



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE CIENCIAS DE LA
COMPUTACIÓN MAESTRÍA EN SISTEMAS DE
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS.**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL
TÍTULO DE MAGISTER EN SISTEMAS DE INFORMACIÓN
E INTELIGENCIA DE NEGOCIOS**

**TEMA: ANÁLISIS DE LA PERTINENCIA DEL PERFIL DE
EGRESO Y SU EFICIENCIA EN EL CAMPO LABORAL DE LOS
GRADUADOS DE LA CARRERA DE INGENIERÍA DE
SISTEMAS DE LA UNIVERSIDAD DE LAS FUERZAS
ARMADAS-ESPE UTILIZANDO TÉCNICAS DE MINERÍA DE
DATOS.**

AUTOR: POMA JAPÓN, DIANA XIMENA

DIRECTOR: DÍAZ ZUÑIGA, MAGI PAÚL

SANGOLQUÍ

2017



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA TECNOLÓGICA
UNIDAD DE GESTIÓN DE POSTGRADOS**

CERTIFICACIÓN

Certifico que el trabajo de titulación, “Análisis de la pertinencia del perfil de egreso y su eficiencia en el campo laboral de los graduados de la carrera de ingeniería de sistemas de la Universidad De Las Fuerzas Armadas-ESPE utilizando técnicas de Minería De Datos” realizado por la señora Diana Ximena Poma Japón, ha sido revisado en su totalidad y analizado por el software anti plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar a la señora DIANA XIMENA POMA JAPÓN para que lo sustente públicamente.

Sangolquí, 22 de septiembre del 2017

Magí Paúl Díaz Zuñiga

DIRECTOR

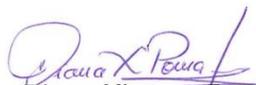


**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA TECNOLÓGICA
UNIDAD DE GESTIÓN DE POSTGRADOS
AUTORÍA DE RESPONSABILIDAD**

Yo, DIANA XIMENA POMA JAPÓN, con cédula de identidad N° 110410171-0, declaro que este trabajo de “Análisis de la pertinencia del perfil de egreso y su eficiencia en el campo laboral de los graduados de la carrera de ingeniería de sistemas de la Universidad de las Fuerzas Armadas-ESPE utilizando Técnicas de Minería de Datos” ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 22 de septiembre del 2017


Diana Ximena Poma Japón

C.C 110410171-0



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA TECNOLÓGICA
UNIDAD DE GESTIÓN DE POSTGRADOS**

AUTORIZACIÓN

Yo, Diana Ximena Poma Japón, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación “Análisis de la pertinencia del perfil de egreso y su eficiencia en el campo laboral de los graduados de la carrera de ingeniería de sistemas de la Universidad de las Fuerzas Armadas-ESPE utilizando Técnicas de Minería de Datos” cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Sangolquí, 22 de septiembre del 2017


Diana Ximena Poma Japón

C.C 110410171-0

DEDICATORIA

La concepción de este proyecto está dedicada a Dios quien ha estado conmigo en cada paso que doy, dándome fortaleza para continuar. A mi esposo Juan José y mis hijos Joaquín y Saraí quienes son pilares fundamentales en mi vida y fueron mi motivación para no rendirme. Sin su apoyo incondicional, jamás hubiese podido conseguir este logro. A mis padres por su ejemplo de lucha constante e insaciable, fomentaron en mí el deseo de superación son mi ejemplo a seguir

Diana Ximena, Poma Japón.

AGRADECIMIENTOS

Quiero agradecer a Dios quien está presente en cada paso que doy por darme la fuerza para superar obstáculos y dificultades y hoy más que nunca sigo segura de que él es fiel. A mi esposo y mis hijos por su paciencia y apoyo incondicional. A mis padres que con sus sabios consejos me enseñaron a no desfallecer ni rendirme ante nada y siempre perseverar.

A mis hermanos Nathaly, Cesar y Andrés por motivarme a seguir y estar pendientes de mis triunfos y derrotas, ustedes son parte importante en mi vida. Agradezco a mi director de Tesis MSc. Paúl Díaz Zúñiga por toda la colaboración brindada durante la elaboración de este proyecto. A los docentes por sus conocimientos impartidos y a esta prestigiosa Universidad que nos abre sus puertas para formar y preparar personas competitivas y de servicio a la comunidad.

Diana Ximena Poma Japón

ÍNDICE DE CONTENIDOS

CERTIFICACIÓN	ii
AUTORÍA DE RESPONSABILIDAD	iii
AUTORIZACIÓN	iv
DEDICATORIA	v
AGRADECIMIENTOS	vi
ÍNDICE DE CONTENIDOS	vii
ÍNDICE DE TABLAS	xii
ÍNDICE DE FIGURAS	xiii
RESUMEN	xv
ABSTRACT	xvi
CAPÍTULO I	1
ASPECTOS GENERALES	1
1. Introducción.	1
1.1. Planteamiento del problema.	1
1.2. Formulación del problema.....	3
1.3. Justificación de la Investigación.....	4
1.4. Objetivos.	5
1.4.1 Objetivo General.	5
1.4.2. Objetivos Específicos.....	5
CAPÍTULO II	7
MARCO TEÓRICO	7
2. Inteligencia de Negocios	7
2.1. Concepto de Inteligencia de Negocios.....	11
2.2. Beneficios al utilizar Inteligencia de Negocios.	12

2.3.	Proceso Business Intelligence.	13
2.4.	Componentes de Business Intelligence.	15
2.5.	Minería de Datos.....	16
2.5.1.	Datos.....	16
2.5.2.	Información.	17
2.5.3.	Conocimiento.	17
2.5.4.	Evolución Historia de la Minería de Datos.	17
2.5.5.	Concepto de Minería de Datos.	18
2.5.6.	Proceso de descubrimiento de conocimiento en minería de datos.	19
2.5.7.	Fundamentos de Minería de Datos.....	19
2.5.8.	Objetivos de la Minería de Datos.	20
2.5.9.	Fases de un Proyecto de Minería de datos.....	21
2.5.9.1.	Filtrado de datos.	21
2.5.9.2.	Selección de variables.	21
2.6.	Algoritmos de Extracción de Conocimiento.....	22
2.7.	Interpretación y evaluación.	22
2.8.	Modelos y Técnicas de explotación de datos.....	22
2.8.1.	Modelos de explotación de datos.....	23
2.8.1.1.	Supervisados o predictivos:	23
2.8.1.2.	No Sup. Descrip. o del descubrimiento del conocimiento.	23
2.9.	Técnicas de explotación de datos.....	24
2.9.1.	Redes neuronales (Neural Networks).	24
2.10.	Análisis Preliminar de datos, usando Query Tools.....	24
2.10.1.	Técnicas de Visualización.....	25
2.10.2.	Reglas de Asociación.....	25
2.10.3.	Algoritmos Genéticos.	25

2.10.4. Redes Bayesianas.....	25
2.10.5. Árbol de Decisión.....	26
2.10.6. Clustering (Agrupamiento).....	26
2.10.7. Segmentación.....	26
2.10.8. Clasificación.....	26
2.10.9. Predicción.....	27
2.11. Herramientas de Minería de Datos.....	27
2.11.1. Técnicas de verificación.....	27
2.11.2. Método de descubrimiento.....	27
2.12. Herramientas Top de Minería de Datos.....	28
2.12.1. Weka (Waikato environment for knowledge analysis).....	29
2.12.2. Características de Weka.....	29
2.12.3. Interfaz gráfica de Weka.....	30
2.12.4. Ventajas al utilizar Weka.....	30
2.12.5. Explorer.....	31
2.12.6. Experimenter.....	32
2.12.7. Knowledge Flow.....	33
2.12.8. Clementine de SPSS.....	33
2.12.9. Utilidades de Clementi.....	34
2.12.10. Knime.....	35
2.12.11. Características de KNIME.....	36
2.12.12. Knime vs Weka.....	36
2.12.13. RapidMiner.....	36
2.12.14. Componentes de RapidMiner.....	37
2.12.15. Características de RapidMiner.....	37
2.12.16. Orange.....	38

2.12.17. Características de Orange.	38
2.12.18. Componentes de Orange.....	39
2.12.19. Metodologías para el proceso de Minería de datos.	40
2.13. KDD (Knowlegde Discovery Databases).....	41
2.13.1. Las Fases del KDD.	41
2.13.2. CRISP-DM (Cross-Industry Standard Process for Data Mining). ...	42
2.13.3. Historia.....	42
2.13.4. Fases de CRISP-DM.....	43
2.13.5. Ventajas de CRISP-DM.	46
2.13.6. SEMMA.....	46
2.13.6.1. Fase de la Metodología SEMMA.	46
2.13.6.2. Particularidad de las Metodologías.	48
CAPÍTULO III.....	50
DESARROLLO DEL CASO DE ESTUDIO	50
3. Análisis de la Situación Actual de la Carrera	50
3.1. Fundamentación teórica.....	61
3.1.1. Pertinencia.	61
3.1.2. Perfil de Egreso.	61
3.1.2. Mercado o Campo Laboral.....	61
3.1.3. Inserción laboral.....	62
3.1.4. Perfil de egreso profesional.	62
3.1.5. Componentes del Perfil de Egreso Profesional.....	63
3.1.6. Validar el Perfil de Egreso.....	64
3.1.7. Desempleo, subempleo y ocupación plena.....	68
3.1.8. Mercado Laboral en Ecuador.....	69
3.1.9. Análisis de la data utilizando la Metodología CRISP-DM.....	70

3.1.10. Tipo de Investigación.	70
3.1.11. Métodos de investigación.....	70
3.1.12. Pre-procesamiento de datos.....	70
3.1.13. Técnicas de minería de datos aplicada al proyecto.	71
3.1.14. Población.	72
3.1.15. Muestra.....	73
3.1.15.1.Cálculo de la muestra.	74
3.1.16. Aplicación de la metodología CRISP-DM al caso de estudio.....	75
3.1.17. Clustering Cuadro de Caja.....	88
3.1.18. Árbol de Decisión - Visor de Árbol.	88
3.1.19. Clustering –Visor de Árbol.	88
3.2. Discusión y Resultados.....	90
CAPÍTULO IV.....	93
CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURO.....	93
4.1. Conclusiones	93
4.2. Líneas de trabajo futuro.	94
BIBLIOGRAFÍA.....	95

ÍNDICE DE TABLAS

Tabla 1. Herramientas de minerías de datos	28
Tabla 2. Metodologías para el proceso de Minería de datos	40
Tabla 3. Características que diferencian a las metodologías de minería de datos	49
Tabla 4. Capacidades técnicas de un profesional en ingeniería en sistemas, requeridas por las empresas, y ofertadas por la facultad de ingeniería en sistemas de la ESPE.....	53
Tabla 5. Empresas activas a nivel nacional que se dedican a actividades Económicas de servicio de la rama de Información y Comunicación	56
Tabla 8. Número de empresas en Quito y Guayaquil que demanda ingenieros en sistemas	58
Tabla 9. Oferta laboral para ingenieros en sistemas, ofertado por empresas ubicadas en las ciudades de Quito y Guayaquil, según cuatro portales web de empleo	58
Tabla 10. Población	72

ÍNDICE DE FIGURAS

Figura 1. Fases del Proceso de B.I.....	15
Figura 2. Interfaz principal	30
Figura 3. Ventana de Comandos	31
Figura 4. Explorer	32
Figura 5. Experimenter	32
Figura 6. Knowledge flow.....	33
Figura 7. Proporción de empresas de actividades económicas; con códigos; J6201.01; J6201.02; J6202.10; J6202.20; J6209.01; J6209.02, por ciudades que mayor número de empresas concentran.	57
Figura 8. Oferta laboral para ingenieros en sistemas, ofertado por empresas ubicadas en las ciudades de Quito y Guayaquil, según el tipo de actividades que se dedican.....	59
Figura 9. Componentes del perfil de egreso	63
Figura 10. Población	73
Figura 11. Fases del modelo CRISP-DM.....	76
Figura 12. Datos de los Graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas – ESPE	79
Figura 13. Widgets o ficheros utilizados para obtener la gráfica de los graduados en relación al campo laboral.	80
Figura 14. Selección de Datos	81
Figura 15..Algoritmo K-means	81
Figura 16. Clustering - Cuadro de Caja de los graduados en relación al campo laboral.....	82

Figura 17. Ficheros utilizados para obtener la gráfica del campo en que se encuentran laborando con mayor incidencia las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE.	83
Figura 18. Selección de la característica cargo y del objetivo variable sexo.	84
Figura 19. Ejecución del algoritmo de árbol de decisión.....	84
Figura 20. Visor de Árbol campo en que se encuentran laborando con mayor incidencia las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE.....	85
Figura 21. Ficheros utilizados para obtener la gráfica que determinara el nivel de incidencia de los graduados de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, en el mercado laboral ecuatoriano.	86
Figura 22. Selección de la característica Nacional/Extranjera en el fichero Selección de datos	86
Figura 23. Ejecución del algoritmo k-means y del algoritmo de árbol de decisión.....	87
Figura 24. Visor de Árbol Nivel de incidencia de los graduados en el mercado laboral ecuatoriano de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE	87

RESUMEN

Según la Ley Orgánica de Educación Superior, LOES, las universidades e institutos superiores deben enfocarse en la formación de aprendizajes significativos, mismos que permitan el desenvolvimiento eficaz en el campo laboral para el futuro profesional. De este modo, muchas instituciones de este nivel educativo, especialmente carreras, han desarrollado estrategias de gestión académica que buscan insertar a sus graduados a un contexto real, por lo que han generado normas y estándares posmodernos que responden a la investigación e innovación tecnológica. Por lo tanto, el presente estudio tiene el objetivo determinar la pertinencia del rol que desempeñan los graduados en las empresas que laboran, de acuerdo al Perfil de Egreso de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, mediante técnicas de minería de datos. De este modo, se pudo determinar que la mayoría de graduados en la carrera de Ingeniería en Sistemas de la Universidad de las Fuerzas Armadas ESPE son quienes ocupan mayor cantidad de plazas laborales en los diferentes campos del mercado ecuatoriano, entre estos se encuentran: Desarrollo de Software, Analistas de Sistema, Administrativa, Sistemas de Información, Consultor-Auditor Informático, Soporte Técnico y Docentes.

PALABRAS CLAVES:

- **GRADUADO**
- **CAMPO LABORAL**
- **PERFIL DE EGRESO**
- **GESTIÓN ACADÉMICA**
- **MINERÍA DE DATOS.**

ABSTRACT

According to the Organic Law of Higher Education, LOES, universities and colleges should focus on the formation of meaningful learning, which allow effective development in the field of work for the professional future. In this way, many institutions of this educational level, especially the careers, have developed academic management strategies that seek to insert their graduates in a real context, reason why they have generated postmodern norms and standards that respond to the research and technological innovation. Therefore, the present study has as objective to determine the relevance of the graduates in the companies that work, according to the Profile of Exit of the Career of Systems Engineering of the Universidad de las Fuerzas Armadas-ESPE, by means of mining techniques of data. In this way, it was possible to determine that the majority of graduateds in the Systems Engineering degree from the University of the Armed Forces ESPE occupy the largest number of jobs in the different fields of the ecuatorian market, among which are: Software, System Analysts, Administrative, Information Systems, IT Consultant-Auditor, Technical Support and Teachers.

KEY WORDS:

- **GRADUATE**
- **FIELD OF WORK**
- **EGRESS PROFILE**
- **ACADEMIC MANAGEMENT**
- **DATAMINING.**

CAPÍTULO I

ASPECTOS GENERALES

1. Introducción.

La Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE se encuentra inmersa en un contexto altamente competitivo, debido a las atractivas propuestas que ésta ofrece a los potenciales estudiantes al momento de la titulación, y no solo por el prestigio institucional de la misma.

De este modo, la carrera se ha visto en la obligación de desempeñar, con mayor eficiencia y eficacia, la gestión interna, así como de buscar la excelencia en la oferta académica y, sobre todo, el compromiso de una formación integral. Por lo tanto, la gestión se basa en realizar un seguimiento al graduado, con el fin de obtener información sobre la pertinencia que existe entre el perfil de egreso y el campo laboral en el que se desempeña el ex alumno.

Con el objetivo de llegar a conocer esta información, el presente estudio trabajará con Orange Canvas, una herramienta de minería de datos, en la cual se efectuarán los análisis de la base de datos de Seguimiento a Graduados de la Carrera de Ingeniería de Sistemas, detallando la existencia o ausencia de Pertinencia del Perfil de Egreso, así como la eficiencia de los profesionales graduados de la carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE.

1.1. Planteamiento del problema.

En el estudio realizado por (Morueta, Tejeda, & Cedeño, 2015) se determinó que el seguimiento del itinerario académico y profesional de los graduados universitarios constituye un insumo fundamental para la planificación estratégica, y permite obtener indicadores de pertinencia acerca de las carreras y planes de estudio, para el diseño y el desarrollo curricular

en universidades. (Gavilanes & Romero, 2015) señalaron que la educación superior se encarga de preparar a los graduados para tareas indeterminadas; además, no sólo tiene que proporcionar el conocimiento de herramientas y normas sino que, tiene que preparar a los estudiantes para cuestionarse, constantemente, las herramientas y normas establecidas.

La Universidad Laica Eloy Alfaro de Manabí (2015) indicó que uno de los grandes problemas en las universidades es el tiempo que transcurre desde que el estudiante inicia el estudio de un programa de tercer nivel hasta que lo termina; para cuando esto sucede las condiciones del país, del mercado y del conocimiento técnico han cambiado y son obsoletas. Así mismo, la crisis de la economía mundial, el crecimiento de la población y las transiciones acontecidas en el mercado laboral e internacional han generado altos índices de desempleo, específicamente en Latinoamérica.

Según la, OIT (Organización Internacional del Trabajo, 2016), en su informe Panorama Laboral, desde el año 2011 América Latina y El Caribe han estado afectados por una desaceleración económica. La tasa de desempleo regional promedio, que había alcanzado mínimos históricos de 6,1% en 2014, subió en el 2015 a 6,6%. En el 2016 la tasa de desempleo regional promedió 8,1%, es decir, un incremento de 1,5 puntos porcentuales. Para el 2016 este programa contuvo una estimación de la tasa de desocupación regional, para el 2017 se espera que esta tasa aumente nuevamente a 8,4%.

La Asamblea Nacional (2010) expresa que en el Ecuador el criterio de seguimiento a graduados es uno de los indicadores de evaluación normado en el Artículo 142 de la LOES, en donde se dispone:

Todas las instituciones del sistema de educación superior, pública y particular, deberán instrumentar un sistema de seguimiento a sus graduados y sus resultados serán remitidos para conocimiento del Consejo de

Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES).

Esto ha permitido que algunas universidades y escuelas ecuatorianas, como la Universidad Casa Grande, la Carrera de Educación de La Tecnológica Indoamérica, la Universidad Católica de Guayaquil, entre otras, tengan un conocimiento del grado de satisfacción de los egresados y de sus empleadores, con respecto a la calidad de formación que ofertan (Universidad Casa Grande, 2009; Dávila, 2012; Seguimiento a Graduados, 2012).

La finalidad del presente proyecto es analizar patrones de información de la base de datos de Seguimiento a Graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, y realizar un tratamiento de los datos, limpieza y normalización de los mismos, mediante CRISP-DM; además, se utilizará Data Mining para la exploración de los datos del ámbito laboral de los graduados, cuyo resultado apoyará a la consolidación de la investigación.

1.2. Formulación del problema.

¿Cuál es el nivel de pertinencia de los graduados en las empresas que laboran, según el Perfil de Egreso de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE?

El visionarse en conocer cuán pertinente es el perfil de egreso de la carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, en cuanto al desenvolvimiento laboral de sus graduados, desencadena el planteamiento de interrogantes, mismos que permitirán orientar el proceso del proyecto, dando respuesta a las mismas. Las preguntas planteadas son las siguientes:

1. ¿La metodología CRISP-DM permite adaptar, de manera concreta las necesidades del proyecto de minería de datos?

2. ¿Cuál es la pertinencia del perfil de egreso de los graduados en relación al campo laboral en el que se desenvuelven?
3. ¿Cuál es el campo laboral de mayor incidencia de acuerdo al perfil de egreso en el que se desenvuelven las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE?
4. ¿Cuál el nivel de incidencia de los graduados en el mercado laboral ecuatoriano de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE aplicando clusterización?

1.3. Justificación de la Investigación.

Esta investigación se justifica por los requisitos que impone la LOES y la sociedad actual que están enfocados en formar conocimientos, aptitudes y actitudes en los estudiantes, para que contribuyan al desarrollo económico y social del país. Este proceso se verá reflejado cuando el graduado se incorpore en el mundo laboral.

Hasta el momento se ha realizado algunos estudios a graduados de la carrera de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, sin embargo, esta investigación analizará cuán pertinente y eficiente es el perfil de egreso de los graduados que se encuentran laborando en las diferentes empresas del Ecuador, determinar el campo laboral más incidente en los graduados de sexo femenino y definir el porcentaje de graduados que prestan sus servicios en empresas ecuatorianas y extranjeras. Además, a través de técnicas de minería de datos se establecerá si los graduados se encuentran trabajando en cargos relacionados al perfil de egreso en el cual fueron formados.

Por lo tanto, se aplicará mecanismos de búsqueda de información que permita analizar patrones de datos y conocer si el perfil de egreso cumple con las necesidades del campo laboral local, regional y nacional; y

determinar indicadores que fomenten a la actualización, reforma o transformación de los perfiles de egreso de la carrera de Ingeniería de Sistemas, para mejorar la calidad de vida de los graduados y su desempeño profesional en el campo laboral. También se aplicará Minería de Datos que, a través de algoritmos supervisados o predictivos, demostrará si el análisis del Perfil de Egreso incide en el campo laboral de los graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE.

El análisis se realizará sobre la base de datos de los graduados, y los resultados obtenidos permitirán mejorar la oferta académica de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE.

1.4. Objetivos.

1.4.1 Objetivo General.

Determinar la pertinencia de los graduados en las empresas que laboran, de acuerdo al Perfil de Egreso de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, mediante técnicas de minería de datos.

1.4.2. Objetivos Específicos.

- Aplicar la metodología CRISP-DM para determinar, de manera concreta, las necesidades del proyecto de minería de datos.
- Analizar la pertinencia del perfil de egreso de los graduados en relación al campo laboral en el que se desenvuelven.
- Analizar el campo laboral de mayor incidencia, de acuerdo al perfil de egreso de las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE.

- Determinar el nivel de incidencia de los graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE en el mercado laboral ecuatoriano, aplicando clusterización.

CAPÍTULO II

MARCO TEÓRICO

2. Inteligencia de Negocios

Hace algo más de 50 años, en octubre de 1958, fue publicado un artículo en la revista *IBM Journal*, donde H.P. Luhn, un investigador de la empresa International Business Machines Corp., IBM y pionero en las ciencias de la información, utilizó el término “Business Intelligence System” para referirse a un sistema automático que acepta información en su formato original, disemina los datos adecuada y rápidamente a los lugares correctos (Luhn, 1958).

Martínez (Martínez H. , 2010) manifiesta que, según Luhn, para cumplir con estos objetivos era necesaria la aplicación de diferentes técnicas:

- Auto abstracción de documentos.
- Auto codificación de documentos.
- Creación y actualización automática de perfiles de usuarios.

Estas técnicas son de carácter estadístico para el procesamiento de los datos, pero no son eficaces si no se cuenta con facilidades en los sistemas de comunicación, así como en los medios de entrada-salida. Para Luhn, desde el punto de vista administrativo, un sistema inteligente para los negocios provee la facilidad en la comunicación que permite la conducción de un negocio, donde la inteligencia permite entender las interrelaciones de hechos pasados, de tal manera que estos guíen las acciones hacia una meta deseada.

Según su trabajo, el objetivo del sistema es: “proporcionar la información adecuada para soportar actividades específicas realizadas por individuos, grupos, departamentos, divisiones o aun unidades más grandes [...] estas actividades se deben realizar rápida y eficientemente.”

Las investigaciones de Luhn pretendían desarrollar un sistema automatizado de análisis de texto, que posibilite gestionar los documentos que se generan en una empresa. Dicho sistema necesitaba de un procesamiento y almacenamiento eficiente que permitan a los empleados realizar búsquedas sobre temas específicos, y el sistema remitía los documentos adecuados a los usuarios, de acuerdo a unos perfiles establecidos.

