UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS

FACULTAD DE CIENCIAS ECONÓMICAS

POSFACE

DIRECCIÓN DE ESTUDIOS DE POSTGRADO

MAESTRIA EN GESTIÓN INFORMATICA

TESIS

“APLICACIÓN DE MINERÍA DE DATOS COMO UNA METODOLOGÍA PARA ESTIMAR LAS PRINCIPALES CAUSAS DE DESEMPLEO Y SUBEMPLEO PROFESIONAL DE LOS EGRESADOS DE LAS UNIVERSIDADES”

SUSTENTADA POR

VANY JETSIBELH CASTILLO GUILLEN

PREVIO A OPTAR AL TÍTULO DE

MÁSTER EN GESTIÓN INFORMATICA

TEGUCIGALPA, HONDURAS

AUTORIDADES UNIVERSITARIAS

LICDA. JULIETA CASTELLANOS RUIZ

RECTORA

ABOG. EMMA VIRGINIA RIVERA MEJÍA

SECRETARIA GENERAL

LICDA. LETICIA SALOMÓN

DIRECTORA DEL SISTEMA

DE ESTUDIOS DE POSTGRADO

LICDA. BELINDA FLORES DE MENDOZA; M.A.

DECANA DE LA FACULTAD

DE CIENCIAS ECONÓMICAS

DR. JORGE ABRAHAM ARITA LEÓN

COORDINADOR GENERAL DE POSTGRADO DE

LA FACULTAD DE CIENCIAS ECONÓMICAS

Dedicatoria

A Dios, por permitirme llegar a este momento tan especial en mi vida, por iluminarme en cada momento y estar conmigo en cada paso que doy y por haber puesto en mi camino a aquellas personas que de muchas formas me han ayudado y me han acompañado durante todo el periodo de estudio. Dedico este y todos mis logros a Dios porque sin El no sería nada, y porque por tanto mis logros también lo son de él.

A mis padres, que son quienes mi inculcaron el amor a Dios y los valores con los que cuento. Gracias a mi Madre, por creer siempre en mí y por darme ánimos y apoyo incondicional sin importar la distancia y nuestras diferencias de opiniones. Gracias a mi Padre, porque a pesar de nuestra distancia física tengo la certeza de que siempre estás conmigo y a pesar de que nos faltaron tantas cosas por vivir juntos, sé que este logro sería tan importante para ti, como lo es para mí. Dedico este logro a mis padres porque dieron todo por mí y porque pensar que ellos estarán orgullosos de mí, me hace intentar ser mejor cada día.

Agradecimientos

Agradezco a Dios por darme la vida para cumplir otra de mis metas, por darme la fuerza, valor y paciencia para continuar cada día y porque nunca me abandonas, y has abierto las puertas necesarias para que yo pueda salir adelante.

Gracias a mi Madre por su confianza y apoyo incondicional, por todo lo que ha dado por mí, y porque sin su ejemplo y esfuerzos nunca hubiera llegado hasta aquí.

Agradezco también a mi Padre, porque pensar en el me hace levantarme cada día y porque sé que estaría orgulloso de la mujer en la que me he convertido, y aun estando lejos lo llevo siempre en mi mente y corazón.

A mis hermanos porque a pesar de nuestros problemas y diferencias, nunca me han abandonado y he aprendido mucho de ustedes. Gracias por estar conmigo.

A mi novio, que durante estos años ha sabido apoyarme y ha estado conmigo siempre, porque no me ha dejado renunciar y ha sacado lo mejor de mí a pesar de cualquier tormenta. Gracias por su amor incondicional y por hacerme mejor persona cada día.

Agradezco finalmente a todas las demás personas que estuvieron involucradas en la elaboración de este proyecto, a mis asesores, maestros y amigos que de alguna forma aportaron su granito de arena para que esto fuera hoy una realidad.

Resumen Ejecutivo

Esta investigación está orientada a estudiar los problemas de desempleo y subempleo profesional universitario y mediante el uso de una metodología de minería de datos poder estimar las principales causas de ambos problemas laborales y de predecir los índices de desempleo y subempleo profesional en los próximos años.

La minería de datos permite preparar, sondear y explorar los datos para sacar la información oculta en ellos de manera automatizada, y de esta forma buscar comportamientos parecidos en los mismos, describirlos, y orientarlos hacia la obtención de algún beneficio; y por esta razón la investigación se fundamenta en las diferentes aplicaciones que tiene la minería de datos, y la efectividad de los algoritmos utilizados en la misma para poder realizar predicciones de interés en los datos analizados.

Adicionalmente se realiza un estudio de las principales causas del desempleo y subempleo profesional que se presentan en el país, y a través del uso de minería de datos realizar predicciones que permitan a las autoridades universitarias o los distintos entes del gobierno realizar un análisis más profundo para conocer y emprender acciones para reducir los índices de desempleo o subempleo profesional.

Palabras Clave: Minería de datos, Desempleo, Subempleo, Profesional, Algoritmos, Metodología.

Abstract

This research focuses on studying the problems of unemployment and underemployment professional by using a data mining methodology to estimate the main causes of work problems and predict unemployment and underemployment professional in the coming years.

Data mining allows to prepare, probe and explore data to get the information hidden in them in an automated manner, and thus find similar behaviors in them, describe them, and guide them towards obtaining some benefit; and for this reason the research is based on the different applications that have data mining, and the effectiveness of the algorithms used in the same order to make predictions of interest in the data.

Additionally, a study of the main causes of unemployment and underemployment professional that occur in the country, and through the use of data mining make predictions that allow the university authorities or other government authorities make a deep analysis to know is performed and execute actions to reduce unemployment or underemployment professional.

Keywords: Data Mining, Unemployment, Underemployment, Professional, Algorithms, Methodology.

ÍNDICE GENERAL

Índice de Figuras

Índice de Tablas

Índice de Gráficos

INTRODUCCION

Los problemas de desempleo en Honduras no son nuevos. Durante los últimos años los niveles de desempleo y subempleo se han caracterizado por su inestabilidad, así se puede apreciar en la Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) realizada por el INE cada año, según el (INE, 2013) para el año 2013 se contaba con una tasa de desempleo abierto del 3% a diferencia del 2012 donde era del 5%,

Según (Parkin & Esquivel, 2008) El ritmo de creación y destrucción de empleos no es constante y fluctúa durante el ciclo económico; y Honduras no es la excepción, en realidad en países como Honduras según (Lopez, 2012) la economía presenta un eterno letargo en su crecimiento, esto no permite la creación de mejores condiciones de vida y por lo tanto de fuentes de empleo.

