado con los rasgos prudencia alto, cooperación alto, autodisciplina alto y además con nivel de actividad bajo y aventuras normal, entonces la predicción es que el índice “IRHT es Alto”.
- Regla 17.** Si el individuo es evaluado con los rasgos prudencia alto, cooperación alto, autodisciplina alto y además con nivel de actividad bajo y aventuras bajo, entonces la predicción es que el índice “IRHT es Alto”.

Regla 18. Si el individuo es evaluado con los rasgos prudencia alto, cooperación alto, autodisciplina alto y además con nivel de actividad bajo y aventuras ninguno, entonces la predicción es que el índice “IRHT es Alto”.

Regla 19. Si el individuo es evaluado con los rasgos prudencia alto, cooperación alto y además con autodisciplina alto y nivel de actividad ninguno, entonces la predicción es que el índice “IRHT es Alto”.

Regla 20. Si el individuo es evaluado con los rasgos prudencia alto, cooperación alto y además con autodisciplina alto y nivel de actividad alto, entonces la predicción es que el índice “IRHT es Alto”.

Regla 21. Si el individuo es evaluado con los rasgos prudencia alto, cooperación alto y autodisciplina normal, entonces la predicción es que el índice “IRHT es Alto”.

Anexo 2. Encuesta aplicada a los expertos para validar la influencia de los rasgos de personalidad en el rendimiento laboral.

Usted ha sido seleccionado como experto en la presente investigación. Se necesita aplicarle una encuesta con el objetivo de identificar cuáles son los rasgos de personalidad, en su opinión, que influyen en los índices de rendimiento laboral establecidos en el GESPRO.

Pregunta sobre la investigación:

- ¿Según su criterio qué peso le daría usted a los rasgos de personalidad de acuerdo a su influencia sobre los índices de rendimiento?

El peso se le asignará un valor en dependencia del criterio del experto en cuanto a:

Si es muy importante para el índice de rendimiento.	5 puntos
Si es importante para el índice de rendimiento.	4 puntos
Si la importancia para el índice de rendimiento es media.	3 puntos
Si no es tan importante para el índice de rendimiento.	2 puntos
Si no tiene importancia alguna para el índice de rendimiento.	1 punto

Tipo de Personalidad	Rasgos de Personalidad	IRHT	IRHE	IRHA	IRHF
Neuroticismo	Ansiedad				
	Ira				
	Depresión				
	Timidez				
	Impulsividad				
	Vulnerabilidad				
Extraversión	Amigabilidad				
	Gregarismo				
	Asertividad				
	Nivel de actividad				
	Búsqueda de emociones				
	Emoción positiva				
Apertura a la experiencia	Imaginación				
	Intereses artísticos				
	Sensibilidad emocional				
	Aventurero				
	Curiosidad intelectual				
	Liberalismo				
Amigabilidad	Confianza				
	Valores éticos				
	Altruismo				

	Cooperación				
	Modestia				
	Simpatía				
Responsabilidad	Autoeficacia				
	Orden				
	Obediencia				
	Capacidad de logro				
	Autodisciplina				
	Prudencia				

Criterio de Evaluación: _____

Evaluación cuantitativa del criterio:

Criterio de Evaluación	Valores
Muy buena	5 puntos
Buena	4 puntos
Regular	3 puntos
Mala	2 puntos