Para situar un marco temporal a las observaciones de Luhn, que para ese entonces el procesamiento al que se hace referencia utilizaba tanto dispositivos mecánicos como computarizados ya que estos últimos recién estaban comenzando su desarrollo, existían restricciones a nivel de procesamiento y almacenamiento.

Entonces, esta referencia puede considerarse como el punto de partida para el desarrollo de los actuales sistemas de Inteligencia de Negocios. Luhn consideró varios aspectos:

- El sistema de inteligencia de negocios incorpora datos generados desde el nivel operativo de la empresa.
- Se planteó una clara distinción de tipos de usuarios para el uso de información, es decir, la información es selectiva, puesto que los intereses de los usuarios son diferentes, dependiendo de sus requerimientos y deseos.
- La distribución de la información se puede dar como un proceso automático o como respuesta a una petición (bajo demanda). En el primer caso, la información distribuida corresponde a elementos preestablecidos. En el segundo caso la información corresponde a preguntas específicas.
- La importancia del uso de los metadatos como ente orientador: la búsqueda de la información se puede orientar mediante otro

tipo de información mucho más generalizada que da guías sobre el contenido de un documento.

- La noción de usuarios expertos que se encargaran de consultas difíciles y los usuarios de negocio.
- La importancia de la seguridad y posesión de los datos. El análisis de la propuesta de Luhn revela muchos elementos fundamentales en la construcción de las bases para los actuales sistemas de inteligencia de negocios, y podría decirse que, en general, para cualquier sistema de información. Es notable el hecho de reconocer que para la época las limitaciones técnicas que, evidentemente hasta varios lustros después con el desarrollo de las tecnologías de la información y comunicaciones, se fueron solventando.

El enfoque propuesto por Luhn considera principalmente la afectación que tiene un sistema de inteligencia de negocios sobre el nivel interno de la organización; no obstante, los sistemas de información pueden afectar el nivel interno, los niveles competitivos y de portafolio de negocios que, en términos generales, se corresponden con el entorno de la organización.

Esta consideración acerca del entorno, como fuente de información e influencia sobre la empresa, fue abordada casi simultáneamente a los planteamientos de Luhn.

(García & Martínez, 2010), citando a (Dill, 1958), señalan: “la mejor forma de analizar el entorno no es tratar de entenderlo como una colección de otros sistemas y organizaciones, sino tratarlo como información a la cual la empresa debe acceder mediante actividades de investigación”. Así mismo, exponen que este autor afirmó: “no son los proveedores ni los clientes en sí mismos los que cuentan, sino la información disponible para la empresa acerca de sus metas, intereses, condiciones bajo las cuales se relacionan con la empresa y muchos otros aspectos de su comportamiento, los que

verdaderamente ayudan a identificar sus características y expectativas para con la organización”.

Este planteamiento va en línea con los postulados y principios de la cibernética, acerca de la forma como se regulan los sistemas y, particularmente, de los efectos que tiene la información en la regulación de los mismos. Estos postulados, pese a que no se desarrollaron a la par con el concepto de Inteligencia de Negocios, tienen una estrecha relación.

Se adjudica la creación del término Inteligencia de Negocios a Richard Green, quien en el año 1966 la definió como: “la información procesada de interés para la administración acerca del presente y futuro del entorno en el cual el negocio debe operar”. El enfoque de Green se acerca mucho más al concepto de inteligencia como forma de espionaje, con una connotación evidentemente militar. En este sentido, el concepto de Inteligencia de Negocios presentó, temáticamente, una bifurcación de su orientación empresarial hacia la inteligencia, desde el punto de vista militar.

Durante los años 70 y 80, el concepto de inteligencia de negocios giró en torno a lo que se conoció como: “escaneo del entorno”, que era una actividad mucho más enfocada hacia el reconocimiento de los entornos externos e internos de las organizaciones, priorizando la actividad de mercadeo sobre las demás funciones organizacionales. Otra denominación basada en el escaneo del entorno es la de “inteligencia competitiva”, enfocada principalmente en los competidores.

En los años ochenta la orientación del discurso administrativo giró hacia el concepto de competitividad y estrategia competitiva, donde se priorizó el poder de negociación con los compradores o clientes, el poder de negociación con los proveedores o vendedores, la amenaza de nuevos competidores, la amenaza de productos sustitutos, y la rivalidad entre los competidores.

Estos elementos de análisis constituyen el modelo “Cinco fuerzas” de Porter, el cual, visto desde la óptica de los sistemas de inteligencia de negocios, permitió dividir los elementos de análisis, bajo un modelo enfocado en la estrategia y conteniendo, a la vez los diferentes niveles de afectación (Interno, Competitivo y de Portafolio de negocios).

La aplicación de este modelo permite identificar claramente qué tipo de información es la que se requiere y, a partir de ello, es posible identificar y dar manejo a las fuentes, desde las cuáles esta información debe ser provista. Desde la perspectiva técnica, para llevar a la práctica la Inteligencia de Negocios, fue necesaria la integración de varios elementos que hasta entonces estaban separados (años noventa). Entre estos elementos están las Bases de Datos de Procesamiento Transaccional, OLTP, Bases de Datos Analíticas, OLAP, Minería de Datos, Sistemas de Generación de Reportes y Visualización de Datos.

Así mismo se desarrollaron interfaces de intercambio de datos para permitir la comunicación entre múltiples sistemas operacionales, con el objetivo de integrar la información de toda la organización (IBM, 2000).

2.1. Concepto de Inteligencia de Negocios.

Se describió a la inteligencia de negocios como: “un conjunto de conceptos y métodos para mejorar el proceso de decisión utilizando un sistema de soporte basado en hechos” (Españeira, Sheldon y Asociados, 2008).

(Vassallo, 2012) Manifiesta que *Bussin Intelligence*, B.I., es un sistema que permiten transformar los datos en conocimiento para obtener una ventaja competitiva. Permite reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento

2.2. Beneficios al utilizar Inteligencia de Negocios.

(Cano J. , 2007) y (Almirón, 2013) describen tres tipos de beneficios que se pueden obtener a través del uso de inteligencia de negocios, los cuales son:

- **Beneficios tangibles:** El hecho de que se tenga disponible la información para la toma de decisiones, hará que más usuarios la utilicen para tomar decisiones y mejorar la posición competitiva, como: mejorar la adquisición de clientes y su conversión, mediante el uso de la segmentación. Reducir la tasa de abandono de clientes, incrementar su fidelidad, teniendo en cuenta cuál es su valor. Incrementar los ingresos por crecimiento de las ventas.

Los beneficios tangibles aumentan los resultados, consiguiendo que los clientes actuales compren más productos o servicios. Así mismo, éstos evitan las pérdidas producidas por las ventas de los competidores, aumentan la rentabilidad por el acceso a información detallada de productos, clientes, etc., conocen mejor cuáles son las características demográficas de la zona de influencia, hacen crecer la participación de mercado, reducen el tiempo de lanzamiento de nuevos productos o servicios, mejoran aquellas actividades relacionadas con la captura de datos, analizan la cesta de la compra y la afinidad de venta entre los productos, y facilitan la adopción de los cambios en la estrategia.

- **Beneficios Intangibles:** El hecho de que se tenga disponible la información para la toma de decisiones, hará que más usuarios la utilicen para tomar decisiones y mejorar la posición competitiva, por ejemplo: optimizar la atención a los clientes, aumentar la satisfacción de los clientes, mejorar el acceso a los datos a través de consultas, análisis, informes, información más actualizada, dotar a la información de mayor precisión, conseguir ventajas competitivas, controlar de mejor manera la información, ahorrar costes, menor

dependencia de los sistemas desarrollados, mayor integración de la información.

- Beneficios estratégicos: La formulación de estrategias especificarán a qué clientes, mercados o con qué productos dirigirse.

Entre los beneficios estratégicos se encuentra la mayor habilidad para analizar estrategias de precios, para identificar y nutrir a aquellos clientes con mayor potencial, mejorar la toma de decisiones, realizándola de forma más rápida informada y basada en hechos, mayor visibilidad de la gestión, dar soporte a las estrategias, aumentar el valor de mercado.

2.3. Proceso *Business Intelligence*.

Para comprender cómo es posible que una empresa u organización cree inteligencia de negocios con los datos que tiene y promuevan acceso a los usuarios finales de la información requerida, se toma en cuenta el siguiente proceso de BI que propone Gestipolis (gestipolis, 2004).

- FASE 1: Dirigir y Planear. Esta fase es el principio y el fin del proceso. Es el principio porque involucra la redacción de los requerimientos específicos, y es el final porque contesta preguntas que guían a otras nuevas. El proceso de BI empieza con los usuarios (Ejecutivos, Directivos, Líderes de Negocio, entre otros.) y aquí se genera las preguntas que les va a ayudar a alcanzar sus objetivos. Ejemplos de esas preguntas son: ¿cuáles son los clientes más rentables? ¿cuál es el margen de cada línea de producto?, entre otras. Estas necesidades son presentadas por diversos analistas. Esto quiere decir que, estos analistas de negocios formulan los requerimientos de los usuarios y dirigen un plan para la recolección de la información y para solución de las respuestas.

- FASE 2: Hay diversas fuentes de información dentro de una compañía. La automatización de los procesos ha creado una fuente de recursos como: puntos de ventas, ERP, CRM, SFA, aplicaciones de servicios al cliente, entre otros. Los diferentes sistemas crean, procesan y almacenan diferentes tipos de información. Este es un proceso continuo y es importante entender que los datos de esas fuentes son simplemente información y no Inteligencia. Frecuentemente los datos en crudo son incompletos y confusos. La información se convierte en inteligente a través de ser procesada y de analizada. El proceso de recolección de información es cuando las diferentes fuentes son analizadas para determinar los datos necesarios para encontrar las respuestas a las preguntas.
- FASE 3: Procesamiento de Datos. Esta fase integra los datos en crudo a un formato utilizable para el análisis. Esto puede ser posible al crear una nueva base de datos, agregando datos a bases de datos existente, o consolidando información. Esta fase, generalmente, es vista como Extracción, Transformación y Carga que ocurre en los ambientes de BI.
- FASE 4: Análisis y Producción. El grupo de análisis de negocios utiliza herramientas y técnicas para ordenar sobre los datos y crear inteligencia. El resultado final es la producción de respuestas “inteligentes”, en un contexto propio. En algunos casos es un proceso simple como la creación de un reporte, y en otros es la creación de indicadores. Tal vez en esta fase sean generados los requerimientos adicionales, pues los analistas pueden encontrar nuevas preguntas que necesiten ser contestadas.
- FASE 5: Difusión. Esta fase consiste en entregar productos inteligentes a los diversos clientes que lo requieren. Esto

básicamente implica el uso de herramientas BI para la publicación de “tableros de indicadores”, reportes o la posibilidad de tener herramientas de fácil uso, para que los mismos usuarios tengan la capacidad de revisar los datos de manera rápida y sencilla.

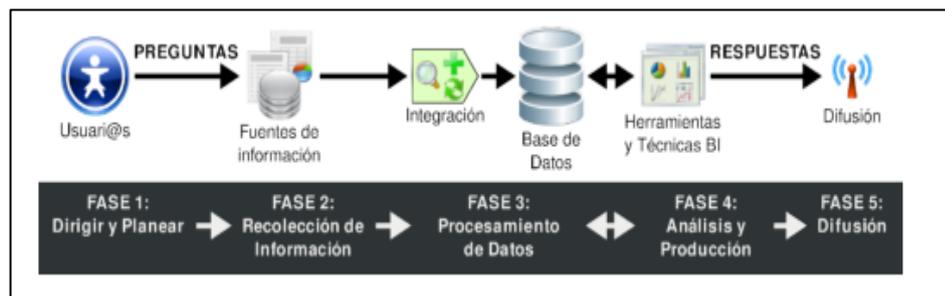


Figura 1. Fases del Proceso de B.I.

Fuente: (DATA PRIX, 2009)

2.4. Componentes de *Business Intelligence*.

(Cano J. L., 2013) Menciona los siguientes componentes:

- **Fuentes de Información:** Generalmente son los datos generados por sistemas operacionales, los cuales se utilizan para alimentar de información al *datawarehouse*.
- **Proceso ETL:** De extracción, transformación y carga de los datos en el *datawarehouse*, pasan por procesos de filtrado, limpieza, transformación y redefinición.
- **El *datawarehouse*:** En él se almacenan los datos de una manera que optimice su flexibilidad, facilidad de acceso y administración, en donde los datos están estructurados para generar informes que ayuden a la toma de decisiones.
- **El motor OLAP:** El motor OLAP, *On-Line Analytical Processing*, es el que proporciona la capacidad de realizar

cálculos, análisis, pronósticos, consultas en grandes volúmenes de datos.

2.5. Minería de Datos.

Antes de estudiar minería de datos, es importante conocer lo que es dato, información y conocimiento.

2.5.1. Datos.

Se trata del elemento constitutivo del conocimiento (Madrid, 2008). Éstos comprenden hechos, representaciones o los mecanismos por los cuales es posible medir e identificar algún aspecto del mundo. Quintanilla y Gil (Quintanilla & Gil, 2016) indican que existen algunos tipos de datos, entre ellos los datos abiertos que se diferencian de los datos personales y de los grandes datos (*big data*).

Los datos personales son el conjunto de datos que pertenecen a cada individuo, y se utilizan para la identificación del mismo. Los datos personales incluyen factores físicos, psicológicos, mentales, económicos, culturales y sociales (*European Union Parliament, 1995*). A los datos personales no se les aplica el movimiento de apertura, dado que este tipo de datos tiene su propia regulación. El número de datos se acrecienta permanentemente, y más aún desde que se utilizan las tecnologías de información y las nuevas herramientas de redes sociales, dando lugar a enormes volúmenes de ellos, conocidos como grandes datos (*big data*), los cuales tienen el potencial de transformar la forma de llevar a cabo las diversas acciones, tomar mejores decisiones y obtener ventajas al recolectarlos, integrarlos y analizarlos.

El término “grandes datos” se refiere a un gran conjunto complejo y cambiante de datos que crea nuevas formas de valor al cambiar los mercados, las organizaciones y las relaciones entre ciudadanos y gobiernos. *Mayer-Schönberger y Cukier (2013)* señalan: “Aprovechar los grandes datos

implica la capacidad de analizar enormes cantidad des de datos acerca de un tema”.

2.5.2. Información.

La Información es una interpretación de los datos, basada en un cambio de las condiciones y en el paso del tiempo -permitámonos pensar también en la perspectiva del observador- (Madrid, 2008), además se gesta al asignar patrones, relaciones y significado a los datos. (Riquelme, Ruiz, & Gilbert, 2006) dicen: “Los datos en bruto raramente son beneficiosos directamente”. Su verdadero valor se basa en la habilidad para extraer información útil, la toma de decisiones o la exploración, y en la comprensión del fenómeno gobernante en la fuente de datos.

2.5.3. Conocimiento.

Se comprende como información organizada dentro de un marco conceptual como lo puede ser: una visión del mundo, un concepto, un principio, una teoría o cualquier otra base de la necesaria abstracción conceptual que permite comprender el entorno, mejorar la capacidad para resolver problemas y tomar decisiones. El conocimiento trata del ámbito de la comprensión, según la cual actúan los individuos. Si bien la información puede llegar a ser abundante y agobiante, el conocimiento es escaso (Madrid, 2008).

2.5.4. Evolución Historia de la Minería de Datos.

La idea de Minería de Datos no es nueva. Ya desde los años sesenta los estadísticos manejaban términos como *data fishing*, *data mining* o *data archaeology*, con la idea de encontrar correlaciones sin una hipótesis previa en bases de datos con ruido.

A principios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros, empezaron a

consolidar los términos de data mining y KDD. A finales de los años ochenta sólo existía un par de empresas dedicadas a esta tecnología; en 2002 existieron más de 100 empresas en el mundo, que ofrecieron alrededor de 300 soluciones. Las listas de discusión sobre este tema las forman investigadores de más de ochenta países. Esta tecnología ha sido un buen punto de encuentro entre personas pertenecientes al ámbito académico y al de los negocios. Es una tecnología compuesta por etapas, que integra varias áreas y que no se debe confundir con un gran software (Rodríguez & Anolandy, 2009).

2.5.5. Concepto de Minería de Datos.

Escarbar en la información almacenada para descubrir elementos de utilidad, desde grandes cantidades de datos almacenadas, con el objetivo de detectar patrones de comportamiento consistente, o relaciones entre los diferentes campos de una base de datos para aplicarlos a nuevos conjuntos de datos, es importante. Puede visualizarse también como un proceso analítico, diseñado para explorar grandes cantidades de datos, con el objetivo de encontrar relaciones entre las diferentes variables, para aplicarlas a nuevos conjuntos de datos, según (Martínez M. C., 2003).

En el libro Congreso Español de Informática, de (Romero, Ventura, & Hervás, 2005), se entiende a la minería de datos como el proceso de descubrir el conocimiento para encontrar información no trivial, previamente desconocida y potencialmente útil de grandes repositorios de datos, Klossgen.

La minería de datos también se define en el libro de *Data, mining. Practical machines*, como el proceso de extraer conocimiento nuevo, útil y comprensible de datos almacenados en diferentes formatos. En otras palabras, la tarea fundamental de la minería de datos es la de encontrar modelos inteligibles, empezando desde los datos (Cardenas, Garrido, Muñoz, & Serrano, 2007).

2.5.6. Proceso de descubrimiento de conocimiento en minería de datos.

Velarde (2003) señala los siguientes pasos que se encuentran involucrados en el proceso de descubrimiento de conocimiento:

1. Entendimiento del dominio de aplicación, el conocimiento relevante a usar y las metas del usuario.
2. Seleccionar un conjunto de datos y enfocar la búsqueda en subconjuntos de variables y / o muestras de datos en donde realizar el proceso de descubrimiento.
3. Limpieza y preprocesamiento de datos, diseñando una estrategia adecuada para manejar ruido, valores incompletos, secuencias de tiempo y otros.
4. Reducción de datos y proyecciones para disminuir el número de variables a considerar.
5. Selección de la tarea de descubrimiento a realizar, por ejemplo: clasificación, agrupamiento, regresión, etcétera.
6. Selección del o los algoritmos a utilizar.
7. Llevar a cabo el proceso de minería de datos.
8. Interpretar los resultados y posiblemente regresar a los pasos anteriores. Esto puede involucrar repetir el proceso, quizás con otros datos, otros algoritmos, otras metas y otras estrategias.
9. Incorporar el conocimiento descubierto al sistema, normalmente para mejorarlo.

2.5.7. Fundamentos de Minería de Datos.

(Fuertes, 2013) Indica que las técnicas de Minería de Datos son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real.

La Minería de Datos toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva. La Minería de Datos está lista para su aplicación en la comunidad de negocios porque está soportado por tres tecnologías que ya están suficientemente maduras.

- Recolección masiva de datos.
- Potentes computadoras con multiprocesadores.
- Algoritmos de *Data Mining*.

2.5.8. Objetivos de la Minería de Datos.

Según (López, 2013), las características principales de la minería de datos son:

- Explorar los datos que se encuentran en las profundidades de las bases de datos, como los almacenes de datos, que algunas veces contienen información almacenada durante varios años.
- En algunos casos, los datos se consolidan en un almacén de datos y en mercados de datos; en otros se mantienen en servidores de Internet e Intranet.
- Las herramientas de la minería de datos ayudan a extraer el mineral de la información enterrado en archivos corporativos, o en registros públicos archivados.
- Muchas veces el minero es un usuario final con poca o ninguna habilidad de programación, facultado por barrenadoras de datos y otras poderosas herramientas indagatorias, para efectuar preguntas ad hoc y obtener rápidamente respuestas.
- Hurgar y sacudir a menudo implica el descubrimiento de resultados valiosos e inesperados.
- Las herramientas de la minería de datos se combinan fácilmente y pueden analizarse y procesarse rápidamente.

- Debido a la gran cantidad de datos, algunas veces resulta necesario usar procesamiento en paralelo para la minería de datos.
- La minería de datos produce cinco tipos de información: Asociaciones, Secuencias, Clasificaciones, Agrupamientos, Pronósticos.
- Los mineros de datos usan varias herramientas y técnicas.

2.5.9. Fases de un Proyecto de Minería de datos.

(Fuertes, 2013) y (Vallejos, 2006) mencionan que los pasos a seguir para la realización de un proyecto de minería de datos son siempre los mismos, independientemente de la técnica específica de extracción de conocimiento usada. El proceso de minería de datos pasa por las siguientes fases: filtrado de datos, selección de variables, extracción de conocimiento, interpretación y evaluación.

2.5.9.1. Filtrado de datos.

El formato de los datos contenidos en la fuente de datos y base de datos, nunca es el idóneo, y la mayoría de las veces no es posible ni siquiera utilizar ningún algoritmo de minería sobre los datos en bruto.

Mediante el preprocesado se filtran los datos, de forma que se eliminan valores incorrectos, no válidos, desconocidos, según las necesidades y el algoritmo a usar, se obtienen muestras de los mismos que están en busca de una mayor velocidad de respuesta del proceso, o se reduce el número de valores posibles

2.5.9.2. Selección de variables.

Aún después de haber sido preprocesados, en la mayoría de los casos se tiene una cantidad infinito de datos. La selección de las características

reduce el tamaño de los datos, eligiendo las variables más influyentes en el problema, sin apenas sacrificar la calidad del modelo de conocimiento, obtenido del proceso de minería.

Los métodos para la selección de características son básicamente dos:

- Aquellos basados en la elección de los mejores atributos del problema.
- Y aquellos que buscan variables independientes mediante test de sensibilidad, algoritmos de distancia o heurísticos.

2.6. Algoritmos de Extracción de Conocimiento.

Mediante una técnica de minería de datos se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema, o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez, para generar distintos modelos, aunque generalmente cada técnica obliga a un preprocesado diferente de los datos.

2.7. Interpretación y evaluación.

Una vez obtenido el modelo, se debe proceder a su validación, comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

2.8. Modelos y Técnicas de explotación de datos.

Es importante notar que no existe un “mejor” modelo o algoritmo de minería de datos, depende del problema en estudio y de los datos

disponibles para decir cuál de todos entrega resultados confiables (Martínez C. , 2012)

2.8.1. Modelos de explotación de datos.

Los modelos de *Data Mining* se clasifican en: supervisados o predictivos, y no supervisados o del descubrimiento del conocimiento, llamados también descriptivos.

2.8.1.1. Supervisados o predictivos:

Predicen el valor de un atributo de un conjunto de datos, permitiendo conocer otros atributos. Los modelos predictivos requieren ser “entrenados”, utilizando un conjunto de datos de entrenamiento, cuyo valor de variable objetivo es conocido. La idea es que el modelo entregue resultados en base a un aprendizaje, en otras palabras, que se vaya ajustando a la realidad conocida (Rodríguez Suárez & Díaz Amador, Herramientas de Minería de Datos, 2009).

2.8.1.2. No Sup. Descrip. o del descubrimiento del conocimiento.

Proporcionan información sobre las relaciones entre los datos y sus características (Beltrán, 2014). El modelo se va ajustando de acuerdo a las observaciones o datos entregados, recurriendo muchas veces a argumentos heurísticos para evaluar la calidad de los resultados, por lo que es usado en los casos en que no se tiene conocimiento previo del resultado al que se va a llegar (Martínez M. C., 2003)

2.9. Técnicas de explotación de datos.

Entre las técnicas estadísticas sencillas se incluyen el estudio de distribución de las variables, estudio de correlaciones entre matrices, tablas de contingencias, entre otros. Por otra parte, entre las técnicas más complejas se incluyen el Análisis de Factores, el Análisis de Grupos, el Escalado Multidimensional, etc. (Rodríguez & Anolandy, 2009).

2.9.1. Redes neuronales (*Neural Networks*).

Las redes neuronales son técnicas analíticas que permiten modelar el proceso de aprendizaje de una forma similar al funcionamiento del cerebro humano, básicamente, la capacidad de aprender a partir de nuevas experiencias. Estas técnicas tuvieron un desarrollo impresionante en la última década, con aplicaciones tanto a la medida como generales, comúnmente llamados Shell, y tienen como objetivo fundamental sustituir la función de un experto humano.

Una de las principales características de las redes neuronales es que son capaces de trabajar con datos incompletos e incluso paradójicos, que dependiendo del problema puede resultar una ventaja o un inconveniente. Además, estas técnicas poseen dos formas de aprendizaje: supervisado y no supervisado.

2.10. Análisis Preliminar de datos, usando *Query Tools*.

Es el primer paso de un proyecto de Minería de Datos, se aplica una consulta SQL al conjunto de datos, para rescatar algunos aspectos visibles antes de aplicar las técnicas.