En Honduras existen aproximadamente 20 universidades según la Dirección de Educación Superior (Superior, 2007) de donde egresan gran cantidad de profesionales de distintas áreas, muchos de estos profesionales llegan a formar parte de las estadísticas de desempleados en el país. Otra parte de estos profesionales logran encontrar un empleo, pero no ejerciendo su profesión, a lo que se le conoce como subempleo profesional o sobreducación.

El desempleo y subempleo profesional es un problema de carácter personal y social, y por lo cual es importante conocer las causas que lo producen para que se puedan tomar medidas tanto a nivel personal como social. Con el avance de la tecnología; hoy en día existen técnicas que permiten evaluar y estudiar grandes cantidades de datos de temas de interés social como lo es el desempleo y subempleo profesional. En esta investigación se usará la minería de datos, para analizar la problemática del desempleo y subempleo profesional.

La minería de datos es el proceso no trivial que permite identificar a partir de los datos, patrones válidos, novedosos y potencialmente útiles; para tomar decisiones o ejecutar cualquier tipo de acción.

Esta investigación pretende a través de la aplicación de técnicas de minería de datos ser de utilidad para que las autoridades universitarias y gubernamentales puedan contar con la información

necesaria para realizar predicciones y análisis de las medidas que deben tomar para disminuir el índice de profesionales desempleados, o con subempleo profesional en el país.

Para esto la investigación se desarrolla en los siguientes capítulos:

1. Planteamiento del problema: aquí se define más ampliamente el problema a abordar en esta investigación, los objetivos de la investigación que surgen a partir del problema analizado y además se definen las limitantes y la viabilidad del estudio.
2. Marco Conceptual: en este se desarrollan los principales conceptos que serán usados a lo largo de esta investigación.
3. Marco Teórico: en este se desarrolla la temática de minería de datos, desempleo, subempleo y demás temas de interés relacionados con la investigación, con base a la literatura de diferentes fuentes validadas y verificables.
4. Enfoque y tipo de investigación: en este capítulo se detalla el tipo de investigación y el enfoque que se le dará a la misma.
5. Variables: Aquí se definen las variables que serán medidas en la investigación y que permitirán cumplir con los objetivos definidos.
6. Estrategia metodológica: Aquí se detalla la población a la cual estará sujeta el estudio, y la muestra que se utilizara para la investigación, además se desarrolla el instrumento de investigación y se realiza la validación del mismo.
7. Plan de análisis: En este capítulo se define el plan a seguir para realizar el análisis de los datos e información recolectada en el estudio, por medio de las encuestas y datos históricos obtenidos por medio del INE.
8. Metodología de minería de datos: Aquí se define y desarrolla la metodología para poder aplicar la minería de datos en la presente investigación.
9. Análisis de resultados: Aquí se procede a realizar el análisis de los resultados obtenidos a través de las encuestas y la aplicación de la minería de datos.
10. Finalmente, con los resultados obtenidos en el análisis y durante toda la investigación, se desarrollaran las conclusiones de la investigación en general.

CAPITULO I: PLANTEAMIENTO DEL PROBLEMA

1.1 Antecedentes

Los índices de desempleo y subempleo profesional en el país han variado con el paso de los años, según la Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) realizada por el INE cada año (INE, 2013), estos índices se ven afectados por distintos factores, que hacen que estos no sean constantes y bajen o aumenten dependiendo de la situación que atraviese el país.

La educación impartida en las universidades no debe disminuir su calidad, sin importar los problemas sociales que existan en el país, sin embargo esto no es del todo cierto, muchas veces la educación se ve afectada por la crisis que pueda atravesar el país, según (Lopez, 2012) en Honduras el retardo del crecimiento de la economía no ayuda a crear mejores condiciones de vida, al contrario el ingreso anual producido y la riqueza se concentran en pocas bolsas, y esto además de afectar a los más necesitados, afecta a los estudiantes de las universidades y el pueblo en general.

El análisis de las causas de los problemas laborales en el país por tanto no solo se refiere a las causas sociales que afecten al país o a los habitantes, también se refiere a los problemas que afecten a las universidades; especialmente porque existe un índice de desempleo y subempleo conformado por profesionales egresados de la universidad, a pesar de que la mayoría de los habitantes asuman que los egresados de las universidades encuentren un empleo más rápidamente.

Para desarrollar esta investigación se utilizará la minería de datos para estudiar la problemática laboral de los profesionales universitarios en el país. La minería de datos ha sido utilizada en diferentes estudios para realizar predicciones asociadas a la educación y el desempleo y ese han obtenido resultados exitosos.

(Azoumana, 2013) Realizó un análisis de la deserción estudiantil en la Universidad Simón Bolívar, y en general según el autor este tipo de análisis a ayudan a tener un conocimiento general del tema para desarrollar trabajos futuros en otras áreas de conocimiento. (PAUTSCH, 2009) También utilizó la minería de datos para realizar otro estudio de deserción estudiantil, y dados los resultados de su análisis, el autor sugirió a la universidad utilizar la minería de datos en el flujo de control de la universidad. En Argentina (Perversi, 2007) utilizó la minería de datos para explorar y detectar patrones delictivos.

Tomando estos ejemplos, más la literatura que se abordará en esta investigación se pretende aplicar una metodología de minería de datos, que permita estimar el nivel de desempleo y subempleo profesional.

1.2 El Problema de Investigación

En Honduras, la tasa de desempleo es uno de los principales problemas de la sociedad, con especial repercusión en jóvenes y egresados universitarios; según el (INE, 2013) con datos de la Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) existen 141,724 personas desempleadas y el 13.5% de estos corresponden a personas con un nivel educativo de educación superior.

De las personas con empleo, existen 408,875 que corresponden a personas con subempleo visible y 1, 422,210 a personas con subempleo invisible, de los cuales 29,591 y 49,019 son profesionales universitarios con subempleo visible e invisible respectivamente.

Este es un problema de interés tanto para los egresados como para el conjunto de la sociedad hondureña, según (Villamandos, Ocerin, & Castro, 2007) el análisis de este problema permite comprobar si el sistema de educación esta convenientemente adecuado a las demandas del mercado laboral. Contar con una situación en la que mayor parte de los jóvenes saliera del sistema educativo y se encontrará sin dificultad un empleo sería ciertamente tranquilizador, tanto para los jóvenes como para la sociedad en conjunto, sin embargo tal como lo describe (Corrales & Rodiguez, 2003) la realidad muestra que una parte importante de los universitarios, tras salir del sistema educativo, se enfrentan a importantes dificultades para encontrar un empleo y, en muchos casos, acaban desempleados o subempleados, es por esto que surge una de las dudas de esta investigación, cuales son las principales causas del desempleo o subempleo en egresados universitarios en Honduras.

Como se cita anteriormente el problema de desempleo es de interés para toda la sociedad hondureña y por ende para la Universidad Nacional Autónoma de Honduras, para el alma máter, como para todas las universidades, el conocimiento es algo imprescindible para tener éxito, por tanto encontrar asociaciones o correlaciones interesantes en los registros de sus estudiantes o egresados puede

ayudar a la toma de decisiones. Según (Azoumana, 2013) en este entorno la minería de datos ofrece la posibilidad de llevar a cabo un proceso de descubrimiento de información automático.

Según un estudio realizado por (Azoumana, 2013) para brindar una solución acorde a las necesidades de las universidades y empresas es necesario entender los objetivos y requerimientos desde la perspectiva de lo que se busca, convirtiendo entonces este conocimiento en la definición de un problema de minería de datos, ya que dependiendo del problema de información que se desea solucionar, existe una serie de técnicas que son aplicadas en la solución de diversos problemas. De esta situación surge, entonces, esta investigación, en la cual se pretende desarrollar una metodología de minería de datos con la cual se puedan estimar las principales causas de desempleo y subempleo profesional de los egresados universitarios, pudiendo de esta misma forma determinar qué modelos o técnicas de minería de datos se adaptan mejor al análisis a realizar en esta investigación.

1.3 Objetivos de la Investigación

1.3.1 Objetivo General

1. Aplicación de técnicas de minería de datos que permita estimar las principales causas de desempleo y subempleo profesional de egresados universitarios.

1.3.2 Objetivos Específicos

1. Determinar los modelos o técnicas de minería de datos se adaptan mejor al análisis del desempleo y subempleo profesional de egresados universitarios.
2. Determinar los principales factores que influyen en el desempleo y subempleo profesional de egresados universitarios.
3. Estimar el nivel de desempleo y subempleo profesional del año 2014 a través de la aplicación de técnicas de minería de datos.

1.4 Preguntas de Investigación

- ¿Qué modelos o técnicas de minería de datos se adaptan mejor al análisis del desempleo y subempleo profesional de egresados universitarios?
- ¿Qué factores influyen en el desempleo de profesionales universitarios?
- ¿Qué factores influyen en el subempleo profesional de egresados universitarios?

1.5 Justificación del Estudio

Desde hace muchos años el desempleo es una realidad latente en Honduras, según el informe del instituto nacional de estadística (INE, 2013), la tasa de desempleo abierto a nivel nacional se estima en un 3.9% de la Población Económicamente Activa (PEA). Aunque este porcentaje parezca pequeño es importante resaltar que el 79.2% de la población nacional está en edad de trabajar, sin embargo, la Población Económicamente Activa, apenas representa el 42.5%, en donde el 3.9% se encuentra desempleada.

Este 3.9% de desempleo no solo está formado por personas con un bajo nivel educativo, según el (INE, 2013) el 13.5% de la población desempleada son profesionales a nivel universitario.

Son distintas las razones por la cual se presenta este porcentaje, entre ellas la problemática que se da en el sistema de educación en el país, los problemas económicos del país, la saturación de profesionales en las distintas carreras, entre otros.

Estas estadísticas afectan a las universidades ya que no todos sus egresados está consiguiendo empleo, es por esta razón que las autoridades deben estar preparadas y tomar las medidas necesarias para solventar las posibles causas que generan estas estadísticas como ser la saturación de profesionales en las distintas carreras y el nivel de educación proporcionado a estos estudiantes que no cumple la demanda del mercado laboral.

Del porcentaje de personas que se encuentran trabajando según estadísticas realizadas por el INE en el 2012, el 40.8% de estas se encuentran en la clasificación de empleos invisibles. Es decir que

tienen ingresos inferiores a un salario mínimo y probablemente varios de ellos cuentan con un título universitario.

Debido a esta problemática laboral en el país, muchos de los profesionales que se gradúan de las universidades se encuentran trabajando, pero no ejerciendo su profesión. Esta investigación pretende realizar un análisis de las principales causas del subempleo profesional y del desempleo de profesionales universitarios.

Para lo anterior se utilizará la minería de datos; ya que esta prepara, sondea y explora los datos para obtener la información oculta en ellos de manera automatizada, y de esta forma buscar comportamientos parecidos en los mismos, describirlos, y orientarlos hacia la obtención de algún beneficio.

En esta investigación el beneficio será proporcionar a las autoridades información para poder estudiar las diferentes causas que producen el desempleo y subempleo profesional, para que puedan emprender acciones para disminuir este tipo de problemas en la educación superior y en el país.

1.6 Delimitación del Problema

La problemática del desempleo y subempleo afecta a todos los niveles educativos, clases sociales, diferentes carreras en todas las universidades el país. Es por esto que esta investigación está orientada a desarrollarse en profesionales a nivel universitario. Así mismo la investigación se desarrolla en la ciudad de Tegucigalpa M.D.C., dentro de esta limitante geográfica es importante destacar que no se puede tener acceso a todos los profesionales en la ciudad, debido a que no se conoce la ubicación de los mismos, y no todos estuvieron dispuestos a colaborar con la investigación.

1.7 Posibles Deficiencias en el Proceso de Investigación

Una de las deficiencias más grandes en la investigación es el recurso del tiempo debido a esto la investigación se realizará de manera general en todos los egresados y no específicamente en la

UNAH, ya que no se cuenta con información de donde se encuentran los egresados, ni el tiempo para localizarlos a todos para poder obtener la información.