2.10.1. Técnicas de Visualización.

Son aptas para ubicar patrones en un conjunto de datos, pueden usarse al comienzo de un proceso de Minería de Datos para determinar la calidad de los datos.

2.10.2. Reglas de Asociación.

Instrumento descriptivo, que tiene el objetivo de encontrar relaciones significativas entre los datos, utilizando probabilidades de ocurrencia de dos objetos. Por ejemplo, es el análisis de los artículos o productos de una canasta de compras en una tienda (Martínez C. , 2012)

2.10.3. Algoritmos Genéticos.

Son técnicas de optimización que usan procesos, tales como combinaciones genéticas y mutaciones, proporcionan programas y optimizaciones que pueden ser usados en la construcción y entrenamiento de otras estructuras, como es el caso de las redes neuronales. Además, los algoritmos genéticos son inspirados en el principio de la supervivencia de los más aptos (Rodríguez Suárez & Díaz Amador, Herramientas de Minería de Datos, 2009)

2.10.4. Redes Bayesianas.

Buscan determinar relaciones causales que expliquen un fenómeno según los datos contenidos en una base de datos. Se han usado principalmente para realizar predicciones (López, 2013). También, las redes bayesianas suelen ser el primer algoritmo para explorar datos. Lo que hace es buscar correlaciones entre atributos.

2.10.5. Árbol de Decisión.

Son estructuras que representan conjuntos de decisiones, y estas decisiones generan reglas para la clasificación de un conjunto de datos. Los árboles de decisión son fáciles de usar, admiten atributos discretos y continuos, tratan bien los atributos no significativos y los valores faltantes. Su principal ventaja es la facilidad de interpretación.

2.10.6. Clustering (Agrupamiento).

Agrupan datos dentro de un número de clases preestablecidas o no, partiendo de criterios de distancia o similitud, de manera que las clases sean similares entre sí y distintas con las otras clases. Su utilización ha proporcionado significativos resultados en lo que respecta a los clasificadores o reconocedores de patrones, como en el modelado de sistemas. Este método, debido a su naturaleza flexible, se puede combinar fácilmente con otro tipo de técnica de minería de datos, dando como resultado un sistema híbrido.

2.10.7. Segmentación.

Consiste en la división de la totalidad de los datos, según determinados criterios. Ejemplo: dividir los clientes en función de su antigüedad.

2.10.8. Clasificación.

Consiste en definir una serie de clases donde agrupar a los diferentes clientes. Ejemplo: definidas las variables de entrada, se produce una determinada salida que clasifica al cliente en un grupo o en otro. Por ejemplo, si la edad está entre 20 y 40, está casado y tiene cuenta de ahorro, entonces contrata hipoteca en un 78% de posibilidades.

2.10.9. Predicción.

Consiste en conocer los resultados futuros a partir de la modernización de los datos actuales. Ejemplo: crear un modelo de variables para saber si el cliente compra o no compra. Aplicar el modelo a un futuro cliente, y ya se puede predecir si comprará o no.

2.11. Herramientas de Minería de Datos.

Las herramientas de minería de datos facilitan el desarrollo de los modelos para la extracción de conocimiento de un dominio establecido, dichas herramientas contienen los algoritmos específicos para la aplicación de técnicas de minería de datos, se los puede utilizar mediante una interfaz gráfica. Algunas herramientas pueden comerciales como de libres distribución, para el desarrollo de modelos de minería de datos (Pérez, 2007), citado por (Córdova, 2014).

Según (Rodríguez Suárez & Díaz Amador, Herramientas de Minería de Datos, 2009) las herramientas de minería de datos empleadas en el proceso de extracción de conocimiento se pueden clasificar en dos grandes grupos:

2.11.1. Técnicas de verificación.

En las que el sistema se limita a comprobar hipótesis suministrada por el usuario.

2.11.2. Método de descubrimiento.

Por medio de este se han de encontrar patrones potencialmente interesantes de forma automática, incluyendo en este grupo todas las técnicas de predicción.

2.12. Herramientas Top de Minería de Datos.

Existen algunas herramientas diseñadas para extraer conocimientos desde bases de datos que contienen grandes cantidades de información, herramientas esenciales en los negocios modernos que son capaces de convertir los datos en inteligencia de negocios, dando así una ventaja de información (JMACOE, 2017).

En la Tabla 1, se muestra las herramientas de minería de datos más populares como *SPSS Clementine*, *Oracle Data Miner*, *Weka*, y *Orange*.

Tabla 1.
Herramientas de minerías de datos

Herramienta de Minería de Datos.	Tipo de Software.	Plataforma	Algoritmos	Tipo de modelo
Weka	Libre	Todas las plataformas	Clustering, Regresión, Asociación (A priori), Agrupamiento Numérico (Algoritmo K-Medias), Clasificación (J48, OneR)	Predictivo
Clementine	Libre	Windows, Linux	Red neuronal, GRI A priori, logística, QUEST, CHAID, KARMA	Predictivo
Knime	Libre	Windows, Linux, Mac Os	Algoritmos segmentación, árboles de decisión, redes neuronales, SVM	Predictivo
IBM SPSS	Comercial	Windows, Linux	Ecuaciones estructurales	Predictivo
RapidMiner	Libre	Windows, Linux	Clustering, árboles de decisión, redes neuronales	Predictivo
Orange	Libre	Multiplataforma	Modelo predictivo: árboles de clasificación, regresión logística, clasificador de Bayes, reglas de asociación. Métodos de descripción de datos: mapas autoorganizados, k-means clustering. Técnicas de validación del modelo: validación cruzada	Predictivo

Fuente: Rodríguez y Díaz (2009)

2.12.1. Weka (Waikato environment for knowledge analysis).

Es una herramienta visual de libre distribución desarrollada por los investigadores de la Universidad de Waikato, en Nueva Zelanda, en 1993, no obstante, su desarrollo original fue hecho en TCL/TK y C. En 1997 pasa a reescribirse todo el código fuente del entorno en Java, una plataforma más universal, y añadir las implementaciones de diferentes algoritmos de modelado (Córdova, 2014)

2.12.2. Características de *Weka*.

(Córdova, 2014) Describe las principales características de Weka:

- Acceso de los datos desde un archivo en formato ARFF, que es un archivo de texto plano; preprocesador de datos. selección, transformación de atributos; visualización del entorno y aplicación de técnicas de aprendizaje (Holmes & Witten, 1994).
- Esta es una herramienta muy versátil que soporta muchas tareas estándar de la minería de datos, en especial tareas de procesamiento de datos, regresión, clasificación, clusterin entre otras, así mismo permite la visualización y la selección de los datos. (Córdoba Fallas, 2011).
- Todas las técnicas en WEKA están basadas en la función de datos, que está disponible en un fichero plano o una relación, en donde cada registro de datos esta descrito por un número fijo de atributos nominales o numéricos (Córdoba Fallas, 2011).
- Permite el acceso a otras instancias de bases de datos por medio de SQL, gracias al JDBC, además puede procesar un resultado generado a base de una consulta hecha a una base de datos. (Córdoba Fallas, 2011).

2.12.3. Interfaz gráfica de Weka.

Weka está compuesta por una serie de herramientas gráficas de visualización y diferentes algoritmos para el análisis de datos y modelado predictivo. Su interfaz gráfica de usuario facilita el acceso a sus múltiples funcionalidades. Ésta potente herramienta de minería de datos se encuentra libremente disponible bajo la licencia pública general de GNU, además, al estar implementada en Java puede ejecutarse prácticamente bajo cualquier entorno. La interfaz gráfica de Weka cuenta con cuatro formas de acceso a las diferentes funcionalidades de la aplicación. En la Figura 2 se muestra la interfaz gráfica principal de WEKA.



Figura 2. Interfaz principal

Fuente: (García M. D., 2013)

2.12.4. Ventajas al utilizar Weka.

(Córdoba Fallas, 2011) Describe las siguientes ventajas:

Es un software de fácil acceso para los usuarios, ya que está disponible libremente bajo licencia pública GNU. Contiene una gran gama de técnicas para modelado y procesamiento de datos. Posee una interfaz de usuario muy sencilla que permite su uso hasta por usuarios sin experiencia. Funciona en

la mayoría de las plataformas actuales ya que es una implementación en java.

Simple CLI (Simple command-line interface). Permite acceder a las opciones de Weka mediante instrucciones por medio de una línea de texto simple (Beltrán, 2014). García (2013) señala que Simple CLI proporciona una consola para poder introducir mandatos. A pesar de que en apariencia sea muy simple, es extremadamente potente porque permite realizar cualquier operación soportada por Weka de forma directa; no obstante, es muy complicada de manejar, ya que es necesario un conocimiento completo de la aplicación. En la Figura 3, se puede observar la ventana de comandos de WEKA conocida como Simple CLI.

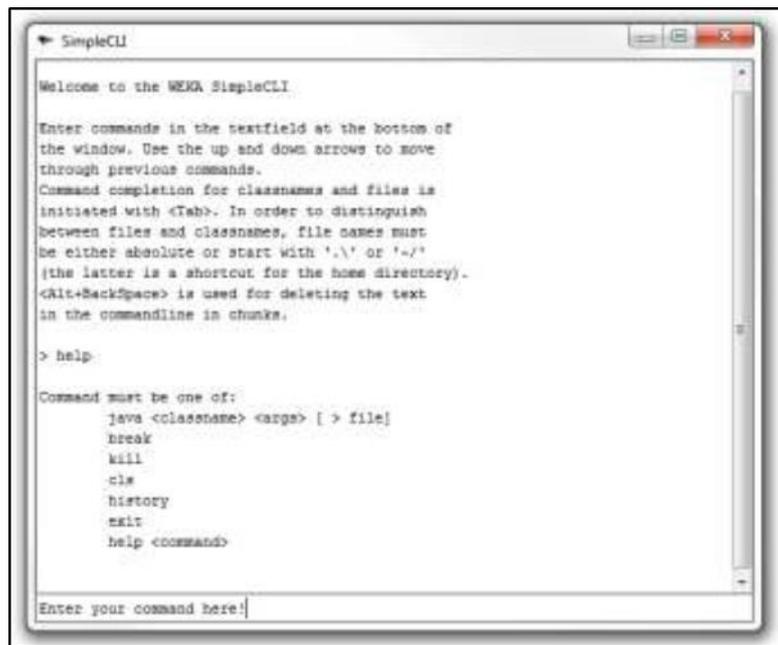


Figura 3. Ventana de Comandos

Fuente: (García M. D., 2013)

2.12.5. Explorer.

Es la interfaz de usuario que permite más accesibilidad a los componentes principales de todo el bando de trabajo, de un modo Figura. (Córdoba, 2011) Indica que ésta es la opción más intuitiva para el usuario,

pues dispone de varios paneles que dan acceso a las principales características del programa. (Calleja & Gómez, 2010). La Figura 4 muestra la ventana de Explorer de WEKA, con sus diferentes paneles o tareas como Preprocesado, Clasificación, Clustering, Búsqueda de Asociaciones, Selección de atributos y Visualización de datos.

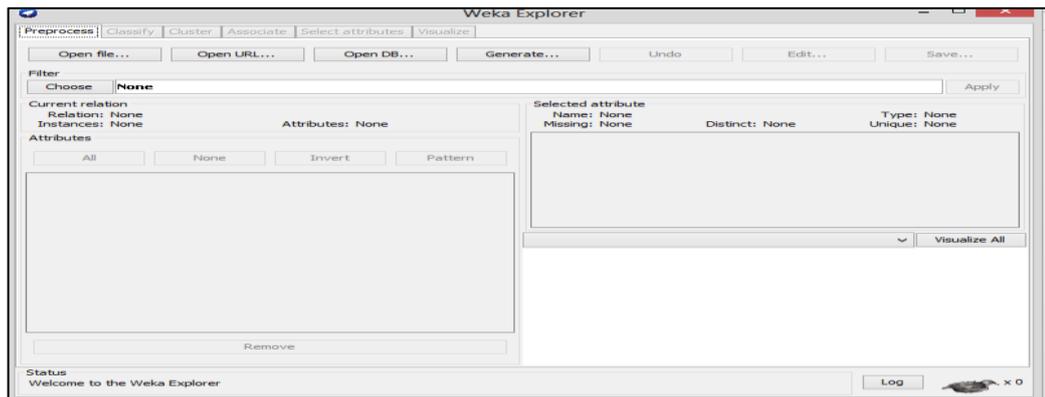


Figura 4. Explorer
Fuente: (García M. D., 2013)

2.12.6. Experimenter.

El modo Experimentador, Experimenter, es un modo muy útil para aplicar uno o varios métodos de clasificación sobre un gran conjunto de datos y, luego poder realizar contrastes estadísticos entre ellos y obtener otros índices estadísticos (García, 2013). En la Figura 5 se puede visualizar la ventana de la aplicación Experimenter.

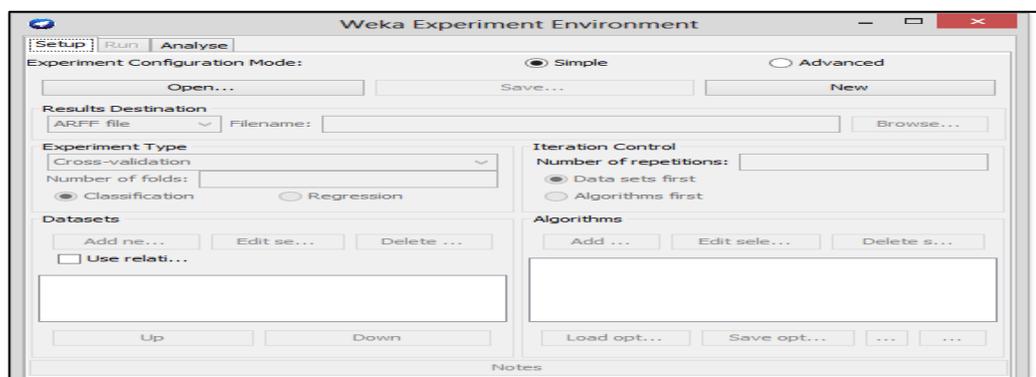


Figura 5. Experimenter
Fuente: (García M. D., 2013)

2.12.7. Knowledge Flow.

Soporta esencialmente las mismas opciones que la interfaz Explorer, pero esta permite “arrastrar y soltar” para la creación de árboles de decisión. Ofrece soporte para el aprendizaje incremental (Beltrán, 2014). La Figura 6 muestra lo siguiente:

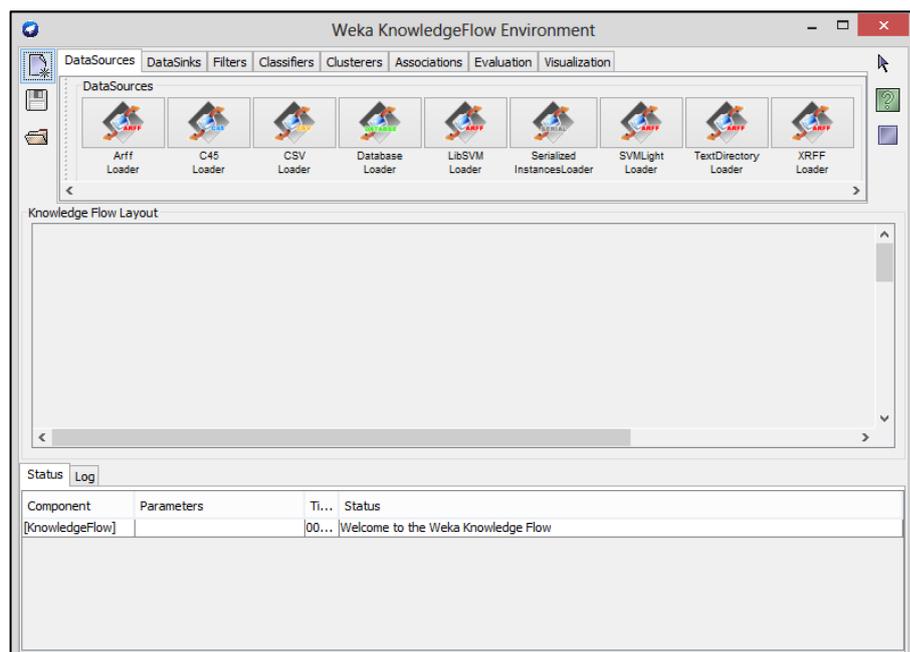


Figura 6. Knowledge flow

Fuente: (García M. D., 2013)

2.12.8. Clementine de SPSS.

(Rodríguez Suárez & Díaz Amador, 2009) Dicen que Clementine se centra en la integración de data mining con otros procesos y sistemas de negocio, para ayudar a entregar inteligencia predictiva en un tiempo eficiente durante las operaciones de negocio diarias. La funcionalidad abierta de data mining en bases de datos que posee Clementine permite que muchos de los procesos de data mining se realicen en entornos que mejoran, tanto el rendimiento como el despliegue de los resultados del mismo. La última versión de Clementine extiende la funcionalidad de data mining al incluir un conjunto de reglas de scoring y modelos de árboles de decisión y carga de resultados de data mining en la base de datos. Es un sistema integrado de

minería de datos que permite encontrar patrones en la información para facilitar la toma de decisiones a los usuarios. Utilizando Clementine se podrá:

- Acceder, preparar e integrar fácilmente datos numéricos de texto, datos provenientes de páginas Web y de encuestas.
- Construir y validar modelos rápidamente, utilizando las técnicas estadísticas y de aprendizaje automático disponibles más avanzadas.
- Implantar eficientemente los modelos predictivos, en tiempo real o según una programación establecida.
- Para personas que toman decisiones y hacen recomendaciones, y para los sistemas que los utilizan.
- Obtener rápidamente un mejor retorno de la inversión y mejores tiempos de respuesta aprovechando las características de rendimiento y escalabilidad.
- Transmitir de forma segura los datos confidenciales a las aplicaciones de data mining en los casos donde la seguridad es crítica.

Esta herramienta permite seleccionar campos o filtrar los datos, permite mostrar propiedades de los datos, encontrar relaciones, ambiente integrado de minería de datos para usuarios finales y desarrolladores. Constituida por algoritmos múltiples de minería de datos y herramientas de visualización. Su compañía es SPSS/*Integral Solutions Limited*, ISL. Funciona sobre todas las plataformas hardware y sistemas operativos, incluyendo Unix, VMS y Windows

2.12.9. Utilidades de *Clementi*.

Las organizaciones utilizan el conocimiento extraído con Clementine para:

- Retener a los clientes rentables.

- Identificar las oportunidades de venta cruzada.
- Detectar fraudes.
- Reducir riesgos y mejorar la prestación de servicios a la administración.
- Alcanzar un mayor nivel de conocimiento de sus clientes online, por lo tanto, mejorar el diseño de sus sitios web.

2.12.10. Knime.

Es un entorno totalmente gratuito para el desarrollo y ejecución de técnicas de minería de datos. KNIME fue desarrollado originalmente en el departamento de bioinformática y minería de datos de la Universidad de Constanza, Alemania, bajo la supervisión del profesor Michael Berthold. En la actualidad, la empresa KNIME.com GmbH, radicada en Zúrich, Suiza, continúa su desarrollo, además de prestar servicios de formación y consultoría. (Ismael)

(Rodríguez Suárez & Díaz Amador, 2009) Mencionan que Knime enteramente está escrito en Java, funciona en las plataformas o sistemas operativos más conocidos. Es un software de código abierto GNU y con licencia GPL. Recientemente fue lanzada la última versión, la cual incluye características como las de implicar nuevos formatos de entrada de datos con operadores para Microsoft Excel y SPSS. Desde la perspectiva de la visualización YALE, ofrece representaciones de datos en dispersión en 2D y 3D; representaciones de datos en formato SOM (Self Organizing Map); coordenadas paralelas y grandes posibilidades de transformar las visualizaciones de los datos.

Antes de empezar a utilizar KNIME se debe asegurar la instalación de los componentes de Weka (IED.Escuela de negocios, 2015).

2.12.11. Características de *KNIME*.

Izquierdo, Bravo, Ceruto y Martín (2015) dan a conocer las siguientes características:

- La herramienta *KNIME* guía el desarrollo del proceso de descubrimiento en los datos a través del diseño de un flujo de trabajo, en el cual se interconectan varios nodos entre sí para desarrollar todas las fases que lo componen.
- El resultado de ejecutar el flujo de trabajo cuando se aplica alguna técnica de MD es representado por patrones y modelos de MD, por ejemplo: reglas de asociación, árboles de decisión, grupos, entre otros (Han & Kamber 2006).
- El uso de los modelos/patrones de MD generados por *KNIME* resulta muy útil para aplicaciones que no se dediquen a guiar y desarrollar el proceso de descubrimiento de conocimiento en los datos.

2.12.12. *Knime vs Weka*.

Berzal (s.f.) compara ambas técnicas según los siguientes atributos:

- *Weka* incorpora un mayor número de componentes.
- La interfaz de *KNIME* es más amigable.
- *KNime* permite usar los nodos de *Weka*.
- *KNIME* también permite otras extensiones, como las ofrecidas por R.

2.12.13. *RapidMiner*.

Microsystem (2010) define a *RapidMiner* como una herramienta de Minería de Datos ampliamente usada y probada a nivel internacional en aplicaciones empresariales, de gobierno y academia. Implementa más de

500 técnicas de pre-procesamiento de datos, modelación predictiva y descriptiva, métodos de prueba de modelos, visualización de datos, etc.

RapidMiner está escrito en el lenguaje de programación Java, proporciona una interfaz gráfica para diseñar y ejecutar flujos de trabajo de análisis. Esos flujos de trabajo se llaman "Proceso" y éstos consisten en múltiples "Operadores". *Rapid Miner* se distribuye bajo la AGPL / Propietaria licencia de código abierto y ha sido organizada por (Rapid Miner, Software Business Intelligence, 2015).

2.12.14. Componentes de *RapidMiner*.

La herramienta cuenta con dos componentes:

- **RapidMiner Studio:** Versión stand-alone para analistas. Implementa todos los operadores de data mining, modelos predictivos, modelos descriptivos, transformación de datos, series de tiempo, etc. (Microsystem, 2010).
- **RapidMiner Server.** Versión Servidor de RapidMiner. Permite el trabajo colaborativo, escalable y concurrente de múltiples usuarios, capacidad de delegar en bases de datos (In-Database Mining) y otras mejoras de funcionalidad como: plataforma Web de publicación de reportes, implementación de sistemas de scoring, diseño y navegación Web de Reportes, Single-sign on, integración vía Servicios Web, entre otras. (Microsystem, 2010)

2.12.15. Características de *RapidMiner*.

En el trabajo realizado por (García González Francisco José, 2013) se describen las siguientes características:

- Está desarrollado en Java.
- Es multiplataforma.
- Representación interna de los procesos de análisis de datos en ficheros XML.

- Permite a los experimentos componerse de un gran número de operadores anidables arbitrariamente, que se detallan en archivos XML.
- Permite el desarrollo de programas a través de un lenguaje de script. Puede usarse de diversas maneras:
 - a. - A través de un GUI.
 - b. - En línea de comandos.
 - c. - En batch (lotes)
 - d. - Desde otros programas, a través de llamadas a sus bibliotecas.
- Extensible.
- Incluye Figuras y herramientas de visualización de datos.
- Dispone de un módulo de integración con R.
- Software de código abierto.

2.12.16. Orange.

(García González Francisco José, 2013) Indica que Orange es una suite para minería de datos y aprendizaje automático, desarrollado en la Facultad de Informática de la Universidad de Ljubljana, de Eslovenia. Es un software libre de aprendizaje automático.

2.12.17. Características de *Orange*.

En la revista Cultura CRM Magazine (2016) se describen algunas características que reside en las funcionalidades de Orange, como la programación visual front-end, para explorar datos y la visualización de resultados, es decir Orange apoya el entrenamiento práctico y las ilustraciones visuales de los conceptos de la ciencia de los datos. Incluso hay widgets que fueron especialmente diseñados para la enseñanza.

Orange se ha utilizado desde su inicio en estudios de biomedicina, bioinformática, investigación genómica e incluso en enseñanza. En estos

sectores, la herramienta ha funcionado como una plataforma de ensayo y error de nuevos algoritmos de aprendizaje automático. Mientras que en la educación, se ha extendido entre estudiantes de biología y biomedicina, con el fin de poner en marcha métodos de aprendizaje automático y análisis data mining (Cultura CRM Magazine , 2016).