Otra limitante es la cantidad de datos que se obtengan de las encuestas aplicadas y del histórico proporcionado por el INE, ya que la minería de datos necesita una gran cantidad de información para poder aplicar algoritmos de manera eficiente y que las predicciones sean más confiables.

El INE únicamente proporciono datos estadísticos de los últimos 6 años y las encuestas aplicadas a profesionales universitarios fueron menos de 500, ya que no se tiene acceso a todas las personas que presentan problemas laborales.

1.8 Viabilidad del Estudio

A pesar de la limitante de datos que se definió en la sección anterior, y a partir de la información que se obtenga; se aplicará la metodología de minería de datos para poder alcanzar los objetivos de esta investigación.

Para aplicar las técnicas de minería de datos se utilizara la herramienta Analysis Services, la cual es nativa del gestor de base de datos SQL Server, por lo cual no representara ningún costo, ya que se tiene acceso a este gestor. Otra de las ventajas de utilizar Analysis Services es que posee distintos tipos de algoritmos, que pueden aplicarse en una sola estructura de minería de datos y realizar comparaciones entre ellos, lo cual permitirá elegir el algoritmo más adecuado para esta investigación.

Además de contar con la herramienta de Analysis Services, se tiene acceso a información teórica, experiencia y el apoyo necesario para poder aplicar la minería de datos y poder cumplir con los objetivos definidos en esta investigación, adicionalmente a esto, se dejaran establecidos los pasos y consideraciones necesarias para poder aplicar la metodología de minería de datos con una mayor cantidad de datos.

Los costos monetarios en esta investigación son mínimos, ya que solo serán gastos en papelería, transporte y la información proporcionada por el INE, esto sumado con las facilidades citadas anteriormente, permite que esta investigación sea viable de desarrollar.

CAPITULO II MARCO CONCEPTUAL

2.1 Minería de Datos

2.1.1 Base de datos

Según (Lapuente, 2013) "Una base de datos es una colección de datos organizados y estructurados según un determinado modelo de información que refleja no sólo los datos en sí mismos, sino también las relaciones que existen entre ellos."

Sistema Gestor de Base de Datos

Para (Garzon Perez, 2010) una base de datos es una colección de programas que permiten a los usuarios crear y mantener una base de datos, facilita los procesos de definición, construcción y manipulación de la base de datos para distintas aplicaciones.

Sistemas OLTP

Una base de datos OLTP es un sistema computarizado que, en general, tiene la finalidad de almacenar información para luego permitir a los usuarios consultar, eliminar y actualizar sus datos. De esta manera ayudan a las organizaciones en el proceso de administración, haciendo que sus operaciones, al menos las más importantes, queden registradas. Estas funciones integran un tipo de proceso denominado OLTP (On Line Transaction Processing o Procesamiento de Transacciones en Línea) (Date, 2001); que son las bases de datos que se mencionaron anteriormente.

2.1.2 Data WareHouse

Según (PAUTSCH, 2009) Data WareHouse es una combinación de hardware de Alto rendimiento y capacidad de almacenamiento que combinado con software especializado, consolida, integra y analiza datos provenientes de distintas fuentes, con el objetivo de apoyar y mejorar la toma de decisiones de los administradores en los niveles estratégicos de las empresas u organizaciones.

Para (W.H., 1992) "Data Warehouse es un conjunto de datos integrado orientados a una materia, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración"

2.1.3 ETL o proceso de consolidación.

Para (PAUTSCH, 2009) ETL es el proceso que se encarga de depurar, clasificar y transformar los datos del OLTP para poder alimentar al DWH. Consolida, sumariza, disgrega y transforma los datos de las aplicaciones que no están integradas.

2.1.4 Sistemas OLAP

OLAP (On Line Analytical Process – Procesamiento Analítico En Línea) es un sistema que trabaja sobre el DWH.

Para (PAUTSCH, 2009) "Son aplicaciones que generan información táctica y estratégica que sirven a la organización como soporte para la toma de decisiones."

Inteligencia de Negocios

2.1.5 Inteligencia de Negocios

Inteligencia de negocios o Business Intelligence (BI) según (Gartner, 2006) "BI es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un datawarehouse), para derivar ideas y extraer conclusiones. El proceso de Business Intelligence incluye la comunicación de los descubrimientos y efectuar los cambios. Las áreas incluyen clientes, proveedores, productos, servicios y competidores."

2.1.6 Minería de Datos

La minería de datos (en Inglés: data mining DM) es un proceso que consigue conocimiento partiendo de un conjunto amplio de datos, a los cuales se le aplica métodos para obtener patrones o tendencia nuevas, generando nuevos conocimientos. (Osorio, 2011) Definió minería de datos como: "un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos"

2.1.7 Descubrimiento de Conocimiento en Bases de Datos (KDD)

Para (Cibertec, 2010) KDD (Knowledge Discovery from Databases) es un proceso que busca extraer información implícita no trivial de las bases de datos, que no era conocida y que sea de

utilidad. Para lograrlo se procesa la información con algoritmos neuronales, árboles de decisión, entre otros.

2.1.8 Algoritmos de Minería de Datos

Según (Microsoft, 2014) un algoritmo de minería de datos es un conjunto de cálculos y reglas heurísticas que permite crear un modelo de minería de datos a partir de los datos. Para crear un modelo, el algoritmo analiza primero los datos proporcionados, en busca de tipos específicos de patrones o tendencias. El algoritmo usa los resultados de este análisis para definir los parámetros óptimos para la creación del modelo de minería de datos. A continuación, estos parámetros se aplican en todo el conjunto de datos para extraer patrones procesables y estadísticas detalladas.

2.2 Problemas Laborales

2.2.1 Desempleo

Desempleo es la situación del trabajador que carece de empleo y, por tanto, de salario. Por extensión es la parte de la población que estando en edad, condiciones y disposición de trabajar (población activa) carece de un puesto de trabajo (Cfr. Samuelson, 2006)

2.2.2 Subempleo

Según (Ramírez Rojas & Guevara Fletcher, 2006) el subempleo se define como "una categoría del mercado de trabajo según la cual, la ocupación que tienen un conjunto de trabajadores es inadecuada respecto a determinadas normas o a otra ocupación posible."

2.2.2.1 Subempleo Visible

Según él (INE, 2013), el subempleo visible comprende a aquellas personas que estando ocupados, trabajan menos de 36 horas a la semana y desean trabajar más.

2.2.2.1 Subempleo Invisible

El subempleo invisible según él (INE, 2013) se define como las personas que trabajan más de 36 horas y tuvieron ingresos inferiores a un salario mínimo.

2.2.3 Subempleo Profesional o Sobreducación

En una investigación realizada por (Moreno & Burga, 2011) define las personas sobreducadas como "aquellos que tienen más educación que la requerida para desempeñar adecuadamente su trabajo (empleo), por lo que sus habilidades no estarían siendo plenamente utilizadas."

"Este cálculo se hace independientemente del salario que reciba por realizar dicha actividad. De manera análoga, un individuo subeducado es aquel que tiene menos educación que la requerida para desempeñar adecuadamente su trabajo."

El (Diccionario de la Lengua Española, 2012) define subemplear como "emplear a alguien en un cargo o puesto inferior al que su capacidad le permitiría desempeñar".

De la misma forma (Moreno & Burga, 2011) se refiere a las personas subempleadas como a los profesionales que desempeña una ocupación que no tiene relación con los estudios que cursaron.

(Gobernado Arribas, 2007) Define subempleo profesional como "el hecho de poseer un nivel educativo que excede las necesidades del puesto de trabajo"

El subempleo profesional generalizado es aquel que alcanza a la totalidad (o a la gran mayoría) de la población. En la medida que la expansión educativa alcanza a más población (expansión horizontal) y que cada vez se estudia durante más años (expansión vertical).

El subempleo profesional relativo es aquel que afecta solamente a una minoría de cada categoría ocupacional: unos pocos tienen un nivel educativo superior al resto de la categoría en cuestión.

CAPITULO III MARCO TEORICO

3.1 Marco Teórico

3.1.1 Bases de Datos

3.1.1.1 Datos e Información

Los datos son todos aquellos símbolos, cifras o instrucciones que por sí solos no significan nada, además se encuentran aislados entre si y sin ningún tipo de orden.

Una vez que los datos pueden ser ordenados y organizados, dejan de ser únicamente símbolos y pasan a convertirse en información, y esta puede ser leída, interpretada y utilizada para analizar y tomar decisiones.

Una vez clara la diferencia entre datos e información, se puede definir lo que es una base de datos. Una base de datos almacena datos, pero a través de un modelo determinado permite convertir estos datos en información, por lo tanto la base de datos puede almacenar tanto datos como información.

3.1.1.2 Características

Como se citó anteriormente una base de datos permite almacenar datos e información, y es diseñada con un propósito específico de tal manera que la información almacenada puede ser compartida por diferentes usuarios y aplicaciones. Es importante además que estos datos se conserven de forma segura e integra, de manera que la información proporcionada sea confiable.

Una base de datos utiliza un conjunto de tablas para representar los datos y las relaciones entre ellos, de esta forma se tiene una tabla llamada Empleados, que contenga el nombre, apellido, identidad y otros datos de los empleados. Si únicamente se cuenta con un nombre, no significa nada; sin embargo al asociar el apellido, identidad y departamento de la empresa en la que ese empleado labora, se tiene una información más detallada y que puede ser utilizada por la empresa.

Las principales características que una base de datos debe tener según (Garzon Perez, 2010) son las siguientes:

- Integridad de los datos: Se refiere a que los datos almacenados deben ser coherentes entre sí, no deben ser alterados con cualquier tipo de contenido y si se presentan a dos usuarios distintos deben ser los mismos para ambos.

- No redundancia de los datos: La redundancia hace referencia a no almacenar más de una vez el mismo dato o conjunto de datos.
- Restricciones de seguridad y confidencialidad: Los datos deben ser almacenados de forma segura, y únicamente los usuarios con los permisos necesarios deben poder acceder a los mismos.

Un sistema gestor de bases de datos es el encargado de administrar y gestionar las bases de datos y almacenar el acceso a los mismos. La mayoría de los gestores de bases de datos relacionales emplean el lenguaje SQL (Structure Query Lenguaje - Lenguaje de Consulta Estructurado). Según (Silberschartz & Sudarshan, 2005) este lenguaje se divide en dos ramas:

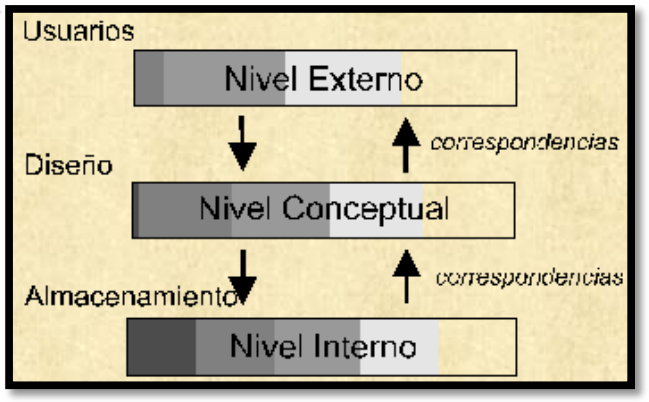
- DDL (Lenguaje de Definición de Datos).
- DML (Lenguaje de Manipulación de Datos).

En una base de datos, las entidades y atributos del mundo real, se convierten en registros y campos en una tabla. Estas entidades pueden ser tanto objetos materiales como libros o fotografías, pero también personas e, incluso, conceptos e ideas abstractas. Las entidades poseen atributos y mantienen relaciones entre ellas.

3.1.1.3 Arquitectura

Según (Lapuente, 2013) existen 3 niveles para definir la arquitectura de un sistema de bases de datos: el nivel externo, conceptual e interno; estos se muestran en la siguiente figura:

Figura 1: Niveles de la arquitectura de un sistema de base de datos.



Fuente (Lapuente, 2013)

- Nivel Interno o Físico

  Es el nivel más bajo de la arquitectura y el nivel real de los datos almacenados. Según (Lapuente, 2013) este nivel define cómo se almacenan los datos en el soporte físico, ya sea en registros o de cualquier otra forma, así como los métodos de acceso. Este nivel a su vez lleva asociada una representación de los datos, que es lo que se denomina Esquema Físico.