2.12.18. Componentes de Orange.

(García González Francisco José, 2013) habla de los siguientes:

- Provee componentes para: Entrada/salida de datos, soportando los formatos C4.5, assistant, retis y tab (nativo).
- Pre procesamiento de datos: selección, discretización, etc.
- Modelado predictivo: árboles de clasificación, regresión logística, clasificador de Bayes, reglas de asociación, etc.
- Métodos de descripción de los datos: mapas autoorganizados, k-means clustering, etc.
- Técnicas de validación del modelo, como la validación cruzada.

De acuerdo al análisis de las diferentes herramientas mencionadas anteriormente, se puede concluir que todas conllevan a una misma finalidad, que es la de utilizar la información que se tiene a la mano para lograr abastecer las necesidades del mercado laboral. Al analizar las características de cada una de ellas, se puede definir que Orange presenta características de solución potente y completas que permiten la explotación de datos y la visualización de los resultados, recalcando además que es una herramienta de acceso libre y completamente gratuita.

2.12.19. Metodologías para el proceso de Minería de datos.

Cobos, Zúñiga, Guarin, León y Mendoza (2010) expresan que existen varias metodologías para orientar el proceso de minería de datos; ellas pretenden facilitar la realización de nuevos proyectos con características similares, optimizar la planificación y dirección de éstos, reducir su complejidad y permitir hacerles un mejor seguimiento a ellos (Gondar, 2004).

Existen varias metodologías para orientar el proceso de minería de datos; ellas pretenden facilitar la realización de nuevos proyectos con características similares, optimizar la planificación y dirección de éstos, reducir su complejidad y permitir hacerles un mejor seguimiento a ellos (Gondar, 2004). Entre esas metodologías se destacan CRISP-DM (2006) y SEMMA Sample, Explore, Modify, Model, Assess (SAS, 2009b).

En la Tabla 2 se muestran las metodologías más relevantes utilizadas para el proceso de minería de datos.

Tabla 2.

Metodologías para el proceso de Minería de datos

Metodología de Minería de Datos	TIPO DE PROCESO	DISTRIBUCIÓN	Fases que Propone
KDD	Iterativo e interactivo.	LIBRE	Selección, Preprocesamiento, Transformación, Minería De Datos y Evaluación e Implantación
CRISP-DM	Proceso Estandar,	LIBRE	Comprensión del negocio, Comprensión de los datos, Preparación de Datos, Modelamiento, Evaluación, Despliegue
SEMA	Proceso de selección, exploración y modelamiento de grandes cantidades de datos para descubrir patrones desconocidos...	LIBRE	Muestra, Exploración, Modificación, Modelo, Evaluación

2.13. KDD (*Knowlegde Discovery Databases*).

Según Fayyad et al. (2002), citados por (Quiroz Gil & Valencia, 2012), el KDD es: “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de los datos”. El objetivo es descubrir patrones y relaciones en los datos que puedan ser usados para hacer predicciones válidas (Gervilla, Jiménez, Montaña, Sesé, & Cajal, 2005).

El proceso KDD es iterativo e interactivo. Se dice iterativo ya que la salida de algunas fases puede retornar a pasos anteriores, y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad. Se habla de interactivo porque el usuario, o generalmente el experto del dominio del problema, debe ayudar en la preparación de los datos, validación del conocimiento extraído, entre otros. (Rodríguez & García, 2016).

2.13.1. Las Fases del KDD.

Según (Nigro, Xodo, Corti, & Terren, 2015), las fases de la metodología KDD son siete:

- Entender el dominio de aplicación, cuál es el problema a resolver, y cuáles son los objetivos.
- Seleccionar del conjunto de datos originales, un subconjunto apropiado, para el problema que deseamos resolver. Eliminando por ejemplo variables irrelevantes
- Implantación del almacén de datos que permita la “navegación” y visualización previa de sus datos, para discernir qué aspectos puede interesar que sean estudiados.
- Selección, limpieza y transformación de los datos que se van a analizar. En esta etapa de limpieza y pre procesamiento se deberían tomar decisiones con respecto a valores faltantes, atípicos, erróneos, ruido, etc.

- Seleccionar y aplicar el método y algoritmo de minería de datos apropiado. Elegir las herramientas de data mining adecuadas al problema a resolver, teniendo en cuenta el objetivo (predecir, explicar, clasificar, agrupar, etc).
- Evaluación, interpretación, transformación y representación de los patrones extraídos, llevada a cabo por el analista.
- Difusión y uso del nuevo conocimiento.

2.13.2. CRISP-DM (Cross-Industry Standard Process for Data Mining).

CRISP-DM [CRISP-DM, 2000], es la guía de referencia más ampliamente utilizada en el desarrollo de proyectos de Data Mining. (Arancibia, 2013).

De Luca (2006) hace referencia a que esta metodología contiene un conjunto de actividades seleccionadas en base a la experiencia de ensayo y error recogida, a través de numerosos proyectos por profesionales de *DaimlerChrysler,SPS YNCR*.

La metodología CRISP-DM constituye, por lo tanto, un mapa de ruta que permite determinar qué actividades desarrollar, en qué etapa y de qué manera alcanzar los objetivos finales del proyecto (De Luca, 2006).

2.13.3. Historia.

Es creada en el 2000 por el grupo de empresas SPSS, NCR y Daimler Chrysler. Es de distribución libre lo que le permite estar en constante desarrollo por la comunidad internacional. Además, resulta independiente de la herramienta que se utilice para llevar a cabo el proceso de MD. Es ampliamente usado por los miembros de la industria (Rodríguez & García, 2016)

2.13.4. Fases de CRISP-DM.

Rodríguez, Álvarez, mesa y González (2003) dan a conocer que CRISP-DM está dividida en cuatro niveles de abstracción, organizados de forma jerárquica en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining, en una serie de seis fases:

1. En la fase I, análisis del problema.

Se trata de comprender el negocio, donde se identifica la expectativa del cliente con el proceso de MD. Se determinan los objetivos y la producción del plan del proyecto. Para determinar los objetivos se debe distinguir los beneficios que se brindaran a la institución, organismo, empresa o cliente de forma general, que desea obtener información útil a partir de datos no supervisados. Es por ello que, esta parte del proceso es imprescindible el diálogo con los expertos del área del conocimiento en que se encuentra, se ofrece explicación de lo que se necesita para lograr lo que ellos esperan; así como las potencialidades de los resultados a obtener.

Por lo tanto, es provechoso definir en esta fase cómo es utilizado el conjunto de datos una vez realizado todo el proceso y categorizadas las instancias. Si con el solo hecho de haberlos dividido en grupos es ya suficiente como para brindar información relevante o esto, por sí solo, no es suficiente y es necesario utilizarlos para crear algún modelo computacional con técnicas estadísticas o de aprendizaje automático, por ejemplo. También este momento se puede tomar la decisión de que herramientas de MD son utilizadas en todo el proceso (Rodríguez & García, 2016).

2. Durante la fase II, comprensión de los datos.

Se obtiene una visión más realista del conjunto de casos del que se extrae el conocimiento. Chapman, et al. (2000), citados por Rodríguez y

García (2016), se orientan mediante la ejecución de tareas como: la recolección de los datos iniciales, describir y explorar los datos y verificar su calidad. Se analiza a continuación como realizar estas actividades en los problemas no supervisados tipo atributo-valor.

3. Es en la fase III, preparación de los datos.

En la que se seleccionan, limpian, construyen, integran y da la forma final a los datos. Los criterios para esta selección incluyen: la importancia para el cumplimiento de los objetivos de la MD, la calidad, y las restricciones técnicas como límites sobre el volumen de datos o los tipos de datos. Este proceso incluye no solo la selección de las variables sino también de los casos. Se puede aplicar filtros que modifiquen la muestra al tratar que la obtenida pueda ser mejor generalizada por los métodos de agrupamiento. Los análisis para trabajar con las instancias pueden incluir buscar en la muestra valores atípicos y extremos.

Estos valores pueden aparecer de cinco formas diferentes: Tipo A (Punto extremo): forma más simples. Tipo B (Extremo contextual): si una instancia de los datos es una rara ocurrencia respecto a un contexto específico de los datos y es normal respecto a otro. Tipo C (Extremo colectivo): es el caso en el que una instancia de los datos individualmente no es anómala, pero de conjunto con la totalidad de los datos sí es un extremo. Tipo D (Extremo real): estas son las observaciones ruidosas que son de interés en el sistema que se analice. Por tanto, no se debe tratar bajo el concepto de ruido sino el de extremos reales. Tipo E: (Extremo de error): es el caso en el que alguna observación es denominada incorrectamente como extremo, debido a algo inherente al problema en cuestión o fallo. (Rodríguez & García, 2016)

4. La fase IV, modelado.

Constituye la cúspide en el proceso de descubrir el conocimiento. Es aquí donde los algoritmos cavarán y se adentrarán en los casos preparados,

extraeyendo el preciado producto: la información. Se cumple esta función al realizar tareas como: seleccionar las técnicas de modelado, generar la prueba de diseño, construir y evaluar el modelo (Chapman, et al, 2000). La selección de los métodos de agrupamiento que se van a utilizar en el proceso de MD es de las partes más difíciles del proceso y de las más determinantes también (Rodríguez & García, 2016).

5. Es en la fase V, evaluación.

Es cuando se analizan los agrupamientos seleccionados como los más promisorios en la etapa anterior, no debe ser únicamente el que queda en primer lugar luego del ordenamiento según los índices de validación, sino una muestra de estos. De conjunto con los especialistas en el dominio, se selecciona el que mejor soluciona las necesidades existentes, y, por tanto, se comprueba el más útil de la modelación realizada. Para ello, pueden realizarse análisis de los centroides de los grupos, comportamientos de las variables dentro de un mismo conglomerado, distribución de las clases, entre otros. En este punto pueden descubrirse cuestiones importantes que pueden hacer que el proceso de MD regrese a la fase de modelado, o incluso anteriores, con un enfoque diferente (Rodríguez & García, 2016).

6. Por último, se tiene la fase VI.

En donde se despliega los resultados de la MD, se concluye que estrategia es el momento de documentar todo lo realizado y unir todos los reportes que el equipo de trabajo ha realizado independientemente. Todo esto es necesario para trazar una estrategia de monitoreo y mantenimiento del proceso, pues pasado un determinado tiempo pueden cometer errores con los resultados de la minería, pues los comportamientos de los datos pueden variar (Rodríguez & García, 2016)

2.13.5. Ventajas de CRISP-DM.

(Guzmán, 2005) Nombra las siguientes ventajas:

- No propietario,
- Independiente de la aplicación o la industria.
- Neutral con respecto a herramientas
- Enfocado en problemas de negocios, así como en el análisis técnico.
- Plataforma guía
- Plantillas para Análisis

2.13.6. SEMMA.

(Rodríguez, Álvarez, Mesa, & González, 2003) señalan que SAS Institute desarrollador de la metodología, SEMA, es el proceso de selección, exploración y modelado de grandes cantidades de datos para descubrir patrones de negocio desconocidos. Además (Rodríguez & García, 2016) dicen que la metodología SEMA pretende facilitar la exploración estadística, las técnicas de visualización, seleccionar y transformar las variables más significativas en la predicción, modelar las variables para predecir salidas y finalmente confirmar la precisión del modelo.

2.13.6.1. Fase de la Metodología SEMMA.

(Kammel, 2015) Detalla las siguientes fases de SEMMA:

1. Fase I *Sample* (Muestreo).

En esta primera fase de la metodología define el ámbito de la muestra de trabajo y garantiza su representatividad, este último punto es crucial ya que

los resultados que se extraigan de la muestra tienen que ser aplicables a toda la población objetivo del estudio.

2. Fase II Exploración.

Definido el ámbito de la muestra, se procede a explorarla para conocer en detalle de qué información se dispone y en qué estado se encuentra, y con ello adelantar los tratamientos que se deberían realizar en la fase de modificación. En esta exploración se realizan análisis descriptivos de las variables, así como procesos de detección de valores desinformados o atípicos (Kammel, 2015).

3. Fase III Modificación.

Tomando como base los resultados obtenidos en la fase de exploración, esta etapa se concentra en la selección y transformación de las variables que serán para la construcción de modelos. Otras tareas a realizar son: la imputación de valores desinformados e atípicos, la reducción de dimensión, la creación de nuevas categorías que aglutinen aquellas sin representatividad (Kammel, 2015)

4. Fase IV Modelado.

El objetivo de esta fase consiste en establecer una relación entre las variables explicativas y las variables objeto del estudio, mismas que posibiliten inferir el valor de las mismas con un 260 de nivel de confianza determinado. Las técnicas utilizadas para el modelado de los datos incluyen métodos estadísticos tradicionales, tales como análisis discriminante, métodos de agrupamiento, y análisis de regresión (Rodríguez, Álvarez, Mesa, & González, 2003). En esta fase las herramientas de software se encargan de realizar una búsqueda completa de combinaciones de datos

que juntos predecirán de una manera confiable los resultados buscados. (Pérez, 2016)

5. Fase V Evaluación de los datos obtenidos.

Después de que la fase de modelación presente los resultados obtenidos de la aplicación de los métodos de minería de datos al conjunto de datos, se deberá realizar un análisis de los resultados para ver si estos fueron exitosos de acuerdo a las entradas que se tuvieron para analizar el problema. (Yoshibauco, 2011).

Kammel (2015) manifiesta que una de las características de esta metodología es que la implementación de las fases en las que se estructura, ya sea de selección, exploración, modificación y modelado, no es rígida, es decir, no es necesario terminar una de sus fases antes de comenzar otra.

2.13.6.2. Particularidad de las Metodologías.

La particularidad de las metodologías radica en las características propias que posee cada una de las metodologías antes nombradas. En la Tabla 3 se muestran las características que permiten establecer la diferencia entre KDD, CRISP-DM Y SEMA.

Tabla 3.

Características que diferencian a las metodologías de minería de datos

KDD	CRISP-DM	SEMMA
Proveen sólo una guía general del trabajo a realizar en cada fase	Orientado a objetivos empresariales.	Orientado al desarrollo del proceso de minería de datos.
Constituyó el primer modelo aceptado en la comunidad científica que estableció las etapas principales de un proyecto de explotación de información.	Se inicia analizando los objetivos del negocio.	Se inicia analizando los datos.
Es iterativo e interactivo.	Metodología abierta y gratuita. Orientado a una metodología de gestión de proyectos.	Ligada a productos SAS.

Las tres metodologías analizadas son de propósito general para realizar procesos de minería de datos a toda la gama de problemas que pueden presentarse, en ninguna de sus fases presentan etapas o tareas como métodos, algoritmos o técnicas a usar en situaciones particulares que se presenten. Sin embargo, se considera a CRISP-DM más oportuna para realizar su adecuación a problemas no supervisados, es de libre distribución y no es dependiente de la herramienta que se utiliza; además es ampliamente la más usada dentro de diversos proyectos a nivel.

CAPÍTULO III

DESARROLLO DEL CASO DE ESTUDIO

3. Análisis de la Situación Actual de la Carrera

(Ordorika, 2015) dice que el seguimiento continuo de la integración social de los graduados universitarios, consiste en una fuente de indicadores que sirve como referencia para la planificación académica institucional. Concretamente, la sistematización de los estudios de seguimiento, como un eje de indagación de la pertinencia, se constituye en una categoría que rige la gestión de la calidad en las Instituciones de Educación Superior en las funciones de docencia, de investigación y de extensión. En este sentido, estudios guiados en esto pueden aportar información para los siguientes propósitos:

- a) Evaluar el nivel de calidad y percepción de la formación recibida, su inserción laboral, el desempeño profesional en los contextos ocupacionales y las necesidades de capacitación continua.
- b) Valorar el flujo de los cambios que suceden en el mercado laboral, en el campo ocupacional específico del profesional en relación a la carrera que se oferta.
- c) Evaluar las vías de integración que desarrollan las carreras a partir de las potencialidades del contexto laboral: local, regional, nacional e internacional.
- d) Evaluar el desempeño de los docentes en la planificación y la ejecución del proceso de enseñanza-aprendizaje, de investigación y de extensión.

(Simón & Arellano, 2009) menciona que los estudios de Seguimiento a Graduados, ESG, tuvieron su origen en los Estados Unidos y en algunos países de Europa. En México dichos estudios se iniciaron en la década de

los setenta, abarcando el área metropolitana de la ciudad de México, siendo pioneras la Universidad Nacional Autónoma de México, UNAM, la Universidad Autónoma Metropolitana–Azcapotzalco, UAM-A, el Instituto Politécnico Nacional, IPN y la Escuela Nacional de Estudios Profesionales–Zaragoza, ENEP-Z, mientras que en provincia las primeras universidades en llevar a cabo estos estudios fueron la Universidad Veracruzana, UV y la Universidad Autónoma de Nuevo León.

De igual manera (Mederos & Vega, 2003) opinan que la universidad debe acercarse al conocimiento de la realidad, utilizando, de forma más intensiva, el mundo del trabajo como recurso pedagógico, de manera tal que se logren acciones conjuntas con empresas y sindicatos. (Fuentes & Espinal, 2015) se refieren al seguimiento a graduados como una prioridad con las siguientes características generales:

- El punto focal es la correspondencia entre la formación recibida con las condiciones del medio laboral.
- Aunque se privilegien los aspectos propios de la formación, hay que considerar los más asociados al mundo del trabajo, tales como: distribución, adaptación, condiciones de utilización del medio laboral (Elementos internos y externos de la IES).
- Acudir a una amplia gama de métodos y técnicas de investigación para lograr aproximación a diferentes niveles: social, institucional e individual.
- Se debe complementar con estudios de campo a partir de experiencias de jóvenes graduados, sus jefes inmediatos, profesionales y dirigentes de experiencias.
- Considerar las diferentes ópticas del proceso y recalcar en la necesidad de realizar análisis tendenciales que permitan comparar diversas fuentes y grupos relacionados.
- Deben de realizarse regularmente, en sintonía con los ciclos de programas.

- Como criterios muestrales se propone: escenarios laborales más representativos, dinámica de los sectores que más receptan graduados, distribución, experiencia laboral de graduados entre uno a cinco años, características de su desarrollo laboral según la formación recibida.
- Proponer análisis comparativos entre las carreras, grupos de carreras, localidades, etc. Esto permite realizar análisis globales y metodológicos que pueden sustentar otras investigaciones.
- Protagonismo e identificación de los máximos responsables del proceso de formación y los docentes, con el fin de asegurar no solo la calidad, si no la divulgación y futura proyección.

(Tirado, Tejeda, & Cedeño, 2015), los estudios de seguimiento a graduados en Ecuador son un requisito, definido por CEAACES, para la acreditación y la calificación de las universidades y carreras universitarias. Al analizar los estudios de seguimiento a graduados en las universidades de Ecuador, pueden localizarse estudios puntuales que resultan de interés para el debate metodológico, ya que cada vez son más las universidades que tienen integrados estos estudios de manera sistemática.

Por otra parte, las universidades que están desarrollando un seguimiento sistemático de sus graduados suelen apoyarse en los recursos tecnológicos, implementando las encuestas de seguimiento en plataformas digitales que suelen integrar servicios informativos de interés para los estudiantes, bolsas de empleo, asociaciones estudiantiles, redes sociales, entre otros servicios.

Actualmente, en la Universidad de las Fuerzas Armadas ESPE se oferta la carrera de Ingeniero en Tecnologías de la Información, misma que tiene una malla descritas por niveles.

A continuación, se presenta la tabla 4 con las capacidades técnicas que un profesional de la carrera de Ingeniería en Sistemas de la Universidad de las Fuerzas Armadas está en conocimientos de resolver, y en otra columna

están las capacidades técnicas requeridas por las empresas al momento de publicar las ofertas laborales que disponen:

Tabla 4.

Capacidades técnicas de un profesional en ingeniería en sistemas, requeridas por las empresas, y ofertadas por la facultad de ingeniería en sistemas de la ESPE

Capacidades técnicas del profesional graduado como Ingeniero en tecnologías de la información	Capacidades técnicas requeridas por las empresas
Programación integrativa de componentes Web	Creación de aplicaciones NET Creación de aplicaciones JAVA Desarrollo de aplicaciones PHP Desarrollo de aplicaciones Node.js
Investigación Ingeniería de Software	Experiencia en desarrollo ágil, mantenimiento y calidad de software
Programación Avanzada	Conocimiento de metodologías de codificación y programación para mejorar el rendimiento de aplicaciones
Tecnología de Seguridad	Auditar regularmente la seguridad de la red
Vulnerabilidades en Sistemas Web	
Gestión de la Seguridad en TI	Seguridad Informática
Minería de Datos, sistemas de base de datos	Manejo de base de datos SQL Administración de base de datos my SQL Administración de base de datos NOSQL (cosmos DB)
Integración de Sistemas y arquitecturas	Mantenimiento de sistemas operativos
Interfaces y Multimedia	-
Redes de Comunicación	-
Ingles técnico	Ingles moderado

Como se pudo observar en la Tabla 4, las empresas solicitan a sus aspirantes capacidades técnicas, mismas que se ajustan a las ofertadas en el pensum de la ESPE, por lo que se podría decir que esta universidad forma estudiantes de acuerdo a los requerimientos del mercado.

Sin embargo, los problemas que se han encontrado al momento de levantar información sobre el número exacto de empresas residentes en Ecuador, que actualmente requieren o demanda profesional con el perfil de Ingeniero en Sistemas, son justamente las limitadas estadísticas y registros

por parte de instituciones públicas, sobre la cantidad de vacantes o puestos de trabajos que requieren las empresas.

Pues, al momento solo existe un portal web, o medio de información pública que registra y publica las ofertas de empleo que las empresas requieren, pero la mayoría de estas ofertas son del sector público.

Por el lado privado existen aproximadamente 9 bolsas de empleo online, estas son:

- Opción empleo
- Bolsa de Trabajo EC
- Por fin empleo
- Multitrabajos
- Computrabajo
- Acción trabajo
- Empleo con Quito
- Jobomas
- Olx.empleo

A pesar el acceder a estos portales en busca de ofertas de trabajo es totalmente gratuito para profesionales en el área de sistemas informáticos y otras áreas acceder, éstos se encuentran con la limitación de que algunas empresas se restringen a publicar sus nombres y los servicios a los que se dedican, por causas ajenas.

Otras fuentes en las que se buscó información de ofertas laborales para profesionales en el área de sistemas informáticos fueron los periódicos, no obstante, el problema que se presentó es que éstos no disponían con los recursos suficientes como para comprar todas las categorías de periódicos que existen en el país.

Un aspecto que se debe tomar en cuenta es que el número efectivo de vacantes que actualmente existan en el país será imposible de medir o calcular, dado que algunas empresas cuando requieren de un profesional, lo mencionan dentro de los mismos empleados de la empresa, sin necesidad de pagar en portales web, periódicos, o agencias de empleo. Y como resultado, esta oferta laboral no se llega a conocer más allá de los mismos

empleados, lo cual dificulta para levantar información del número real de vacantes que existen en el país.

Es importante mencionar que, para la selección de las vacantes de profesionales en sistemas, se procedió a seleccionar con mucho cuidado el número de estas, ya que una misma empresa puede haber pagado la publicación de su o sus vacantes, en más de una fuente (portales web, periódicos). Por mencionados argumentos que se describieron anteriormente, se procedió a estimar el número de ofertas que actualmente están disponibles para profesionales en el área de sistemas.

Antes de estimar el número de ofertas laborables disponibles en la carrera de Ingeniería en Sistemas, se procedió a estimar un número aproximado de la población total de empresas residentes en Ecuador, mismas que requerirán en algún momento como parte de su nómina a un profesional en la carrera de Ingeniería en Sistemas. Para ello, se ha buscado información en la Superintendencia de Compañías, sobre el número de empresas que actualmente se encuentran registradas y activas, pero que se dedican a las siguientes actividades:

- J6201.01 - actividades de diseño de la estructura y el contenido de los elementos siguientes (y/o escritura del código informático necesario para su creación y aplicación): programas de sistemas operativos (incluidas actualizaciones y parches de corrección), aplicaciones informáticas (incluidas actualizaciones y parches de corrección), bases de datos y páginas web.
- J6201.02 - adaptación de programas informáticos a las necesidades de los clientes, es decir, modificación y configuración de una aplicación existente para que pueda funcionar adecuadamente con los sistemas de información de que dispone el cliente.
- J6202.10 - actividades de planificación y diseño de sistemas informáticos que integran equipo y programas informáticos y tecnología de las comunicaciones.

- J6202.20 - servicios de gestión y manejo in situ de sistemas informáticos y/o instalaciones de procesamiento de datos de los clientes, y servicios de apoyo conexos.
- J6209.01 - actividades relacionadas a la informática como: recuperación en casos de desastre informático, instalación de programas informáticos.
- J6209.02 - actividades de instalación (montaje) de computadoras personales.

Se eligió empresas que se dedican a estas actividades económicas, dado que éstas serían las plazas potenciales que estarían obligadas a contratar un profesional con las características de un ingeniero en sistemas, por el mismo hecho de que en las actividades que realizan, necesitan capacidades y conocimientos técnicos de un profesional con el perfil de ingeniero en sistemas. A continuación, en la Tabla 5 se presenta el número de empresas de acuerdo a la actividad que se mencionó:

Tabla 5.
Empresas activas a nivel nacional que se dedican a actividades Económicas de servicio de la rama de Información y Comunicación

Cód. Actividad Económica	Número de empresas Activas	Porcentaje
J6201.01	534	42,7%
J6201.02	152	12,1%
J6202.10	404	32,3%
J6202.20	83	6,6%
J6209.01	66	5,3%
J6209.02	13	1,0%
Total	1252	100,0%

Fuente: Superintendencia de Compañías, Valores y Seguros (2017)

De acuerdo a la tabla anterior, se puede ver que las empresas que se dedican a los servicios de información y comunicación, a nivel nacional, son

las que mayor número presentan (42,7%) (Ver Tabla 4), cuyas actividades son;

- Actividades de diseño de la estructura y escritura de códigos informáticos.
- Programas de sistemas operativos., aplicaciones informáticas
- Bases de datos y páginas web.

Sin embargo, se procedió a ver la tendencia de mercado solo para las ciudades donde mayor número de empresas se concentran, siendo estas Quito con un 57%, y Guayaquil con un 30%, el 13% restante comprende las otras ciudades (Ver Figura 1).

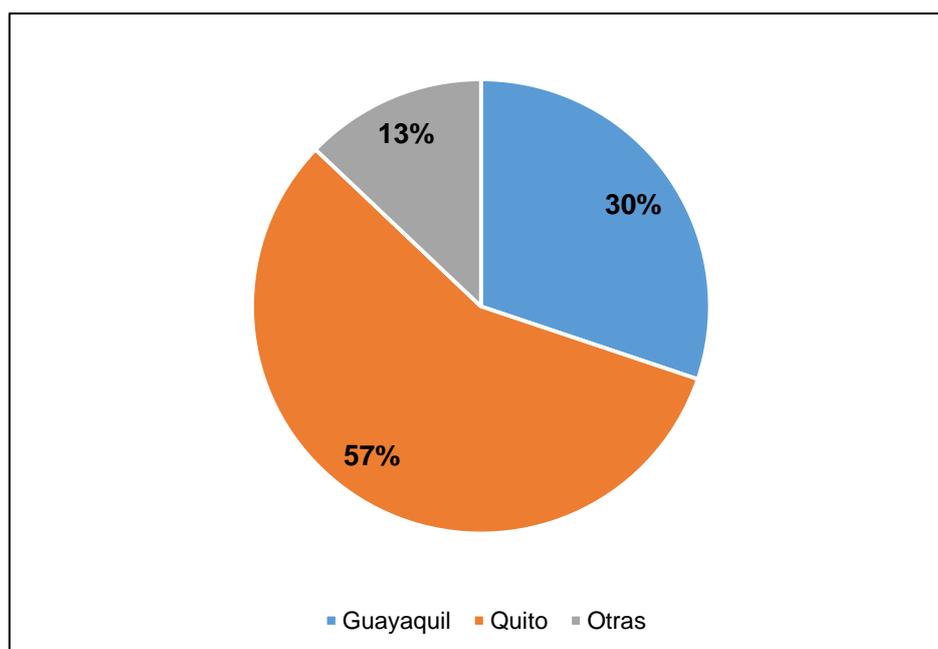


Figura 7. Proporción de empresas de actividades económicas; con códigos; J6201.01; J6201.02; J6202.10; J6202.20; J6209.01; J6209.02, por ciudades que mayor número de empresas concentran.

Fuente: Superintendencia de Compañías, Valores y Seguros (2017)

Una vez que se ha determinado una estimación sobre el número de empresas que demandan o demandaría los ingenieros en sistemas en Guayaquil y Quito, se ha supuesto que por lo menos 1 profesional será

requerido por cada empresa, dándonos como resultado de que en Guayaquil la demanda actual o futura es de 378 plazas de trabajo para un Ingeniero en Sistemas, y en Quito es de 713 plazas de trabajo (Ver Tabla 6).

Tabla 8.
Número de empresas en Quito y Guayaquil que demanda ingenieros en sistemas

Ciudades	Demanda	Porcentaje
Guayaquil	378	30%
Quito	713	57%
Otras	161	13%
Total	1252	100%

Fuente: Superintendencia de Compañías, Valores y Seguros

Sin embargo, para estimar los puestos de trabajo que realmente las empresas están demandando se procedió a calcular una muestra de 295 empresas, tomando como población las 1.252 empresas existentes en Guayaquil y Quito. Pero en base a un levantamiento de información en portales web (computrabajo, multitrabajo, Jobomas, opcionempleo,) se encontró que las empresas están realmente ofertando 102 plazas de trabajo en las dos ciudades (Ver Tabla 7).

Tabla 9.

Oferta laboral para ingenieros en sistemas, ofertado por empresas ubicadas en las ciudades de Quito y Guayaquil, según cuatro portales web de empleo

Ciudad	Multitrabajo		Computrabajo		Jobomas		Opcionempleo		Total	
	n	%	n	%	n	%	n	%	n	%
Quito	15	56%	24	71%	9	60%	15	65%	65	64%
Guayaquil	12	44%	10	29%	6	40%	8	35%	37	36%
Total	27	100%	34	100%	15	100%	23	100%	102	100%

En base a la Tabla 7, también se puede ver que el mayor número de ofertas laborales para ingenieros en sistemas, se encuentran en la ciudad

de Quito con un 64% (65 puestos de trabajo), mientras que Guayaquil solo alcanza un 36% (37 puestos). Por otro lado, los tipos de empresas que más ofertan son justamente como se había mencionado las empresas que se dedican a las actividades de servicios informáticos, es así que este grupo representa el 70,59 % del total de empresas ubicadas en Quito y Guayaquil (Ver Figura 2). Sin embargo, es importante mencionar que estos valores, solo son una aproximación del total del número de puestos de trabajo realmente ofertados, por cuestiones consideradas anteriormente, y además el número real de ofertas puede variar cada día, por el hecho de que en estos portales se publican anuncios de empleo diariamente.

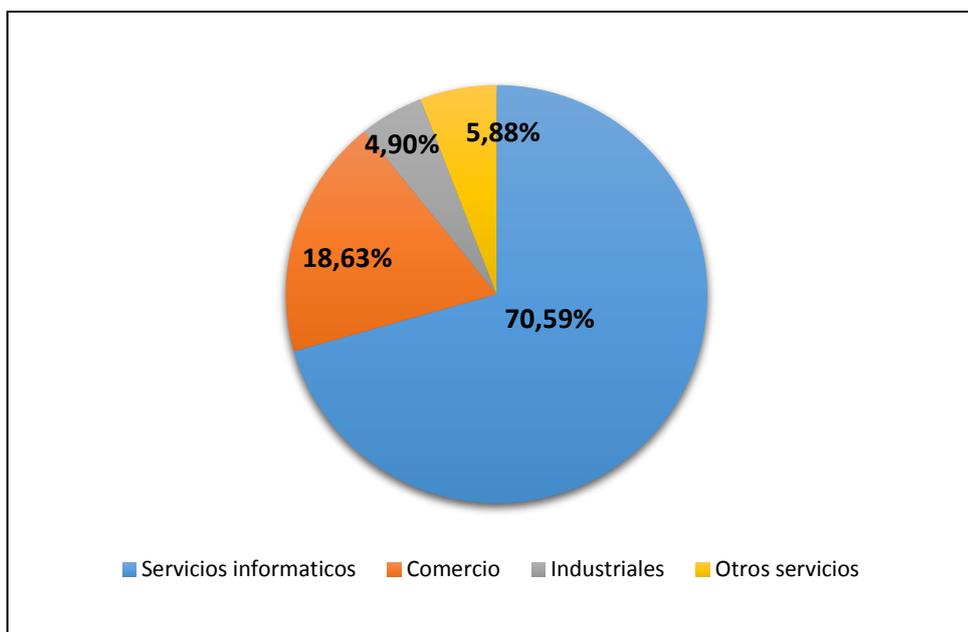


Figura 8. Oferta laboral para ingenieros en sistemas, ofertado por empresas ubicadas en las ciudades de Quito y Guayaquil, según el tipo de actividades que se dedican

Con lo expuesto hasta el momento, solo estamos en la posibilidad de poder hacer una estimación de la sobreoferta o sobredemanda que exista en el sector de servicios informáticos, dado que se dispone solo de una muestra del total de empresas en este sector, a pesar de que aproximadamente

existan un 18,63% de ofertas laborales para ingenieros en sistemas en Quito y Guayaquil, según información de los portales webs.

Con esto no se quiere decir que es difícil obtener información del número de empresas en otros sectores, es más la “SUPERCIAS” siempre registra esta información, pero lo que sí es difícil es saber si todas estas empresas requerirán por lo menos de los servicios de un ingeniero en sistemas, a diferencia del sector de servicios informáticos, que por el mismo hecho de que sus servicios se centren el área informática y tecnológica, necesitaran de un Ingeniero en esta área.

Por lo tanto, con la información que tenemos se estima que en el sector de empresas que se dedican a prestar servicios informáticos y tecnológicos ubicadas en las ciudades de Quito y Guayaquil, existe una demanda del 8,15% y una oferta del 24%, por ende, hay una sobreoferta del 15,85%, equivalente a decir que posiblemente puede haber 193 plazas de trabajo que podrían estar disponibles en los siguientes días por parte de estas empresas.

Para concluir, entre las capacidades técnicas de un profesional en ingeniería en sistemas solicitadas por estas empresas al momento de publicar las ofertas de empleo en los portales webs, se tienen las siguientes:

- Creación de aplicaciones NET
- Creación de aplicaciones JAVA
- Desarrollo de aplicaciones PHP
- Desarrollo de aplicaciones Node.js
- Desarrollo de aplicaciones JAVA
- Mango de base de datos SQL
- Administración de base de datos mySQL
- Administración de base de datos NOSQL (cosmos DB)
- Seguridad Informática.
- Mantenimiento de sistemas operativos.

- Experiencia en desarrollo ágil, mantenimiento y calidad de software.

Como se había mencionado en párrafos anteriores, y con respecto a la Universidad de las Fuerzas Armadas - ESPE, esta universidad se encuentra formando profesionales dentro de esta área, con conocimientos acordes a lo que requieren las empresas, sin embargo una limitación con lo que se encuentran los graduados de esta universidad, así como de otras es la nula experiencia profesional que las empresas solicitan, ya que los años de experiencia solicitadas por las empresas fluctúa de 2 a 3 años como mínimo. Características que se vuelve imposible tener para un recién graduado como ingeniero en sistemas.

3.1. Fundamentación teórica.

3.1.1. Pertinencia.

“La pertinencia constituye el medio por el cual se establecen las múltiples relaciones entre las universidades y el entorno” (Malagón, 2003).

3.1.2. Perfil de Egreso.

El perfil de egreso es una estructura descriptiva que representa la promesa y el compromiso institucional hacia la sociedad y los estudiantes, en términos de habilitar a éstos en los principales dominios de la profesión. Como contenido del contrato social entre la universidad y el estudiante y la sociedad, representa aquello que la universidad respaldará y certificará en el acto de graduación (Torres, Esquivel, López, & Cardona, 2011).

3.1.2. Mercado o Campo Laboral.

“El mercado laboral es el entorno en el que se interrelacionan las ofertas de trabajo y las demandas de empleo. Por tanto, constituye el ámbito en el que se define y desarrolla todo lo relacionado con las ocupaciones: empresas, personas que buscan empleo, contratos instituciones, etc.” (Santomé, 2004; p. 1).

3.1.3. Inserción laboral.

La inserción laboral es un proceso integral, en el que intervienen distintos factores para su realización, desde las características que presentan las personas y las oportunidades que ofrece el mercado laboral se organizan en un encuentro efectivo entre la empleabilidad y la ocupabilidad, Bisquerra (1992), citado por (Pelayo, 2012).

3.1.4. Perfil de egreso profesional.

Irigon (2006), citado por Sánchez, Molina, del Valle, PaScual y Alvear (2009), manifiesta que una carrera universitaria de pregrado necesita tener un “Perfil de egreso Académico Profesional”. Es decir, la declaración pública de aquellas competencias que un estudiante debe tener para que la universidad le otorgue un determinado grado y título al egresar de su carrera, y para las cuales se espera haberlo formado. Ante las necesidades que plantea el mundo se incluye el mejorar la empleabilidad de los egresados, y que estos sean profesionales competentes y destacados que puedan responder a las necesidades que tiene el mundo, las de ser transformado, para lograr superar la pobreza e inequidad, resolver pacíficamente los conflictos sociales e interculturales, mejorar las condiciones de vida de las personas y los pueblos y desarrollar el conocimiento.

A esto se agrega la necesidad que tienen los estudiantes universitarios y todos los seres humanos, la de tener una vida con sentido que se pueda sentirse realizado a plenitud y a la vez ser un aporte a la sociedad. Tobón (2006) dice que el perfil de cada carrera está formado tanto desde la academia como desde el mundo profesional. Y más aún, debe recoger los aportes de otros actores de la sociedad y de la comunidad universitaria, constituyéndose así en el Perfil del ciudadano que se quiere contribuir a formar.

3.1.5. Componentes del Perfil de Egreso Profesional.

(Urra & Jiménez, 2011) indican que el formato del perfil de Egreso está integrado por los siguientes componentes: Áreas de Conocimiento, Habilidades y Destrezas Profesionales, Actitudes y Valores, Competencias y Resultados de Aprendizaje. En la Figura 9 se puede observar cómo se encuentran distribuidos los componentes antes mencionados.

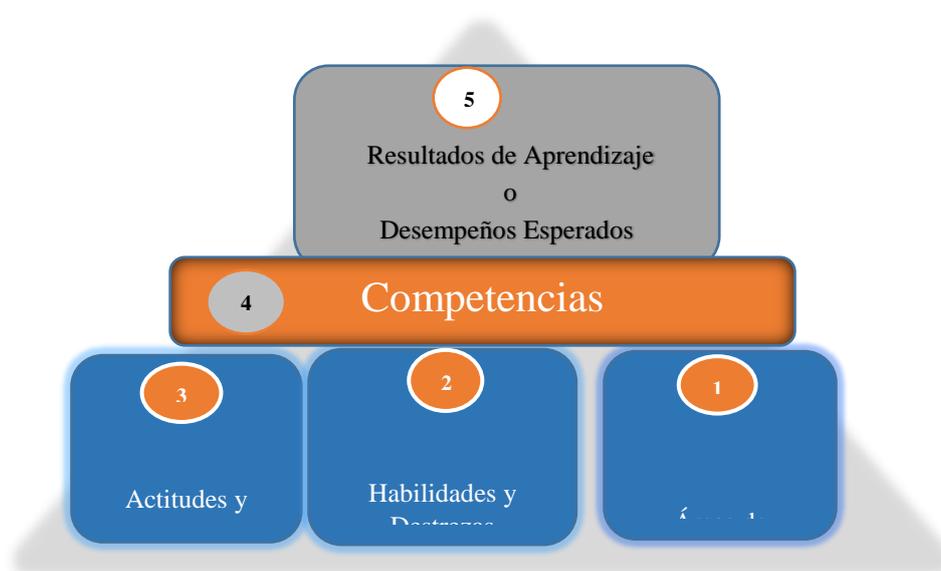


Figura 9. Componentes del perfil de egreso

Fuente: (Urra & Jiménez, 2011)

1. Áreas de Conocimiento.

Son las áreas esenciales de la o las disciplinas que otorgan el sustento científico y profesional a una carrera. (Urra & Jiménez, 2011).

2. Habilidades y Destrezas Profesionales.

Son formas de pensar u operar que condicionan altamente la actuación profesional, y no necesariamente deben asociarse a tareas muy determinadas (Urra & Jiménez, 2011).

3. Actitudes y Valores.

Son disposiciones adquiridas para actuar selectivamente o conducirse de determinada manera frente a una situación, y operan como parte de un

sistema de representación de la realidad, una vez incorporadas regulan la conducta (Urra & Jiménez, 2011).

4. Competencias.

Son un saber actuar complejo, producto de experiencias integradoras de aprendizaje, que se sustentan en la movilización y combinación eficaz de una variedad de recursos internos -conocimientos, habilidades y disposiciones- y externos, para el desempeño exitoso en una determinada situación o actividad real (Urra & Jiménez, 2011).

5. Resultados de Aprendizaje.

Cabe destacar que las competencias declaradas en el perfil de egreso contemplan las áreas de conocimiento, habilidades y destrezas, actitudes y valores declarados en los otros apartados del perfil. Es la forma en la que se generan las bases para el desarrollo de resultados de aprendizaje, y en el caso de las competencias, desempeños esperados que el plan de estudios debe asegurar (Urra & Jiménez, 2011)

Además, las carreras pueden optar por no incluir alguno de estos componentes, según el currículo del plan de estudios. Se señala sólo a modo de ejemplo que un perfil puede contener áreas de conocimiento, habilidades y destrezas profesionales y actitudes; mientras que otro perfil de egreso puede ser elaborado en base a áreas de conocimiento y competencias. La decisión deberá acordarla cada unidad académica, en consistencia con el tipo de plan de estudios que aspire a lograr (Urra & Jiménez, 2011).

3.1.6. Validar el Perfil de Egreso.

A continuación, se describen las siguientes fases:

1. Aplicar los criterios generales de diseño curricular.

Es importante revisar el cumplimiento de los hitos claves y el grado de consistencia que tiene el proceso de diseño curricular, con la formulación del perfil de egreso (Urra & Jiménez, 2011).

2. Verificar la consistencia del perfil de egreso con los objetivos de la carrera.

En este punto se debe considerar especialmente el contexto laboral o formativo de egreso, y los principios institucionales (Urra & Jiménez, 2011).

3. Uso del estado del arte en las definiciones del perfil de egreso.

(Urra & Jiménez, 2011) señalan que se debe garantizar que se ha empleado el conocimiento más pertinente y actualizado en su formulación. Para este propósito, la consulta a grupos de interés corresponde a una metodología altamente utilizada y recomendable. Puede hacerse a través de encuestas, entrevistas, *focus groups* u otros medios, e incluye la consideración de dos tipos de opiniones, para evaluar el perfil de egreso a posterior:

- Opiniones internas de alumnos, académicos y directivos.
- Opiniones externas de egresados, empleadores y expertos.

Esta es una etapa a mediano plazo, de alta relevancia, esencial para el Aseguramiento de la Calidad.

1. Verificar la factibilidad general de implementar un plan de estudios consistente con el perfil de egreso.

Si bien esto no es posible de hacerse con precisión hasta no haber elaborado en detalle el plan de estudios, la experiencia institucional debería permitir estimar a priori esa factibilidad. Factores esenciales en esa estimación son los recursos humanos y materiales requeridos, así como la posibilidad de evaluar los aprendizajes e impactos esperados, y de contar con el soporte administrativo necesario. La evaluación a posteriori del perfil de egreso es la verificación tanto del logro efectivo de los aprendizajes comprometidos, como del cumplimiento de los objetivos de la carrera. (Urra & Jiménez, 2011)

Es decir, se necesita verificar si los estudiantes al egresar lograron desarrollar las capacidades y atributos preestablecidos en el perfil de egreso, y si dichas capacidades y atributos, en el caso que se hubieran logrado, tuvieron el impacto esperado tanto en los egresados como en la sociedad. Del resultado de estas evaluaciones deben emerger antecedentes que permitan a cada carrera revisar sus perfiles de egreso, en caso de ser necesario, para mejorar su pertinencia social. (Urra & Jiménez, 2011)

2. Mercado o campo laboral.

Se denomina mercado de trabajo o mercado laboral al mercado en donde confluyen la demanda y la oferta de trabajo (Mercado Laboral GTH, 2011).

3. Trayectorias del Mercado.

Jiménez (2008) presenta su enfoque en relación a las siguientes trayectorias que se encuentran en el mercado.

4. Trayectoria Laboral.

Su propósito es conocer el impacto de la formación profesional en el desempeño de los egresados, y el recorrido laboral que desarrollan a lo largo de un periodo determinado (Jiménez, 2008)

5. Trayectoria ocupacional.

Dan cuenta de las ocupaciones que desempeñan los egresados en el mercado de trabajo, y se reconstruyen para explicar la movilidad ocupacional de los mismos. Cuando se alude al concepto trayectoria ocupacional se ha observado que la importancia del trabajo radica en la identificación de las actividades desarrolladas por los egresados, que pueden ser de una carrera en particular o de varias, y que, además, realizan actividades que pueden estar relacionadas o no con su formación profesional. (Jiménez, 2008)

6. Características del mercado laboral actual.

El mercado laboral ha sufrido cambios importantes en los últimos años (SOC Generalitat de Catalunya, 2013), por lo que se describe al mercado laboral con las siguientes características:

- Incremento de la presencia de la mujer en más sectores profesionales.
- Aumento de la producción, pero reducción de los puestos de trabajo.
- Aparición de nuevas profesiones.
- El sector servicios es el que ocupa más personas.
- Aumento de las pequeñas empresas en grandes núcleos urbanos, a consecuencia de una descentralización productiva.
- Fuerte competencia entre empresas.
- Aumento de la mecanización, uso generalizado de robots y automatismos industriales.
- Implicación de las personas trabajadoras en el funcionamiento de la empresa, cada vez se valora más la capacidad de organización y la iniciativa.
- Necesidad de aprender y adaptarse a cambios tecnológicos en muchas profesiones u oficios.
- Reducción de la jornada laboral.
- Aumento de oportunidades laborales dónde se trabaja con información, Figuras, datos, estudios.
- Disminución de la actividad del sector agrario.

3.1.7. Desempleo, subempleo y ocupación plena.

Serrano (2013) expresa que el desempleo surge debido a que existen personas que quieren trabajar, pero no encuentran trabajo. Este desequilibrio es ocasionado por varios factores, el principal, es el ciclo económico; existen períodos de contracción y otros expansivos en la economía; además pueden surgir choques de demanda u oferta, factores que en su conjunto ocasionan que la demanda laboral se contraiga y se genere desempleo. Por otra parte, se presenta un tipo de desempleo temporal debido a la normal desaparición de empleos; este período de búsqueda temporal de trabajos nuevos ocasiona que siempre exista una tasa positiva friccional de desempleo, también conocida como la tasa natural de desempleo (Aguar, 2007).

Así mismo, Serrano (2013) estipula que las situaciones de desempleo mencionadas anteriormente, se enmarcan en un mercado laboral sin imperfecciones, tales como sindicatos, rigideces legales o un gobierno intervencionista; en donde el salario es completamente flexible y se ajusta hasta el nivel de pleno empleo, en donde la oferta es igual a la demanda laboral. No obstante, cuando se relajan estos supuestos, el desempleo puede surgir por diversas razones; una de ellas es la rigidez de los salarios o el establecimiento de salarios mínimos.

Por otro lado, se encuentra el enfoque que considera un mercado laboral monopsónico, con cierto poder por parte de los empleadores, los cuales pagan a sus trabajadores por debajo de su producto marginal. Un ejemplo dentro de esto son ciertos trabajadores cualificados con habilidades especiales, los cuales prefieren trabajar en empresas en donde puedan aplicar sus conocimientos, aunque su salario este por debajo de su producto marginal. En este caso, el nivel de empleo puede incrementarse, debido al establecimiento de salarios mínimos ya que las empresas pasan a ser tomadoras de precios.

En América Latina el subempleo es asociado a la insuficiencia de recursos que deben ser utilizados por las unidades productivas; lo que decanta en actividades de baja productividad e ingreso que son características del sistema económico en la región y, por tanto, determinan una situación de "subempleo estructural" (Aguar, 2007).

Finalmente, la ocupación plena está constituida por personas ocupadas que se encuentran en edad de trabajar, que trabajan como mínimo la jornada legal de trabajo, y tienen ingresos superiores al salario unificado legal y no desean trabajar más, o bien que trabajan menos de 40 horas y sus ingresos son superiores al salario unificado legal y no desean trabajar más horas (INEC, 2014).