- Nivel conceptual

  Es el nivel intermedio y como se muestra en la figura anterior corresponde al diseño de la base de datos, es decir que es como se muestra la misma en el mundo real. Este nivel es el diseño desarrollado por la empresa y que recoge todos los datos de los diferentes usuarios y aplicaciones. (Lapuente, 2013) Afirma que este nivel se trata con la entidad u objeto representado, sin importar como está representado o almacenado éste y que a su vez lleva asociado un Esquema Conceptual.

- Nivel Externo o de visión

  Este nivel es la parte a la que los usuarios de la base de datos tienen acceso, ya que los mismos no pueden acceder a todo el nivel conceptual. Un ejemplo de esto como menciona (Lapuente, 2013) sería el caso del empleado de una organización que tiene acceso a la visión de su nómina, pero no a la de sus compañeros. El esquema asociado a éste nivel es el Esquema de Visión.

3.1.2 Sistemas OLTP

Los sistemas OLTP son básicamente aquellos que almacenan la información que los usuarios pueden consultar, modificar o eliminar, por esta razón su diseño está orientado al procesamiento de transacciones y pueden almacenar grandes volúmenes de datos. Sin embargo esta información no necesariamente es utilizada para la toma de grandes decisiones.

3.1.2.2 Inconvenientes

En la actualidad la mayoría de las empresas compiten entre sí y por eso intentan sacar la mayor utilidad de la información disponible en sus bases de datos. En las bases de datos OLTP las empresas cuentan con un gran volumen de información; el problema radica en que las empresas no pueden garantizar la buena calidad de los datos almacenados, ya que los datos son procesados por los usuarios y pueden contener errores; y por lo tanto esta información no puede utilizarse para la toma de las decisiones.

Los mayores inconvenientes que se presentan en las bases de datos OLTP según (PAUTSCH, 2009) son los siguientes:

- Por lo general no existe una BD única donde se pueda consultar sobre los distintos temas, sino que se cuenta con varios sistemas independientes en las distintas áreas (por ejemplo: sistema de facturación, contaduría, stock).
- Datos que se representan con distinto formato o codificación (por ejemplo: el estado civil, el sexo o fecha).
- Datos históricos que a menudo son resguardados, comprimidos y sacados del sistema OLTP.
- En las bases de datos OLTP no existe un detalle apropiado que permita a los administradores de las organizaciones tomar las decisiones adecuadas.

3.1.3 Data Warehouse

Un Data Warehouse (DWH) es una base de datos corporativa que replica los datos de una base de datos una vez que estos han sido seleccionados, depurados y convertidos en forma de objetos para realizar actividades como la construcción de reportes o consultas.

Es decir que un DWH es una base de datos con gran capacidad de almacenamiento que permite a los directivos tener datos históricos necesarios para la toma de decisiones.

Según (PAUTSCH, 2009) el objetivo de un DWH es convertir datos en información. En ese proceso de conversión, se toman datos provenientes de distintas fuentes (incluidas en estas los sistemas OLTP), se los consolida y almacenan en un DWH.

Ya que el DWH presenta la información ya procesada y de forma consolidada, el mismo debe tener por lo tanto un entorno amigable para el usuario, debe ser fácil de utilizar y permitir exportar e imprimir datos del sistema.

3.1.3.2 Características

Entre las principales características de una Data WareHouse citadas por (PAUTSCH, 2009) se pueden mencionar las siguientes:

- Un sistema Integrado

  Un Data WareHouse permite integrar datos de diferentes fuentes, ya sean internas o externas a la organización; estos datos a su vez deben ser almacenados de forma coherente y consistente; por ejemplo si se utilizan nomenclaturas para definir tipos en la información, los datos deben procesar de todas las fuentes considerando la adaptación de los mismos a la misma nomenclatura.

- Orientado

  De manera de consultar de forma eficiente la información en las bases de datos de Data WareHouse, esta información debe estar orientada como menciona (PAUTSCH, 2009) a consultar, es decir que debe contener toda la información y no únicamente la del área de la empresa que la necesita.

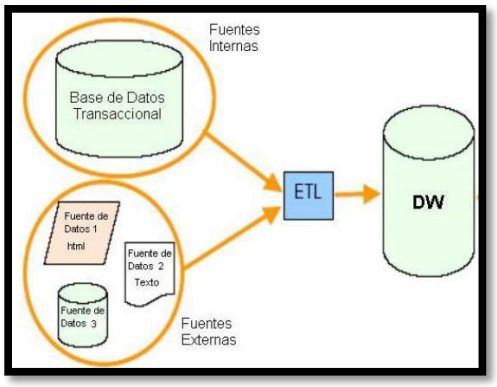
- Sistema no volátil

  Según (PAUTSCH, 2009) en un Data wareHouse los datos siempre son agregados y nunca removidos, tampoco deben ser actualizados. Esto permite analizar las diferentes áreas y a su vez ver el historial de la empresa con el paso del tiempo.

3.1.3.3 Arquitectura

Un DWH según (W.H., 1992) está formado por la arquitectura que se puede apreciar en la siguiente figura, consta de una base de datos transaccional o OLTP, fuentes externas y un proceso de transformación y depuración de los datos (ETL)

Figura 2: Arquitectura de un DWH.



Fuente: (W.H., 1992)

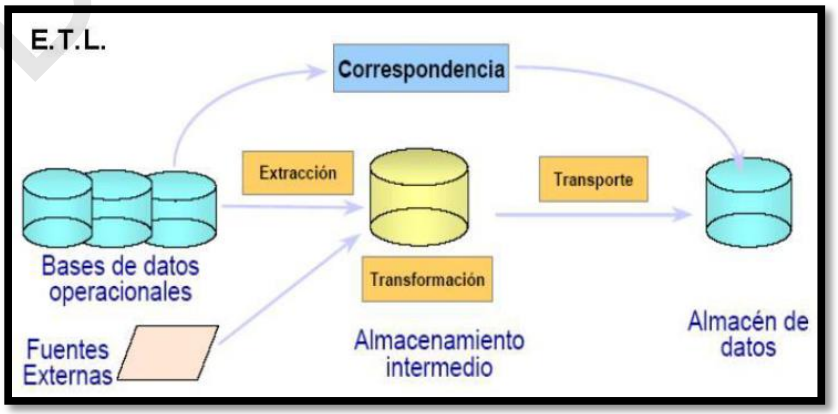
Como se observa en la figura un DWH es alimentado por medio de las bases de datos OLTP y fuentes externas; ya sean bases de datos, archivos de texto, xml u otros. Todos estos datos sin embargo no son almacenados directamente a la base de datos; si no que pasan por un proceso de depuración o transformación llamado ETL, que se define a continuación.

3.1.3.4 ETL o Proceso de Consolidación.

ETL por sus siglas en ingles Extract, Transform, Load, es el proceso que se encarga de extraer los datos de diferentes fuentes, luego los depura y transforma a datos válidos y coherentes para finalmente cargarlos a la base de datos final.

En la siguiente figura se muestra cual es el proceso que se realiza en el ETL:

Figura 3: Extracción, Transformación y Transporte de datos.

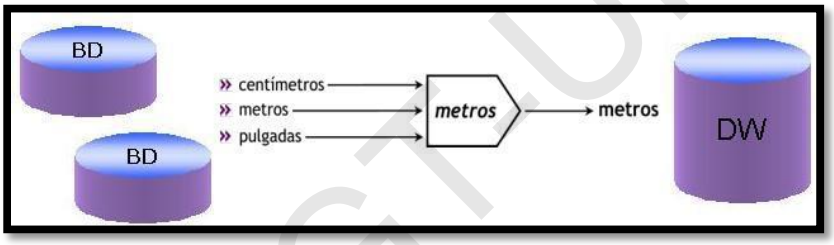


Fuente (PAUTSCH, 2009)

- El primer paso es la extracción de las diferentes fuentes externas o bases de datos operacionales, estos datos se trasladan a una base de datos intermedia.

- El segundo paso es la transformación, este se realiza en la base de datos intermedia, aquí es donde los datos son depurados en caso de que existan inconsistencias, o tipos de datos inválidos; además se realiza la transformación de los mismos a las nomenclaturas o tipos necesarios, tal como se muestra en la figura siguiente, en ese caso se obtienen datos de diversas bases de datos, ya sean centímetros, pulgadas o metros; todos se transforman finalmente a metros para ser almacenados en la base de datos final.

Figura 4: Transformación de los Datos.



Fuente (PAUTSCH, 2009)

- El tercer y último paso es la carga o transporte, una vez que los datos son transformados los mismos son cargados al almacén de datos o Data wareHouse a través de un proceso Batch. La periodicidad con la que este proceso se lleva a cabo depende de la empresa y las necesidades de la misma.

3.1.4 Sistemas OLAP

Los sistemas o bases de datos OLAP son aquellas que permiten consultar grandes volúmenes de datos de una forma rápida, un sistema OLAP trabaja sobre un Data warehouse para proporcionar a la empresa la información que la misma necesite para la toma de decisiones.

Algo interesante en este tipo de bases de datos o sistemas es que no se necesita conocer la estructura del DWH, lo que se hace es presentar al usuario una interfaz en la que él pueda realizar consultas al seleccionar los diferentes atributos; para esto se utilizan lo que son los cubos, estos son

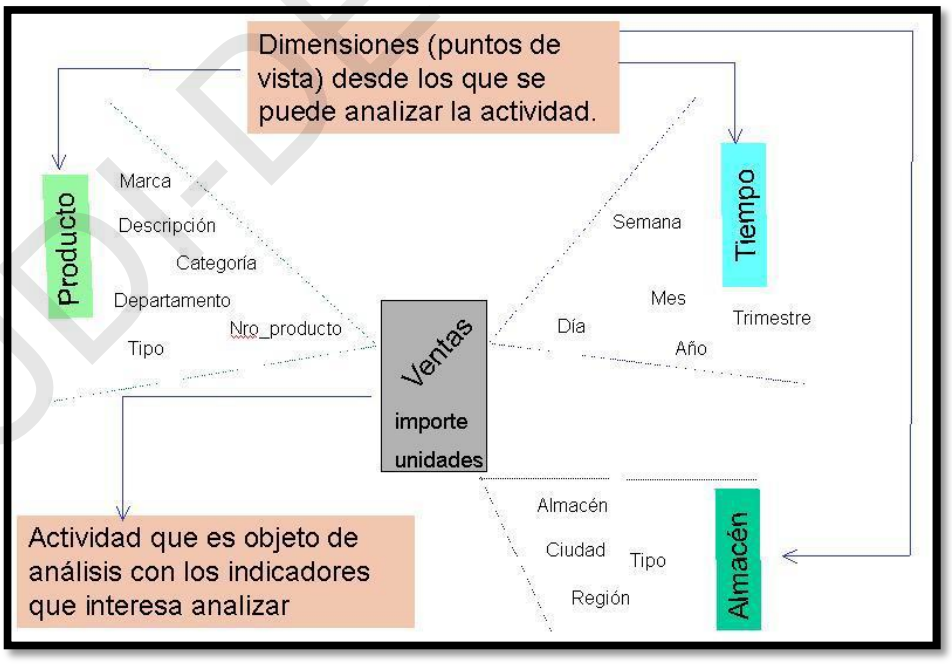
estructuras multidimensionales que permiten presentar y relacionar información de diferentes tablas del DWH, es así que el usuario selecciona atributos de los cubos e internamente el sistema OLAP realiza la consulta al gestor de base de datos.

Según (PAUTSCH, 2009) la estructura multidimensional o cubo consta de una tabla de sucesos o hechos, cuyos atributos describen la actividad que es el objeto del análisis y varias tablas llamadas dimensiones. Los atributos de cada dimensión tienen el objetivo de aportar información particular sobre cada tupla de la tabla de hechos.

Por ejemplo en la siguiente figura, se puede apreciar la tabla de hechos ventas, las dimensiones que son Tiempo, Producto y Almacén y desde los cuales se puede analizar la información, y los atributos de cada una de las dimensiones, por ejemplo en el caso de los productos se tienen marcas, categorías y tipos.

Si observamos todo el cubo, el usuario podría realizar consultas de cuál es el producto más vendido en el mes X en la ciudad Y; esto combinando los atributos producto, mes, ciudad de las tres dimensiones.

Figura 5: Estructura Multidimensional.