3.1.8. Mercado Laboral en Ecuador.

El Instituto Nacional de Estadística y Censos (INEC) (2016), cumpliendo con su calendario estadístico, publicó que según la última Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), Ecuador registró en septiembre 2016 una tasa de desempleo del 5,2% a nivel nacional. En lo que fue del año 2016, la tasa de desempleo se mantuvo estadísticamente estable. Desagregado por áreas, la tasa de desempleo urbano se ubicó en 6,7% y en el área rural en 2,3% en septiembre del 2016.

Por su parte, la tasa global de participación laboral se ubicó en 69,2%; cifra que refleja una mayor oferta laboral o que existen más personas ingresando al mercado de trabajo. Consistente con lo anterior, la tasa de empleo bruto –proporción de la población en edad de trabajar que tiene un empleo asalariado o independiente- que mide la capacidad de absorción del empleo en la economía, alcanza el 65,6% el noveno mes del año. El subempleo -personas ocupadas que reciben ingresos inferiores al salario básico, y/o trabajaron menos de la jornada legal, pero tienen el deseo y disponibilidad de trabajar más- se ubicó en 19,4%.

3.1.9. Análisis de la data utilizando la Metodología CRISP-DM.

En esta sección se explica la metodología que se utilizó para realizar el análisis de la pertinencia del perfil de egreso y su eficiencia en el campo laboral de los graduados de la carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, aplicando técnicas de minería de datos. En primera instancia se da a conocer el tipo de investigación, luego se observa la elección y aplicación de la metodología y las etapas que la comprende, que van desde el tratamiento y procesamiento de los datos, hasta la selección de los modelos y algoritmos de *Data Mining*.

3.1.10. Tipo de Investigación.

El tipo de investigación es cualitativa, porque permite recoger, procesar y analizar datos cualitativos. Se puede visualizar una serie de datos clasificados que se enmarcan en el propósito de este proyecto, que es el de demostrar la incidencia del perfil de egreso en el campo laboral de los graduados de la carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas- ESPE.

3.1.11. Métodos de investigación.

El método utilizado en el proyecto de investigación es el analítico descriptivo, pues éste permite que los grupos de datos sean analizados de la manera adecuada en el software (Orange) y poder cumplir con los parámetros de la metodología de minería de datos CRISP-DM, facilitando aplicar los algoritmos y mostrar los resultados esperados.

3.1.12. Pre-procesamiento de datos.

El pre procesamiento de datos es uno de los primeros pasos que se toman en cuenta al momento de decidir el algoritmo a emplear en minería de datos. Existen técnicas destinadas a detectar y manipular datos considerados imperfectos; y por otro lado, se consideran aquellas técnicas

cuya finalidad es transformar los datos para hacerlos más manejables (Cárdenas, 2015).

Según (Cárdenas, 2015) se aplican técnicas de agrupamiento de datos, integración de datos, limpieza de datos, selección de variables y atributos, reducción de la dimensionalidad, filtrado de datos.

En este proyecto el agrupamiento de datos permitió reunir todos los datos disponibles para la resolución del problema. En la integración de datos se agrupó todos los datos, inclusive aquellos que se obtuvieron y recopilaron de otras fuentes, como páginas de redes sociales, LinkedIn, Senescyt, Registro Civil, etc, tomando en cuenta que esta recopilación de datos sea coherente. La limpieza de datos permitió detectar algunos datos erróneos o irrelevantes, y se procedió a descartarlos, analizando la inconsistencia de estos en relación al conjunto de datos de la base de datos. En la selección de variables y atributos se descartó y eliminó atributos de los datos que no eran relevantes para la toma de decisiones, en este caso no permitían obtener los resultados que se buscaba.

Se construyó nuevos atributos a partir de los atributos originales, siempre tratando de no perder información esto en la técnica de la reducción de la dimensionalidad en el filtrado de datos se eliminó información redundante para obtener una buena base de datos con un volumen de datos manejable, facilitando una mejor interpretación de la información.

3.1.13. Técnicas de minería de datos aplicada al proyecto.

Las técnicas de minería utilizadas en el proyecto son K-means clusterización, árbol de decisión y un análisis descriptivo empleando reglas de asociación.

3.1.14. Población.

Méndez (1998) define a la población como el total de elementos sobre el cual se quiere hacer una inferencia, para lo que se debe basar en la información relativa a ciertas especificaciones.

Para este proyecto se consideró a los graduados de la Facultad de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas –ESPE, de los años 2002 hasta el 2016 comprendiendo un total de 743 graduados que son los que se encuentran registrados en la base de datos del sistema de seguimiento a graduados de la carrera con fecha de corte 2016. (Autora)

En la Tabla 8 se muestra la Población existente, la misma que se encuentra establecida en columnas por el año de graduación, la cantidad de graduados por año y el porcentaje que representa. Observar el Figura 10 para diferenciar de forma detallada el número de graduados por año.

Tabla 10.
Población

AÑO DE GRADUACIÓN	CANTIDAD	PORCENTAJE
2002	21	2.83
2003	10	1.35
2004	4	0.54
2005	7	0.94
2006	10	1.35
2007	5	0.67
2008	7	0.94
2009	9	1.21
2010	38	5.11
2011	73	9.83
2012	159	21.40
2013	46	6.19
2014	93	12.52
2015	160	21.53
2016	95	12.79
S/R	3	0.40
OTROS	3	0.40
TOTAL	743	100.00

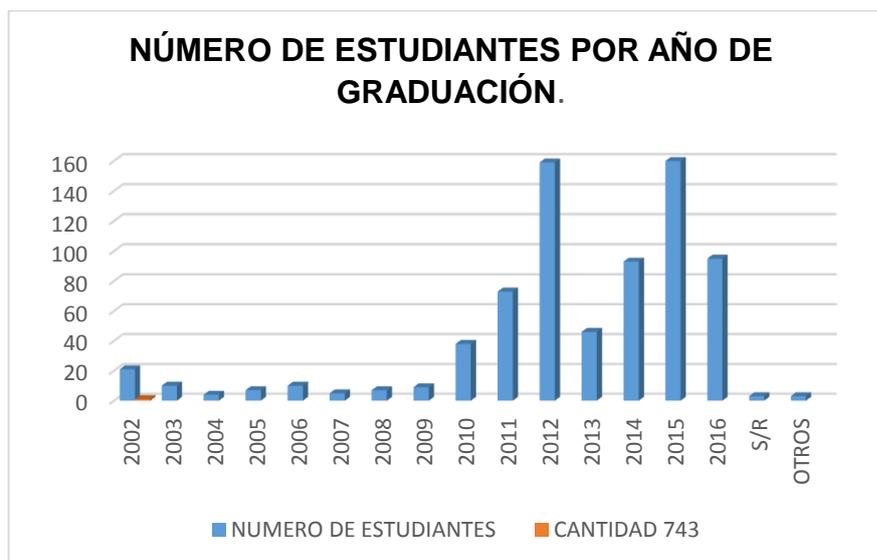


Figura 10. Población

3.1.15. Muestra.

Se aplicó un muestreo probabilístico aleatorio estratificado, debido a que la población está dividida en diferentes subgrupos o estratos de grupo (SUPO, 2014). Esto quiere decir que los grupos son heterogéneos entre sí, pero homogéneos dentro de cada grupo.

Heterogéneos porque existen los estratos años de graduación, que se diferencian por su tiempo de corte y homogéneo debido a que los campos a evaluar de cada año de graduación son los mismos. (Autora)

Saraí (2005) dice que en las investigaciones donde la variable principal es de tipo cualitativo, que se reporta mediante la proporción del fenómeno en estudio en la población de referencia, la muestra se calcula a través de las fórmulas para población finita.

Por lo tanto, la fórmula empleada para calcular la muestra de este proyecto es de tipo cualitativo para población finita.

$$n = \frac{NZ^2 pq}{d^2(N-1) + Z^2 pq}$$

3.1.15.1. Cálculo de la muestra.

$$n = \frac{NZ^2 pq}{d^2(N-1) + Z^2 pq}$$

$$n = \frac{743 \times (1.96)^2 \times (0.5 \times 0.5)}{[(0.05)^2 \times (743-1)] + [(1.96)^2 \times 0.5 \times 0.5]}$$

$$n = \frac{743 \times 3.8416 \times 0.25}{[0.0025 \times (742)] + [3.8416 \times 0.25]}$$

$$n = \frac{713,5772}{(1.855) + (0.604)}$$

$$n = \frac{713,5772}{2.8154}$$

$$n = 253.45315052$$

$$m = 254$$

m=254 graduados.

n= Tamaño de la muestra.

N= Tamaño de la población.

Z $\alpha/2$ = Calculado en las tablas del área de la curva normal llamado también nivel de confianza 1,96.

d = nivel de precisión absoluta. Referido a la amplitud del intervalo de confianza deseado en la determinación del valor promedio de la variable en estudio.

p = proporción aproximada del fenómeno en estudio en la población de referencia. 0,5

q = proporción de la población de referencia que no presenta el fenómeno en estudio (1-p).

Es importante recalcar que para este proyecto no se tomaron en cuenta los 254 graduados que es el resultado que se obtuvo al calcular la muestra, si no que se tomaron 359 graduados o datos que se adquirieron al validar, limpiar y filtrar la base de datos de los 743 que conforman la población, reiterando que en un proyecto de minería de datos es favorable trabajar con muestras grandes puesto que entre más datos y variables tengamos es mejor, esto permitirá obtener descripciones más coherentes y entender mejor el comportamiento del grupo de graduados de la Facultad de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas –ESPE.

3.1.16. Aplicación de la metodología CRISP-DM al caso de estudio.

CRISP-DM (Cross – Industry Standard Process for Data Mining) (Pete Chapman (NCR), 1999-2000), es una metodología que guía apropiadamente el desarrollo de un proyecto de minería de datos (Cobos, Zuñiga, Guarín, León, & Mendoza, 2010).

En este proyecto se aplicará la metodología CRISP-DM, tratando de obtener el mejor provecho al uso de Data Mining al evaluar las seis fases, logrando entender de manera completa lo que se desea resolver y permitiendo recolectar de manera correcta los datos, así como realizar un análisis adecuado e interpretar correctamente los resultados, alcanzando los objetivos planteados

En la Figura 11 se observar las *Fases del modelo CRISP-DM*.

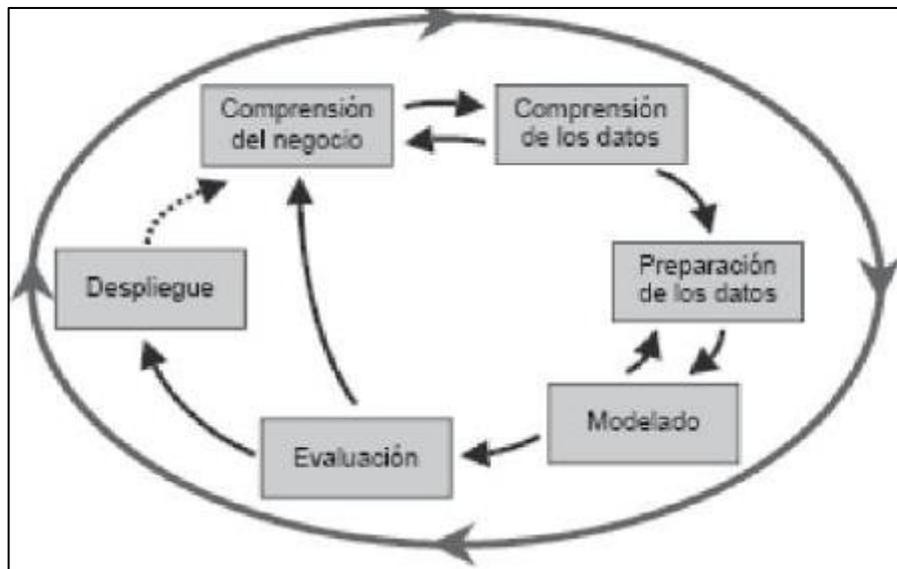


Figura 11. Fases del modelo CRISP-DM.

Fuente: Pete Chapman, INCR (1999-2000)

1. Fase I: Comprensión del Negocio o Problema.

El análisis a ejecutarse es de la pertinencia del perfil de egreso y su eficiencia en el campo laboral de los graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE-. En esta fase se analiza los siguientes parámetros:

- Establecimiento de los objetivos del negocio. El objetivo es determinar la pertinencia de los graduados en las empresas que laboran de acuerdo al Perfil de Egreso.
- Evaluación de la situación. La carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, actualmente cuenta con un Departamento de Seguimiento a Graduados, el mismo que tiene un repositorio de los datos de los graduados en el cual consta el campo laboral donde se desempeñan, nombre del lugar donde residen actualmente, tipo de empresa entre otros.

- Establecimiento de los objetivos de la minería de datos. Uno de los objetivos principales es saber si el campo laboral que se encuentran desempeñando los graduados tiene pertinencia con el perfil de egreso de la carrera que se formaron académicamente, así como el determinar el campo laboral de mayor influencia en los graduados de género femenino, con lo cual la minería de datos permitirá detectar posibles campos de trabajo, tomando en cuenta los campos laborales actuales en los que se encuentran inmersos los graduados.
- Generación del plan del proyecto. - Detalladas las fases, se propone analizar la base de datos de graduados utilizando la herramienta Orange Canvas y algoritmos que incluyen la misma.

2. Fase II: Comprensión de Datos.

- Recolección de datos iniciales. La Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE posee una base de datos de graduados, en la cual se alojan los siguientes datos de los graduados que se detallan a continuación:
- Cédula de Identidad. Documento que identifica a un individuo jurídicamente.
- Nombre y Apellido del Graduado. Nombre y apellido del ex estudiante que culminó sus estudios.
- País. Lugar de procedencia de una persona.
- Sexo. Diferenciación génica entre hombre y mujer.
- Dirección. Referencia del lugar en donde se puede localizar a un individuo.
- Provincia. Sector en donde que se encuentra en uno o varios cantones. Referencia del sector de ubicación de una persona

- Número de Teléfono. Secuencia de dígitos que se encuentran en una red telefónica.
- Mail Personal. Medio de comunicación e interacción de información.
- Cargo que desempeña en la empresa que labora. Realizar las funciones propias de un empleo o que un individuo se encuentra designado a realizar de forma específica en su empresa.
- Nombre de la Empresa en que trabaja. Nombre de la entidad a la que presta sus servicios.
- Ingreso mensual. Ingreso económico con el que cuenta mes a mes una persona.
- Tipo de empresa. Hace referencia a si labora en una empresa pública o privada
- Posición actual del graduado en el campo laboral. Si el graduado se encuentra trabajando o no.
- Años de trabajo. Años que labora en la empresa.

La información y resultados que buscamos de los Graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas - ESPE se muestran en la Tabla 9.

Info	edad	Cargo Empresa	iAR DE NACIMIEN	Sexo	ÑO GRADUACIÓI	CARGO	Ingreso Mensual	acional / Extranje	UBLICO / PRIVAD	empleadores trab.
359 instances 10 features (2.9% missing values) No target variable. 2 meta attributes (3.2% missing values)	1 24	Desarrollador de software	Pichincha	FEMENINO	2016.000	Desarrollo de s...	Menor a \$1.000	Nacional	PUBLICA	? ?
Variables	2 32	ADMINISTRADOR DE TECNOLOGIAS	Pichincha	MASCULINO	2010.000	Coordinación T...	Entre \$1.000 y \$...	Nacional	PRIVADA	5,000 6,000
<input checked="" type="checkbox"/> Show variable labels (if present)	3 29	Especialista de Servicios CSI Forence	Pichincha	MASCULINO	2014.000	SEGURIDAD DE ...	Entre \$1.000 y \$...	Nacional	PRIVADA	2,000 2,000
<input checked="" type="checkbox"/> Visualize continuous values	4 38	analista de tecnología geoinformatico	Pichincha	MASCULINO	2006.000	OTRO	Entre \$1.000 y \$...	Nacional	PUBLICA	6,000 11,000
<input checked="" type="checkbox"/> Color by instance classes	5 36	Thechnical Project Manager LATAM. G...	Cotopaxi	MASCULINO	2004.000	Administrativo	Entre \$2.001 y \$...	Nacional	PRIVADA	4,000 10,000
Selection	6 33	docente	Pichincha	MASCULINO	2011.000	DOCENTE	Entre \$1.000 y \$...	Nacional	PUBLICA	2,000 3,000
<input checked="" type="checkbox"/> Select full rows	7 27	Desarrollador	Loja	MASCULINO	2014.000	Desarrollo de s...	Menor a \$1.000	Extranjera	PRIVADA	1,000 1,000
Restore Original Order	8 41	Ingeniero de Soporte	Tungurahua	FEMENINO	2002.000	Soporte Técnico	Entre \$1.000 y \$...	Nacional	PRIVADA	7,000 16,000
Report	9 50	Gerente General de Operaciones	Pichincha	MASCULINO	2015.000	Administrativo	Entre \$3.001 y \$...	Nacional	PRIVADA	8,000 20,000
<input checked="" type="checkbox"/> Send Automatically	10 27	Consultor Técnico	Pichincha	MASCULINO	2015.000	Consultor_Aud...	Entre \$1.000 y \$...	Extranjera	PRIVADA	2,000 3,000
	11 41	COORDINADOR DE LAS TICs	Pichincha	MASCULINO	2015.000	DOCENTE	Entre \$1.000 y \$...	Nacional	PUBLICA	2,000 0,000
	12 28	Analista de Base de Datos	Pichincha	MASCULINO	2014.000	SISTEMAS DE L...	Menor a \$1.000	Nacional	PUBLICA	? 0,000
	13 30	Research Assistant Asistente Investiga...	Pichincha	MASCULINO	2010.000	DOCENTE	Entre \$2.001 y \$...	Extranjera	PRIVADA	2,000 3,000
	14 33	mantenimiento	Pichincha	MASCULINO	2016.000	Soporte Técnico	Menor a \$1.000	Extranjera	PRIVADA	6,000 4,000
	15 36	webmaster desarrollador	Pichincha	MASCULINO	2005.000	Desarrollo de s...	Entre \$1.000 y \$...	Nacional	PUBLICA	3,000 9,000
	16 28	Analista de Sistemas	Manabí	MASCULINO	2015.000	ANALISTA DE SL...	Entre \$1.000 y \$...	Nacional	PRIVADA	1,000 1,000
	17 26	Desarrolladora de Software	Carchi	FEMENINO	2015.000	Desarrollo de s...	Menor a \$1.000	Nacional	PRIVADA	2,000 1,000
	18 35	PROPIETARIO Gerente	Pichincha	MASCULINO	2015.000	Administrativo	Entre \$1.000 y \$...	extranjera	PRIVADA	1,000 8,000
	19 27	Desarrollador	Bolivar	FEMENINO	2014.000	Desarrollo de s...	Menor a \$1.000	Nacional	PUBLICA	? 0,000
	20 35	Lider de proyectos	Pichincha	MASCULINO	2012.000	PROYECTOS	Entre \$1.000 y \$...	Nacional	PRIVADA	1,000 3,000
	21 27	Consultora Técnica	Pichincha	FEMENINO	2015.000	Consultor_Aud...	Menor a \$1.000	Extranjera	PRIVADA	1,000 1,000
	22 43	docente universitario	Chimborazo	FEMENINO	2003.000	DOCENTE	Menor a \$1.000	Nacional	PUBLICA	4,000 10,000
	23 40	Jefe de desarrollo de Software	Pichincha	MASCULINO	2004.000	Desarrollo de s...	Entre \$2.001 y \$...	Nacional	PUBLICA	7,000 15,000
	24 44	Asesor Informático	Tungurahua	MASCULINO	2015.000	Consultor_Aud...	Entre \$1.000 y \$...	Nacional	PRIVADA	6,000 19,000
	25 32	especialista de base de datos	Cotopaxi	MASCULINO	2013.000	SISTEMAS DE L...	Entre \$1.000 y \$...	Nacional	PUBLICA	1,000 5,000
	26 48	docente	Pichincha	MASCULINO	2002.000	DOCENTE	Entre \$2.001 y \$...	Nacional	PUBLICA	5,000 15,000
	27 27	Ingeniero de Desarrollo	Chimborazo	MASCULINO	2014.000	Desarrollo de s...	Entre \$1.000 y \$...	Nacional	PRIVADA	3,000 2,000

Figura 12. Datos de los Graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas – ESPE

3. Fase III: Preparación de los Datos.

1. Se analiza las columnas referentes a:

1. Sexo
2. Cargo Empresa
3. Tipo de Empresa
4. País
5. Provincia
6. Nacional/Extranjera

4. Fase IV: Modelado de Datos.

En esta fase se hace uso de los algoritmos que posee Orange Canvas, estos se aplican de acuerdo a las necesidades del proyecto, y permitirán evaluar la información que posee la base de datos, enfocado a los resultados que se requiere obtener y los objetivos que se busca cumplir, los modelos y algoritmos a utilizar serán K-means, clusterización, árboles de decisión

En el Figura 13 se puede observar los ficheros que posee el software Orange utilizados para obtener la gráfica de los graduados en relación al campo laboral.



Figura 13. Widgets o ficheros utilizados para obtener la gráfica de los graduados en relación al campo laboral.

En el análisis de la pertinencia del perfil de egreso de los graduados en relación al campo laboral en el que se desenvuelven actualmente, se emplean ficheros que van relacionados acorde a los resultados que se busca obtener.

El proceso utilizado en este análisis es el siguiente: el widget-archivo, lee los datos de entrada y envía a la Tabla de Datos, en donde se puede visualizar el conjunto de datos que posee la base de datos, luego se hace uso del fichero Selección de Columnas para seleccionar las característica de datos, las clases o las variables meta que se va a emplear en el cálculo, en este particular se selecciona Cargo, que en la data hace referencia a una característica de tipo nominal .

Esto es enviado al fichero K-means que hace posible la selección del algoritmo siendo este el más apropiado a utilizar en el análisis, la información de la base de datos evaluada con clustering-kmeans se presentar en 4 clusters siendo estos los más relevantes. En el grafico 4 y 6 se muestra la selección y aplicación de este algoritmo, luego de esto se efectúa la relación con el fichero de visualización “Cuadro de Caja” para obtener y representar el resultado

En el Figura 14 se visualiza la selección de la característica de tipo nominal “Cargo” y el Figura 15 muestran la selección y ejecución del algoritmo k-means.

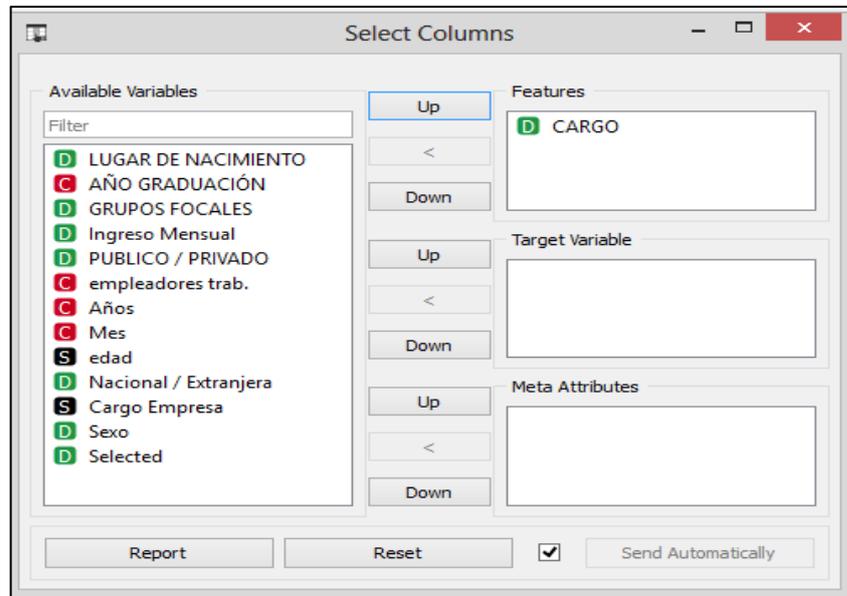


Figura 14. Selección de Datos

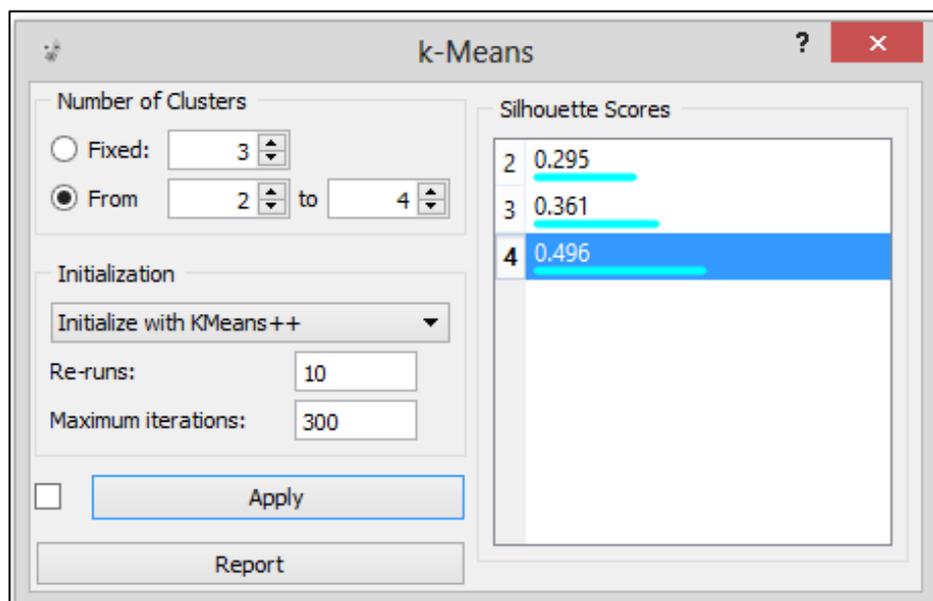


Figura 15. Algoritmo K-means

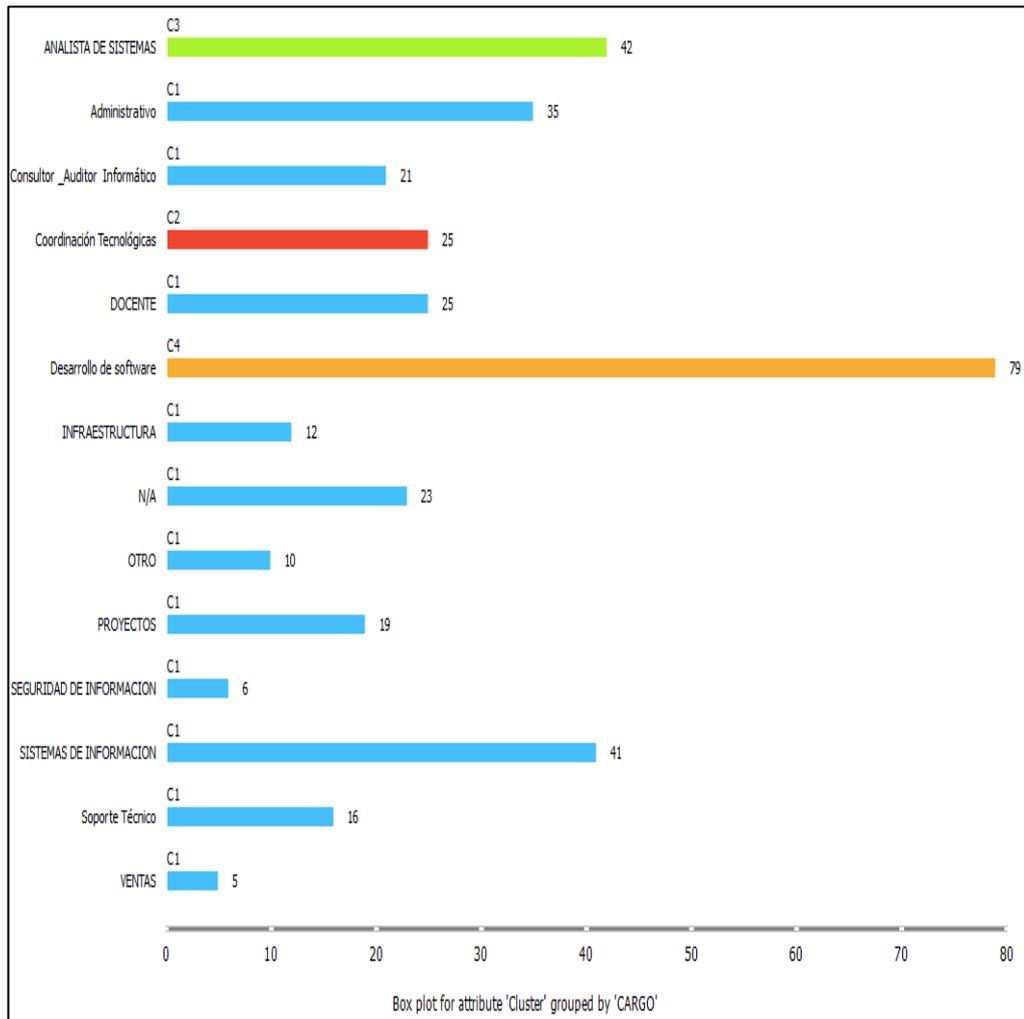


Figura 16. *Clustering* - Cuadro de Caja de los graduados en relación al campo laboral

La ejecución de ficheros para obtener resultados sobre el campo en que se encuentran laborando con mayor incidencia las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE. Se lo puede visualizar en el Figura 16.

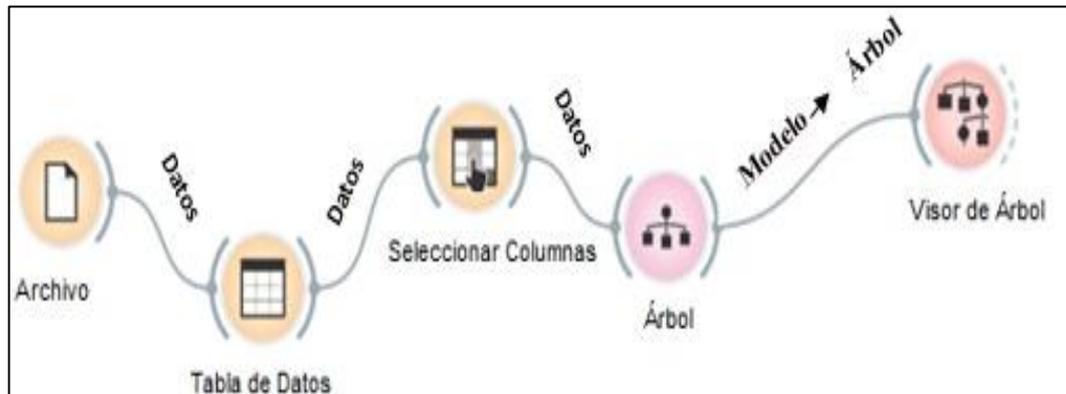


Figura 17. Ficheros utilizados para obtener la gráfica del campo en que se encuentran laborando con mayor incidencia las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE.

Para realizar el análisis sobre el campo en que se encuentran laborando con mayor incidencia las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE. En el Figura 17, se observa que los datos de entrada son leídos en el fichero Archivo y enviados a la Tabla de Datos en donde se puede observar el conjunto de datos dispuestos a ser valorados y seleccionados en el fichero Seleccionar Columnas, para este análisis se elige como característica: Cargo y como objetivo variable: Sexo, siendo estas dos variables de tipo “nominal” visualizar Figura 17. El siguiente parámetro de conexión es el fichero árbol de decisión, en donde se hace uso del algoritmo para analizar estos datos. Ver Figura 18, siendo sus resultados visualizados en un Figura de Visor de Árbol. (Ver Figura 18).

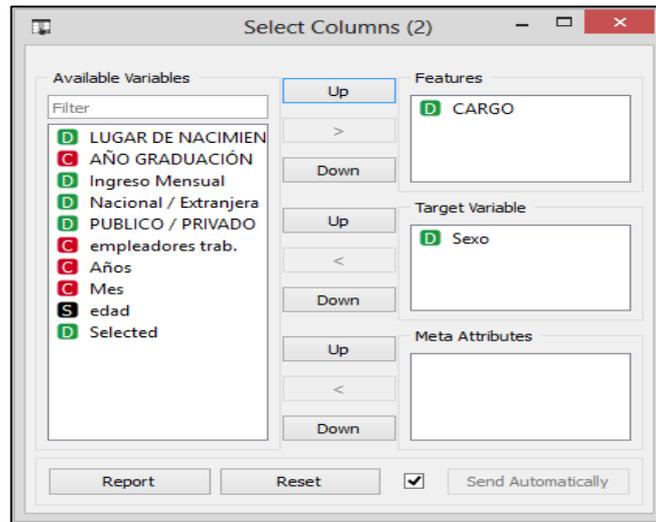


Figura 18. Selección de la característica cargo y del objetivo variable sexo.

El Figura 19 visualiza la selección y ejecución del algoritmo de árbol de decisión.

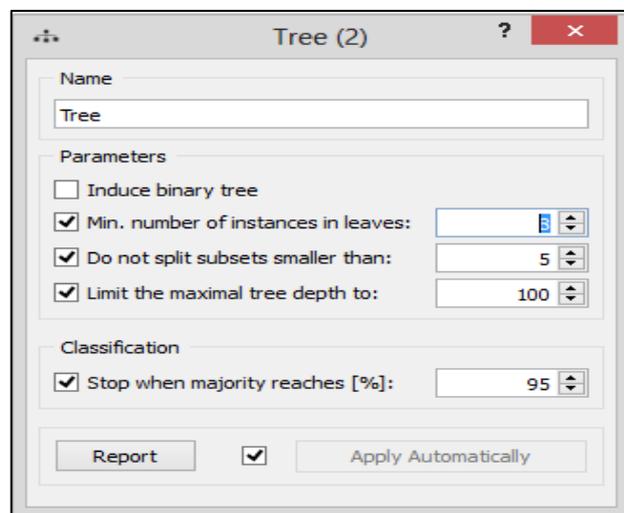


Figura 19. Ejecución del algoritmo de árbol de decisión

En el Figura 20 se encuentra representado el resultado al analizar el campo en que se encuentran laborando con mayor incidencia las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE

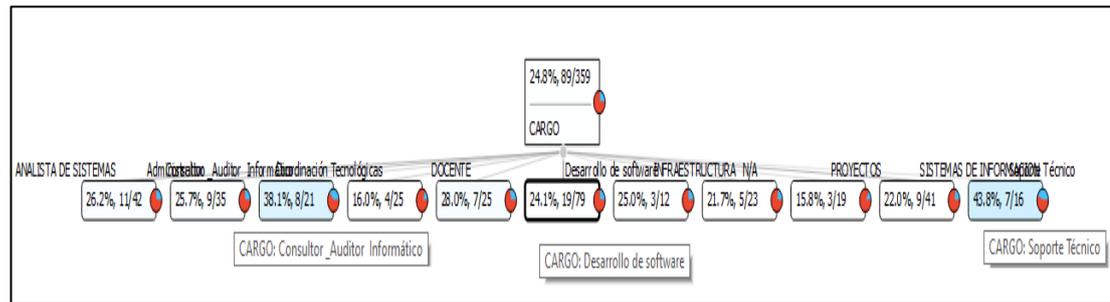


Figura 20. Visor de Árbol campo en que se encuentran laborando con mayor incidencia las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE

En el Figura 20 se puede observar cómo se realiza el proceso para determinar el nivel de incidencia de los graduados de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE en el mercado laboral ecuatoriano. Para ello los datos de entrada son cargados y leídos en el fichero Archivo y enviados a la Tabla de Datos, que contiene un conjunto de datos que en la ficha Seleccionar Columnas serán analizados y dispuestos a ser valorados y seleccionados.

Nacional/Extranjera hace referencia a la ubicación de la empresa en donde labora el graduado, esta característica se encuentra especificada en la data como un tipo de dato nominal (Ver Figura 21). El siguiente parámetro de conexión es el fichero que hace uso del algoritmo k-means para clusterizar haciendo uso de dos clusters y del árbol de decisión, para establecer los nodos predominantes (Ver Figura 22), siendo sus resultados visualizados en un Figura Visor de Árbol, representado en la Gráfica 21.

En el Figura 21 se observa la selección de los Ficheros utilizados para obtener el nivel de incidencia de los graduados en el mercado laboral ecuatoriano.

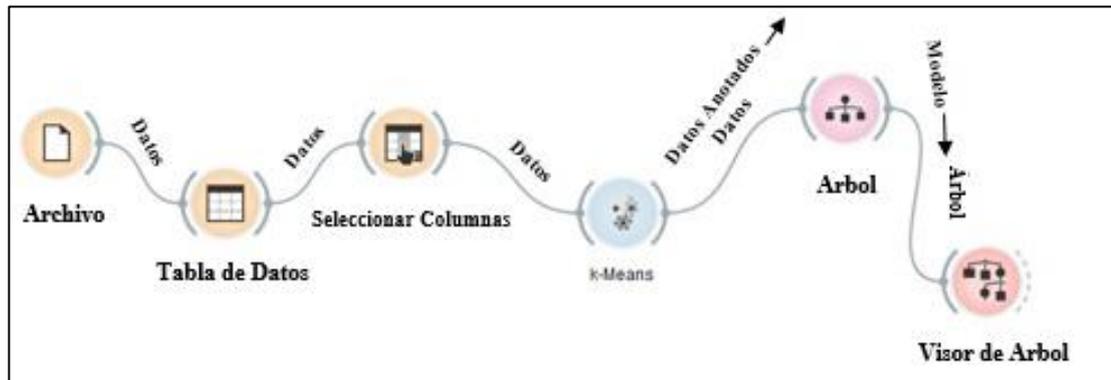


Figura 21. Ficheros utilizados para obtener la gráfica que determinara el nivel de incidencia de los graduados de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, en el mercado laboral ecuatoriano.

La característica electa en el fichero Seleccionar Datos que permitirá realizar el respectivo análisis para dar cumplimiento al objetivo planteado se visualiza en el Figura 13.

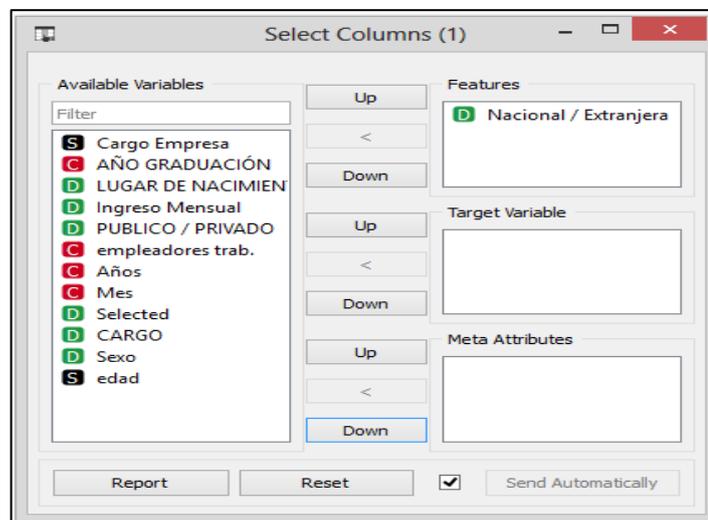


Figura 22. Selección de la característica Nacional/Extranjera en el fichero Selección de datos

En el Figura 23 se presentan los algoritmos k-meas y el de árbol de decisión aplicados al análisis de este proyecto.

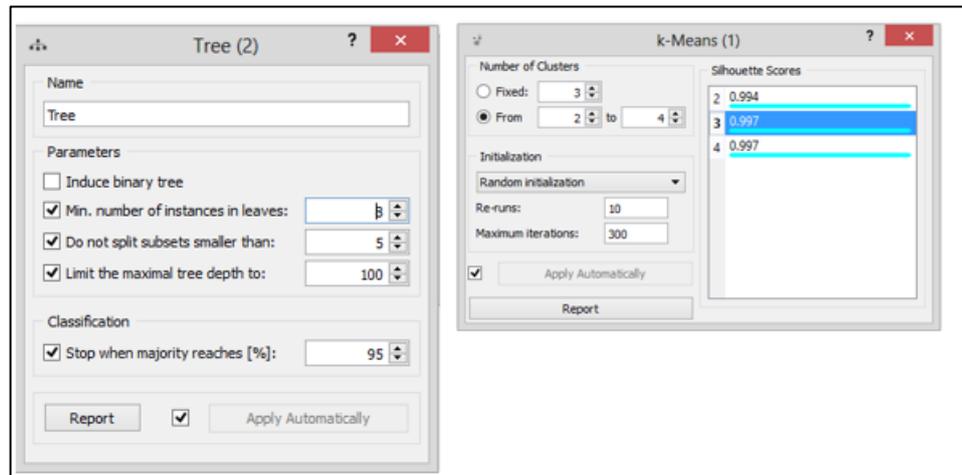


Figura 23. Ejecución del algoritmo k-means y del algoritmo de árbol de decisión

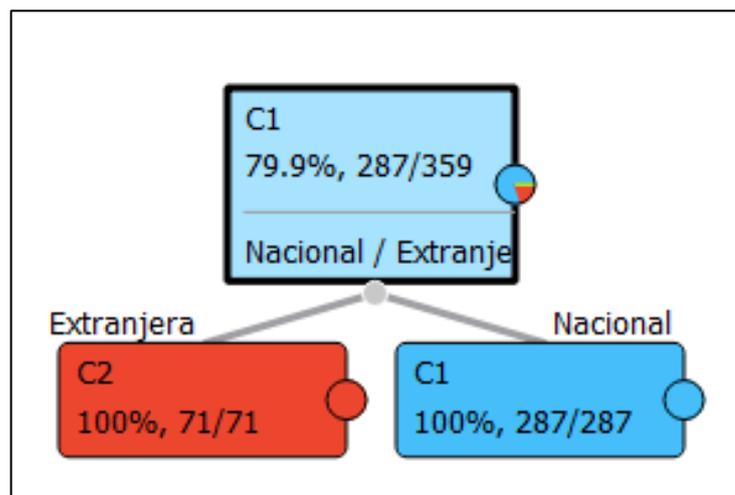


Figura 24. Visor de Árbol Nivel de incidencia de los graduados en el mercado laboral ecuatoriano de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE

5. Fase V: Evaluación de Datos.

Esta es una evaluación técnica, basada en el resultado del modelado, resume los resultados de esta tarea, lista la calidad de los modelos generados.

3.1.17. Clustering Cuadro de Caja.

El Figura 7, en donde los datos se encuentran representados en la Gráfica denominada Cuadro de Caja, se denota el modelo llamado clusterización que evalúa 4 clusters: C1 que se visualiza de color celeste, C2 de color rojo, el color anaranjado corresponde al cluster C3 y el cluster C4 está representado por el color verde.

Dentro de los más representativos está el Desarrollo de Software con 79 estudiantes, que vendría a ser el 22% de graduados que se encuentran desempeñando ese cargo. Seguido se encuentra Analista de Sistemas con 42 estudiantes, que representa el 12% de los graduados, como tercera opción se encuentra Sistemas de Información con 41 graduados, equivalente a un 11%, con 35 graduados que representa un 10% se visualiza el área Administrativa. El resto de campos toma valores de 25 hasta 5 graduados por campo.

3.1.18. Árbol de Decisión - Visor de Árbol.

El análisis ejecutado en el árbol de decisión representado en el Figura11, denominado Visor de Árbol, se puede recalcar la cantidad de género femenino, siendo las 89 mujeres las que se visualizan en la gráfica que representa al 24.8%. Los nodos más relevantes que señalan los campos con mayor número de graduadas laborando son en: Desarrollo de Software con 19 graduadas, Analistas de Sistema 11, en el área Administrativa 9, Sistemas de Información 9, Consultor_Auditor Informático 8, Soporte Técnico 7, Docentes 7, el resto de campos tienen parámetros de 5,4 y 3.

3.1.19. Clustering –Visor de Árbol.

La ejecución de la clusterización visualizada en la gráfica 15 denominada Visor de Árbol se puede observar como el nodo con una ruta más relativa y acentuada es el nodo Nacional con 287 graduados que representa el 79.9%,

laboran en el mercado ecuatoriano y 71 graduados que viene hacer el 20.1.%, laboran en empresas extranjeras.

6. Fase VI: Desarrollo.

La facilitación de un reporte o informe de los resultados obtenidos en el que se puede especificar que los estudiantes graduados de la carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE. Se encuentran laborando en áreas acorde al perfil de egreso de su carrera esto en cuanto al Figura de Clustering- Cuadro de Caja.

Si se analiza el reporte del Figura Visor de Árbol se puede informar que los graduados de género femenino se encuentran desempeñando cargos en Desarrollo de Software seguido de Analistas de Sistemas y de Sistemas de Información. Una de las expectativas muy claras en el último del Figura de Árboles es que se visualiza como el mercado laboral ecuatoriano es quien cubre la mayoría de las vacantes de los graduados en las áreas antes mencionadas. Se puede informar que se debe tratar de vincular a los estudiantes que realizan vinculación con la sociedad o sus horas pre profesionales en los campos profesionales que no están siendo copados en nuestro medio y así ampliar la inserción laboral

El Departamento de Seguimiento a Graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, actualmente deberían tener el conocimiento de la factibilidad de emplear nuevas metodologías y procesos como es Data Mining, que facilitan la extracción de los datos, que asociados se convierten en información. La misma que si es empleada de manera correcta, ejecutando las necesidades específicas requeridas podrá proporcionar resultados que peritaran aportar y satisfacer razonablemente las necesidades de los ex y futuros graduados.

Sin embargo, es necesario que los delegados de Seguimiento a Graduados de las diferentes áreas de estudio se comprometan con el

proceso que se encuentra a su cargo, solicitando se proporcione una capacitación de uso de nuevos procesos, estrategias y herramientas informáticas que les facilite obtener los resultados solicitados en menor tiempo, esto permitirá aportar valor agregado en cada uno de los procesos solicitados y dar respuesta inmediata para que se ejecute los cambios solicitados.

3.2. Discusión y Resultados

En este punto se valorará cada una de las fases de CRISP-DM, que se encuentran enlazadas en los objetivos específicos de este proyecto.

En la fase I, en donde se detalla la comprensión del negocio o problema, se pudo comprender que los objetivos que se plantearon en el proyecto determinaron el nivel de alcance de los cargos desempeñados por los graduados, con respecto al sexo y al lugar de residencia (país, provincia), así como también establecer posibles técnicas y algoritmos a emplear como K-means, clustering y árboles de decisión.

En la fase II, la cual se puntualiza en la comprensión de los datos, permitió recolectar información de 743, datos a través de la existencia y disponibilidad de una base de datos de los graduados desde el año 2002 hasta el 2016, se analizaron cada uno de los datos, verificando su consistencia, veracidad y calidad.

Con la fase III, que corresponde a la preparación de los datos, se logró seleccionar los registros y atributos válidos, así como transformar y limpiar los 743 datos. Para ejecutar la limpieza de la data se recurrió, en su mayoría, a completar datos nulos e inválidos, a través de consultas en portales web como LinkedIn y Facebook que permitieron verificar y completar datos como: el nombre del lugar de trabajo del graduado, desempeño laboral y su residencia actual.

En la página de la Senescyt se consultó el año de graduación del graduado, y si la mención del título profesional tenía coincidencia con los datos expuestos en la data. La web del Registro Civil permitió verificar el número de cédula y la coincidencia de los datos personales, además se organizaron los datos iguales o parecidos de la data es decir, asignando un mismo nombre que identifique a ese conjunto de datos semejantes. De esta forma se lograron catorce designaciones en los cargos que desempeñan. Se eliminaron datos atípicos y ruidosos, después de todo el proceso se construyó una nueva data con 359 datos validos que fueron los dispuestos a ser valorados, modelados y analizados en Orange.

En la fase IV, que hace referencia al modelado de datos, se hizo uso de los modelos y algoritmos de Orange, por ejemplo: para evaluar la pertinencia del perfil de egreso de los graduados con respecto al cargo laboral que desempeña, se empleó K-means – clustering, el siguiente modelo y algoritmo utilizado fue el de árbol de decisión, para determinar el campo laboral más demandante, de acuerdo al perfil de egreso de las mujeres graduadas de la facultad de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas – ESPE. Por último, para determinar el nivel de incidencia de los graduados en el mercado laboral ecuatoriano, se aplicó clusterización-kmeans.

La fase V, que es la evaluación de los datos, condescendió a certificar el cumplimiento de los objetivos con los resultados y porcentajes obtenidos al ejecutar el proceso de la fase IV. Y para finalizar, en la última fase, que corresponde al desarrollo, se emitió un informe del cumplimiento y porcentaje obtenido en cada modelo ejecutado y posibles sugerencias.

Proporcionado este análisis se puede concluir que el estudio de la pertinencia del perfil de egreso, con respecto a los graduados de Universidad de las Fuerzas Armadas-ESPE, ha sido valorativa, porque se puede conocer que la oferta educativa propuesta ha entregado buenos resultados al mundo laboral, permitiendo cubrir necesidades laborales y

económicas del entorno, al mismo tiempo que permitió conocer y emplear herramientas de minería de datos, lo que antecede un autoaprendizaje que desarrolla la capacidad creativa, dando solución a problemas y generando productividad.

CAPÍTULO IV

CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURO

4.1. Conclusiones

1. La metodología que se ejecutó en este proyecto, me permitió valorar el ciclo de vida de este trabajo, permitiéndome moldear los datos conforme a mi necesidad y lograr generar a través de los modelos y algoritmos de minería de datos la información requerida para dar cumplimiento a los objetivos planteados.
2. Los 349 graduados de la carrera de ingeniería de sistema de la Universidad de las Fuerzas Armadas- ESPE, se encuentran actualmente laborando en empleos acorde al perfil de egreso de la carrera de ingeniería de sistemas siendo 10 graduados los que no se encuentran laborando en campos acorde a su perfil, y 23 de los mismos no registran campo laboral. Los resultados obtenidos dan cumplimiento a uno de los objetivos de la unidad de titulación normada en el “Proyecto de Diseño de la Carrera de Ingeniería en Tecnologías de la Información”, que se enfoca en que los graduados estén aptos a Integrar conocimientos teóricos y prácticos que conduzcan a la solución de problemas de conectividad e integración de diferentes plataformas e infraestructuras tecnológicas.
3. El análisis de la pertinencia de perfil de egreso en el campo laboral permitió validar que las mujeres graduadas de la Universidad de las Fuerzas Armadas-ESPE, se desempeñan profesionalmente en el campo de Desarrollo de Software, cubriendo un 24.1% en relación a los demás campos.

4. La incidencia de los 287 graduados de la Carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE es de un 79.95% en el mercado laboral ecuatoriano.

4.2. Líneas de trabajo futuro.

1. Dado que la carrera de Ingeniería de Sistemas de la Universidad de las Fuerzas Armadas-ESPE, se ha rediseñado, siendo actualmente la carrera de Ingeniería en Tecnologías de la Información y la Comunicación, es pertinente indicar que se debe tomar en cuenta este análisis para evaluar las competencias, destrezas, conocimientos y habilidades propuestas en la malla curricular de la carrera, permitiendo superar la trascendencia y excelencia que posee, y el resultado obtenido sea el desempeño profesional y competente que la caracteriza.
2. Se recomienda evaluar y agregar variables en la base de datos de este proyecto, analizar la data aplicando otros softwares como rapidminer, weka entre otros y ejecutar modelos predictivos, algoritmos de regresión lineal en datos numéricos.

BIBLIOGRAFÍA

- Vallejos, S. (2006). *Minería de Datos*. Obtenido de Minería de Datos:
http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_Datos_Vallejos.pdf
- ACR, D. e. (09 de 08 de 2016). Universidades generan Bolsa de Empleo.
Universidades generan Bolsa de Empleo.
- Aguiar, V. (2007). El Mercado Laboral Ecuatoriano: Propuesta de una reforma. En V. Aguiar, *El Mercado Laboral Ecuatoriano: Propuesta de una reforma* (págs. 9,10). Edición electrónica.
- Aguiar, V. (2007). El Mercado Laboral Ecuatoriano: Propuesta de una reforma. En V. Aguiar, *El Mercado Laboral Ecuatoriano: Propuesta de una reforma*. (pág. 12). El Mercado Laboral Ecuatoriano: Edición electrónica.
- Aguiar, V. (2007). El Mercado Laboral Ecuatoriano: Propuesta de una reforma. En A. Víctor, *El Mercado Laboral Ecuatoriano: Propuesta de una reforma*. (pág. 69). eumed.net. Obtenido de El Mercado Laboral Ecuatoriano: Propuesta de una reforma.: <http://www.eumed.net/libros-gratis/2007a/240/240.pdf>
- Almirón, A. (2013). *Ofinal Final Trabajo Inteligencia de Negocios Aa*. Obtenido de Inteligencia de Negocios:
<https://es.scribd.com/presentation/322599922/Ofinal-Final-Trabajo-Inteligencia-de-Negocios-Aa>
- Arancibia, J. A. (s.f.). *Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM*. Obtenido de Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM:
http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRISP-DM.2385037.pdf
- Asamblea Nacional. (12 de Octubre de 2010). *Sección Tercera Del Funcionamiento de las Instituciones de Educación Superior*. Obtenido

de Art. 142.- Sistema de seguimiento a graduados:

<http://educaciondecalidad.ec/leyes-sistema/ley-educacion-superior-loes.html>

Beltrán, C. A. (Julio de 2014). *Interfaces de usuario “Explorer” y “Knowledge*.

Obtenido de Interfaces de usuario “Explorer” y “Knowledge:

<http://www.fce.unal.edu.co/media/files/documentos/uifce/proyectos/Interfaces%20de%20usuario%20Explorer%20y%20Knowledge%20Flow%20en%20Weka.pdf>

Berzal, J. C. (s.f.). *DECSAI Sistemas Inteligentes de Gestión: KNIME* .

Obtenido de DECSAISistemas Inteligentes de Gestión: KNIME |:

<http://elvex.ugr.es/decsai/intelligent/workbook/D1%20KNIME.pdf>

Calleja, J., & Gómez, A. (Diciembre de 2010). *Minería de datos con weka para la*. Obtenido de minería de datos con weka para la.

Cano, J. (2007). business intelligence - Esade. *Business*

intelligence: competir con información, 32. Obtenido de

http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf

Cano, J. L. (2013). Business intelligence: competir con información. En J. L.

Cano, *business intelligence: competir con información* (págs. 93-94).

Obtenido de

http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf

Cardenas, J., Garrido, M., Muñoz, E., & Serrano, E. (2007). *CEDI 2007:*

Mineria de datos aplicada a la modelizacion de un sistema hibrido.

ZARAGOZA: ISBN 978-84-9732-602-5 Editorial Thomson. Obtenido

de <http://www.lsi.us.es/redmidas/CEDI07/%5B18%5D.pdf>

Cárdenas, M. (2015). *PREPROCESADO DE DATOS PARA MINERIA*.

Obtenido de PREPROCESADO DE DATOS PARA MINERIA:

<http://wwwae.ciemat.es/~cardenas/docs/lessons/PreprocesadoDatos.pdf>

- Ciro Rodríguez León, D. C. (2016). Adecuación a metodología de minería de datos para aplicar a problemas no supervisados tipo atributo-valor. *SCIELO*. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202016000400005
- Cobos, C., Zuñiga, J., Guarín, J., León, E., & Mendoza, M. (2010). CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *redalyc.org.Ingeniería e Investigación*, 47. Obtenido de <http://www.redalyc.org/articulo.oa?id=64316140004>
- Cobos, C., Zuñiga, J., Guarín, J., León, E., & Mendoza, M. (3 de diciembre de 2010). CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. *redalyc.org. Red de Revistas Científicas de América Latina y el Caribe, España y Portugal*, pp. 45-56. Obtenido de <http://www.redalyc.org/articulo.oa?id=64316140004>
- Córdoba Fallas, L. (junio de 16 de 2011). *Minería de Datos*. Obtenido de Minería de Datos: <http://cor-mineriadedatos.blogspot.com/2011/06/weka.html>
- Córdova, C. (2014). *“Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes que pertenecen al colegio fisco-misional “san francisco” de la ciudad de ibarra”*. IBARRA.
- Cultura CRM Magazine . (27 de Julio de 2016). *Orange Data Mining, análisis de datos científicos*. Obtenido de Orange Data Mining, análisis de datos científicos: <http://culturacrm.com/crm/recursos-crm/orange-data-mining-analisis-datos/>
- DATA PRIX. (12 de Mayo de 2009). *Business Intelligence*. Obtenido de Business Intelligence: <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/1-business-intelligence/13-proceso-bi>

Dávila, I. (2012). *Informe del Estudio de Seguimienot a Graduados*. Obtenido de <http://www.uti.edu.ec/documents/SeguimientoGraduadosEducacion2012.pdf>

EcuRed. (5 de junio de 2017). *EcuRed*. Obtenido de EcuRed: <https://www.ecured.cu/Weka>

Espiñeira, Sheldon y Asociados. (2008). *La inteligencia de negocios - (business intelligence) - PwC*. Obtenido de La inteligencia de negocios - (business intelligence) - PwC: <https://www.pwc.com/ve/es/asesoria-gerencial/boletin/assets/boletin-advisory-edicion-10-2008.pdf>

Fuentes, D., & Espinal, R. (2015). Proceso de seguimiento a graduados en la facultad de ciencias económicas de la universidad laica eloy alfaro de manabí. *Revista Electrónica Formación y Calidad Educativa (REFCaE)ISSN 1390-9010*, 63.

Fuertes, L. (11 de 07. de 2013). *Minería de Datos Data Mining*. Obtenido de Academia.edu: http://www.academia.edu/8111501/Miner%C3%ADa_de_Datos_Data_Mining

García González Francisco José. (2013). *Aplicación de técnicas de Minería de Datos a datos obtenidos por el Centro Andaluz de Medio Ambiente (CEAMA)*. Obtenido de http://masteres.ugr.es/moea/pages/tfm-1213/tfm_garciagonzalezfrancisco_1/

García González Francisco José. (2013). *Aplicación de técnicas de Minería de Datos a datos obtenidos por el Centro Andaluz de Medio Ambiente (CEAMA)*. Obtenido de Aplicación de técnicas de Minería de Datos a datos obtenidos por el Centro Andaluz de Medio Ambiente (CEAMA): http://masteres.ugr.es/moea/pages/tfm-1213/tfm_garciagonzalezfrancisco_1/

GARCÍA, J. H. (2010). *Docentes - Universidad Nacional de Colombia*. Obtenido de Docentes - Universidad Nacional de Colombia:

http://www.docentes.unal.edu.co/hrumana/docs/TESIS_JHMG_Inteligencia_de_Negocios_2010.pdf

García, J., & Martínez, H. (2010). *Business intelligence as a tool for strategic decision making in business. Analysis of its applicability in the Colombian corporate context*. Obtenido de Business intelligence as a tool for strategic decision making in business. Analysis of its applicability in the Colombian corporate context.:

<http://docplayer.es/4991931-Business-intelligence-as-a-tool-for-strategic-decision-making-in-business-analysis-of-its-applicability-in-the-colombian-corporate-context.html>

García, M. D. (s.f.). *Manual de weka*. Obtenido de manual de weka:

<http://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/InteligenciaDeNegocio/weka.pdf>

Gavilanes, J. E., & Romero, A. (2015). La empleabilidad de graduados universitarios en el contexto. *Atenas. Revista Científico Pedagógica*, 4.

Gervilla García, E., Jiménez López, R., Montaña Moreno, J. J., Sesé Abad, A., & Cajal. (200). La metodología del Data Mining. Una aplicación al consumo de alcohol en adolescentes. *redalyc.org*, pp. 65-80.

Obtenido de <http://www.redalyc.org/articulo.oa?id=289122882009>

gestiopolis. (14 de Julio de 2004). *Procesos de la inteligencia de negocios*.

Obtenido de Procesos de la inteligencia de negocios:

<https://es.scribd.com/presentation/322599922/0final-Final-Trabajo-Inteligencia-de-Negocios-Aa>

Guzmán, E. L. (s.f.). *Metodologías aplicadas al proceso de Minería de Datos*.

Obtenido de Metodologías aplicadas al proceso de Minería de Datos:

http://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf

Héctor Oscar Nigro, D. X. (s.f.). *Un proceso centrado en el usuario*.

Obtenido de Un proceso centrado en el usuario :

http://sedici.unlp.edu.ar/bitstream/handle/10915/21220/Documento_completo.pdf?sequence=1

IED.Escuela de negocios. (s.f.). *Manual básico KNIME*. Obtenido de http://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/InteligenciaDeNegocio/manual_basico_knime.pdf

INEC. (2014). *ecuadorencifras*. Obtenido de Sistema Integrado de Encuestas Encuesta Nacional de Empleo, Desempleo y Subempleo Indicadores Laborales : http://www.ecuadorencifras.gob.ec/documentos/web-inec/EMPLEO/Empleo-mar-2014/10_anios/201403_EnemduPresentacion_10anios.pdf

INEC. (17 de Octubre de 2016). *INEC publica cifras del mercado laboral de septiembre 2016*. Obtenido de INEC publica cifras del mercado laboral de septiembre 2016: <http://www.ecuadorencifras.gob.ec/inec-publica-cifras-del-mercado-laboral-de-septiembre-2016/>

Ismael, T. V. (s.f.). *Manual Básica Knime*. Obtenido de Manual Básica Knime: http://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/InteligenciaDeNegocio/manual_basico_knime.pdf

Izquierdo Iglesias, A., Bravo Ilisástigui, L., Ceruto Cordovéz, T., Martín, & Diana, R. (2015). Nuevos plugins para la herramienta Knime para el uso de sus flujos de trabajo desde otras aplicaciones. *Ciencias de la Información*, 48. Obtenido de <http://www.redalyc.org/pdf/1814/181439409007.pdf>

Jiménez, M. S. (6 de Marzo de 2008). Tendencias y hallazgos en los estudios de trayectoria: una opción metodológica para clasificar el desarrollo laboral. *SCIELO*.

Kammel, L. K. (Noviembre de 2015). *La educación, el empleo y los hábitos*. Obtenido de la educación, el empleo y los hábitos: <http://eprints.ucm.es/34882/1/20151118%20TFM%20Loubna%20Khali%20fi.pdf>

- Kammel, L. K. (Noviembre de 2015). *La educación, el empleo y los hábitos*.
Obtenido de la educación, el empleo y los hábitos:
<http://eprints.ucm.es/34882/1/20151118%20TFM%20Loubna%20Khali%20fi.pdf>
- López, D. (2013). Análisis de las posibilidades de uso de Big Data en las organizaciones. *Analysis of the possibilities of use of Big Data in organizations*. Obtenido de
<https://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADaS.pdf?sequence=1>
- Luhn, H. (1958). A Business Intelligence System. *IBM JOURNAL*, 315-317.
Obtenido de <http://altaplana.com/ibmrd0204H.pdf>
- Madrid, E. P. (2008). LOS INSUMOS INVISIBLES DE DECISIÓN: DATOS, INFORMACIÓN Y CONOCIMIENTO. *Anales de Documentación*, 3. Obtenido de
<http://www.redalyc.org/articulo.oa?id=63501110>
- Malagón, A. (2003). La Pertinencia en la Educación Superior: Elementos para su comprensión. *Revista de la Educación Superior*,
<http://publicaciones.anuies.mx/acervo/revsup/127/03.html>.
- MANABÍ, U. L. (2015). *Esudio de Graduados de la Carrera de Enfermería*. Chone - Ecuador.
- MARTÍNEZ ÁLVAREZ, C. (Octubre de 2012). Aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión en entel. *Aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión en entel*. Santiago, Chile.
- Martínez, H. (2010). *La inteligencia de negocios como herramienta para la toma de*. Obtenido de la inteligencia de negocios como herramienta para la toma de:
<http://www.bdigital.unal.edu.co/3098/1/940607.2010a.pdf>

- Martínez, M. C. (2003). Minería de Datos . *Conciencia Tecnológica*, 1.
- Mederos, J. F., & Vega, E. (2003). El desarrollo laboral de los jóvenes profesionales: 20 años de estudios de seguimiento de los egresados de las universidades cubanas. *Conferencia Magistral presentada en el Congreso de Ciencias de la Educación, Facultad de Educación, UADY*, (pág. 100). Mérida, Yuc.
- Méndez, M. (2014). Mercado Laboral.
<https://insercionlaboralegresados.jimdo.com/marco-te%C3%B3rico/definici%C3%B3n-de-conceptos-b%C3%A1sicos/>
- Mercado Laboral GTH. (28 de Julio de 2011). *Mercado Laboral Definición*.
 Obtenido de <http://mercadolaboraledu.blogspot.com/2011/07/se-denomina-mercado-de-trabajo-o.html>
- microsystem*. (s.f.). Obtenido de microsystem:
<http://www.microsystem.cl/plataforma/rapidminer/>
- Morueta, R. T., Tejeda, R., & Cedeño, G. (2015). Implementación institucional de un modelo cooperativo para el seguimiento a graduados en el Ecuador. *Revista de la Educación Superior*, 126.
- Ordorika, I. (2015). Revista de la Educación Superior. *Revista de la Educación Superior*, 135-136.
- Organización Internacional del Trabajo. (2016). Panorama Laboral 2016. América Latina y el Caribe. *Panorama Nacional 2016*, 8.
- Pelayo, M. (2012). Capital social y competencias profesionales: Factores condicionantes para la inserción laboral. En M. B. Pelayo Pérez, *Capital social y competencias profesionales* (pág. 18). Tepic, Nayarit, México.: Edición electrónica.
- Pérez, G. J. (s.f.). *Minería de datos Metodología Sema*. Obtenido de <https://www.emaze.com/@ALRQLILT>
- Pete Chapman (NCR), J. C. (1999-2000). *CRISP-DM 1.0 Step-by-step data mining guide*. Obtenido de CRISP-DM 1.0 Step-by-step data mining

guide: <http://www-staff.it.uts.edu.au/~paulk/teaching/dmkdd/ass2/readings/methodology/CRISPWP-0800.pdf>

Ps.Ind.Acosta, P., & Dra. Escobar, N. (2011). *Informe de Estudio de Seguimiento a Graduados*. Ambato.

Quintanilla, G., & Gil, J. R. (2016). Gobierno abierto y datos vinculados: *Revista del CLAD Reforma y Democracia*, 75. Obtenido de <http://www.redalyc.org/articulo.oa?id=357546620003>

Quintanilla, G., & Gil-García, J. R. (junio de 2016). Gobierno abierto y datos vinculados: conceptos, experiencias y lecciones con base en el caso mexicano. *Revista del CLAD Reforma y Democracia*, 76 y 93. Obtenido de <http://www.redalyc.org/articulo.oa?id=357546620003>

Quiroz Gil, N. L., & Valencia, C. A. (2012). Aplicación del proceso de KDD en el contexto de bibliomining: El caso Elogim. *Revista Interamericana de Bibliotecología*, vol. 35, núm. 1., pp. 97-108. Obtenido de <http://www.redalyc.org/articulo.oa?id=179024991009>

Rapid Miner, Software Business Intelligence. (26 de Marzo de 2015). Obtenido de Rapid Miner, Software Business Intelligence: <http://www.postecnologia.com/2015/03/rapid-miner-software-business-intelligence.html>

Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias|q. *Revista Iberoamericana de Inteligencia Artificial*, 3. Obtenido de <http://www.redalyc.org/articulo.oa?id=92502902>

Rodríguez Montequín, M. T., Álvarez Cabal, J. V., Mesa Fernández, J. M., & González Valdés, A. (s.f.). *Metodologías para la realización de proyectos de data mining*. Obtenido de metodologías para la realización de proyectos de data mining : http://www.aepro.com/files/congresos/2003pamplona/ciip03_0257_0265.2134.pdf

- Rodríguez Suárez, Y., & Díaz Amador, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 75.
- Rodríguez Suárez, Y., & Díaz Amador, A. (3-4 de julio-diciembre de 2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 77.
- Rodríguez, Y., & Anolandy, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 75. Obtenido de <http://www.redalyc.org/articulo.oa?id=378343637009>
- Rodríguez, Y., & Díaz, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 5-6.
- Romero, C., Ventura, S., & Hervás, C. (2005). *Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web*. Córdoba: ISBN 84-9732-449-8 Editorial Thomson. Obtenido de <file:///C:/Users/Diana/Downloads/Nueva%20carpeta/MINERIA%20DE%20DATOS/Taller%20Nacional%20de%20Miner%C3%ADa%20de%20Datos%20y%20Aprendizaje,%20TAMIDA2005.pdf>
- Sánchez, T., Molina, P., del Valle, R., Pascual, G., & Alvear, M. (2009). *Guías para la Renovación Curricular*. Obtenido de Orientaciones para la renovación curricular: http://www.academia.edu/16732792/Gu%C3%ADas_para_la_Renovaci%C3%B3n_Curricular
- Saraí, A. B. (2005). Fórmulas para el cálculo de la muestra en investigaciones de Salud. *redalyc.org*, 333-338. Obtenido de <http://www.redalyc.org/articulo.oa?id=48711206>
- Seguimiento a Graduados. (2012). *Estudio de empleabilidad de los graduados*. Obtenido de <http://www2.ucsg.edu.ec/transparencia-de-la-informacion/acreditacion-medicina/documentos-criterio-pertinencia/estado-actual-y-prospectiva/documentos-de-estado-actual-y-prospectiva/374-informe-empleabilidad-ucsg-2011-2012/file.html>

- Serrano, A. (Julio de 2013). *Análisis de condiciones de vida, el mercado laboral y los medios de producción e inversión pública*. Obtenido de Cuaderno de trabajo SENPLADES Subsecretaría de Inversión Pública Dirección de Objetivos y Manejo Sostenible de la Inversión Pública: <http://www.planificacion.gob.ec/wp-content/uploads/downloads/2013/08/An%C3%A1lisis-de-condiciones-de-vida-el-mercado-laboral-y-los-medios-de-producci%C3%B3n-e-inversi%C3%B3n-p%C3%BAblica-Cuaderno-de-trabajo-N.-3-SENPLADES1.pdf>
- Simón, J., & Arellano, L. (2009). Calidad profesional del Técnico Superior Universitario en Administración. Una visión de graduados y de empleadores. *Actualidades Investigativas en Educación*, 5. Obtenido de <http://revistas.ucr.ac.cr/index.php/aie/article/view/9542/17901>
- SOC Generalitat de Catalunya. (2013). El mercado laboral. *SOC Generalitat de Catalunya*, 1,2. Obtenido de EL MERCADO LABORAL: https://www.oficinadetreball.gencat.cat/socweb/export/sites/default/socweb_es/ciudadans/_fitxers/OrientacioTrobarFeina_06_EIMercatLaboral.pdf
- SUPO, D. J. (2014). Cómo elegir una muestra. En d. J. Supo, *cómo elegir una muestra* (págs. 13-31). Perú: bioestadístico eirl.
- Tirado, R., Tejada, R., & Cedeño, G. (2015). Implementación institucional de un modelo cooperativo para el seguimiento a graduados en Ecuador. *Revista de la Educación Superior*, 132,133.
- Torres, C., Esquivel, I., López, L., & Cardona, M. d. (27 y 28 de January de 2011). Señales de los alumnos sobre la efectividad de la. *Coloquio ANFECA: Hacia un espacio común de Educación Superior en las Facultades y Escuelas de Negocios en México PLESNA Siglo XXI*, (pág. 7). Veracruz, México. Obtenido de https://www.researchgate.net/publication/236935754_Senales_de_los

_alumnos_sobre_la_efectividad_de_la_imparticion_del_programa_educativo_Sistemas_Computacionales_Administrativos

Universidad Casa Grande. (2009). *Estudios de Seguimiento a Graduados*.

Obtenido de Estudio de Seguimiento:

<http://www.casagrande.edu.ec/casagrande/graduados/estudios-de-seguimiento/>

Urra, P., & Jiménez, J. (2011). *Perfil de Egreso Capítulo 2*. Obtenido de

MANUAL DE REVISIÓN Y DISEÑO: <http://www.unie.usach.cl/wp-content/uploads/download-manager-files/II.%20Perfil%20de%20egreso.pdf>

Vassallo, J. (2012). Business Intelligence y la importancia para la toma de decisiones gracias a la información en tiempo real. *BI Como herramienta*, 1. Obtenido de

<https://es.scribd.com/document/133980184/Business-intelligence-la-gestion-inteligente-de-la-informacion>

<https://es.scribd.com/document/133980184/Business-intelligence-la-gestion-inteligente-de-la-informacion>

Velarde, A. (2003). Minería de Datos. Una Introducción. *Conciencia*

Tecnológica, 3. Obtenido de

<http://www.redalyc.org/articulo.oa?id=94402303>

yoshibauco. (2011). *Metodología para desarrollo de Proyectos de Minería de*

Datos . Obtenido de Metodología para desarrollo de Proyectos de

Minería de Datos :

<https://yoshibauco.wordpress.com/2011/03/07/metodologia-para-desarrollo-de-proyectos-de-mineria-de-datos/>