Fuente (PAUTSCH, 2009)

3.1.4.2 Diferencias entre OLTP y OLAP

Para hacer un resumen de lo citado en los sistemas OLTP y OLAP y poder comprender mejor los mismos, se establecen las principales diferencias entre ambos sistemas, según (PAUTSCH, 2009).

Como se observa en la tabla, los datos en un sistema OLTP están en constante movimiento porque son los datos que la organización esta diariamente actualizando, por lo cual también son un menor volumen de datos, ya que no necesariamente se mantienen históricos del mismo, por esta razón los tiempos de respuesta de las consultas son más bajos.

En el caso de los sistemas OLAP los datos son más históricos, ya que se almacenan todos los datos que pudieron existir en la base de datos OLTP, por esta razón se manejan grandes volúmenes de datos y los tiempos de respuesta no son tan rápidos, pero no necesariamente lentos.

En la siguiente tabla se muestra a mayor detalle las diferencias entre ambos sistemas.

Tabla 1: Diferencias entre OLTP y OLAP.

Sistemas OLTP	Sistemas OLAP
Datos actuales.	Datos actuales más históricos.
Volumen de datos acotados.	Gran Volumen de datos.
Actualización continua.	Estable.
Actualización On Line.	Actualización Batch.
Atomizado.	Sumarizado.
Un registro a la vez.	Varios registros a la vez.
Orientado a la Aplicación.	Orientado al sujeto.
Orientado a la Operación.	Orientado a la Información Estratégica.
Muchos usuarios concurrentes.	Pocos usuarios concurrentes.
Consultas predefinidas y actualizables.	Consultas complejas no anticipadas.
Requerimientos de respuesta inmediata.	Los tiempos de respuesta no son críticos.
Datos con relaciones.	Datos Multidimensionales.
Baja Redundancia.	Alta Redundancia.

Fuente (PAUTSCH, 2009)

3.1.5 Inteligencia de Negocios

Antes de definir lo que es la minería de datos, es importante mencionar que este concepto procede de la inteligencia de negocios, en inglés: business Intelligence (BI). La inteligencia de negocios es un conjunto de aplicaciones o tecnologías que tienen como principal objetivo dar soporte a la toma de decisiones en las empresas. Según (Microsoft, 2010) la Inteligencia de Negocios o Business Intelligence (BI) es un conjunto de herramientas enfocadas a la administración mediante la entrega de información precisa y útil, en un plazo de tiempo óptimo para apoyar una toma de decisiones eficiente

En la actualidad se vive en una sociedad de la información, se encuentran datos e información en cualquier lugar y gracias al desarrollo de los sistemas de información y el uso del Internet los directivos de las empresas pueden acceder a mucha más información, de más calidad y con mayor rapidez que hace unos años. Esto ofrece un gran potencial para guiar a las empresas hacia la consecución de sus objetivos y para mejorar la toma de decisiones.
Según (Chavez & Bavera, 2013) esto se consolido a partir del año 2000, ya que se logró que muchas aplicaciones se unificaran en unas pocas plataformas y es a partir de esto que las herramientas de BI empezaron a dar soluciones reales a las empresas, tal como lo tenemos hoy en día.

La minería de datos nace a partir del BI, ya que las empresas cuentan con grandes volúmenes de información y la misma se encuentra en diferentes formas, con toda esta información los directivos se enfrentan al problema de que cada día la misma va aumentando pero no tienen el tiempo para analizarla, Business Intelligence pretende convertir y procesar toda esta información y a partir de la misma descubrir conocimiento.

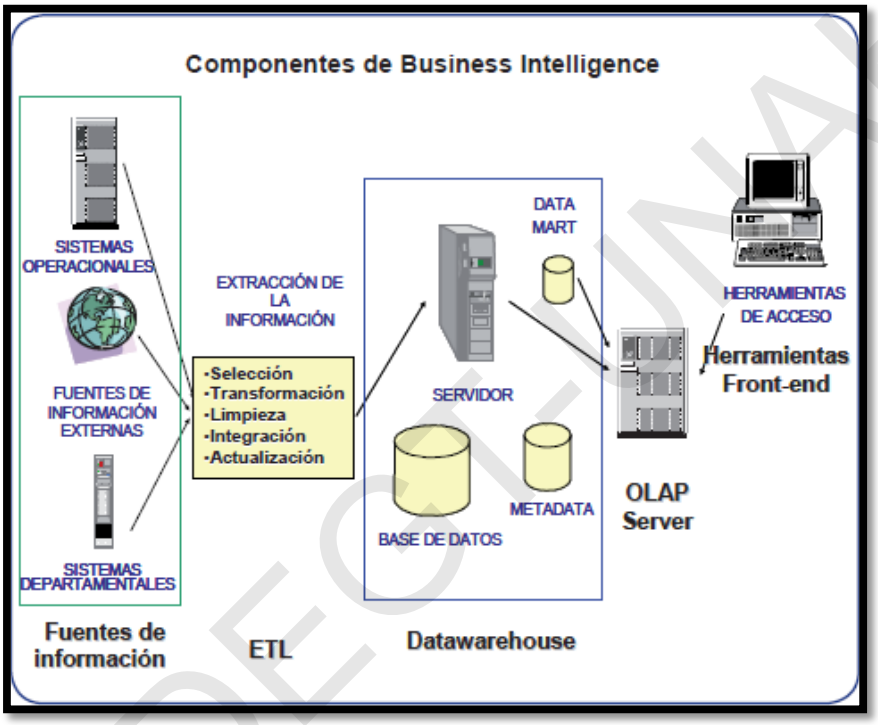
3.1.5.1 Componentes de la Inteligencia de Negocios

El proceso de inteligencia de negocios está compuesto por varios componentes, que ya se han citado con anterioridad, y que se muestran en la siguiente figura:

- Fuentes de información: que pueden ser internas o externas a la organización.
- Proceso de ETL: que se encarga de limpiar, transformar, integrar y actualizar los datos.
- DatawareHouse: que contiene la información ya procesada y depurada.

- OLAP: que permite consultar los grandes volúmenes de datos almacenados en el Data WareHouse.
- Herramientas Front-end: aquellas herramientas que permiten consultar la base de datos OLAP.

Figura 6: Componentes de Business Intelligence



Fuente (Chavez & Bavera, 2013)

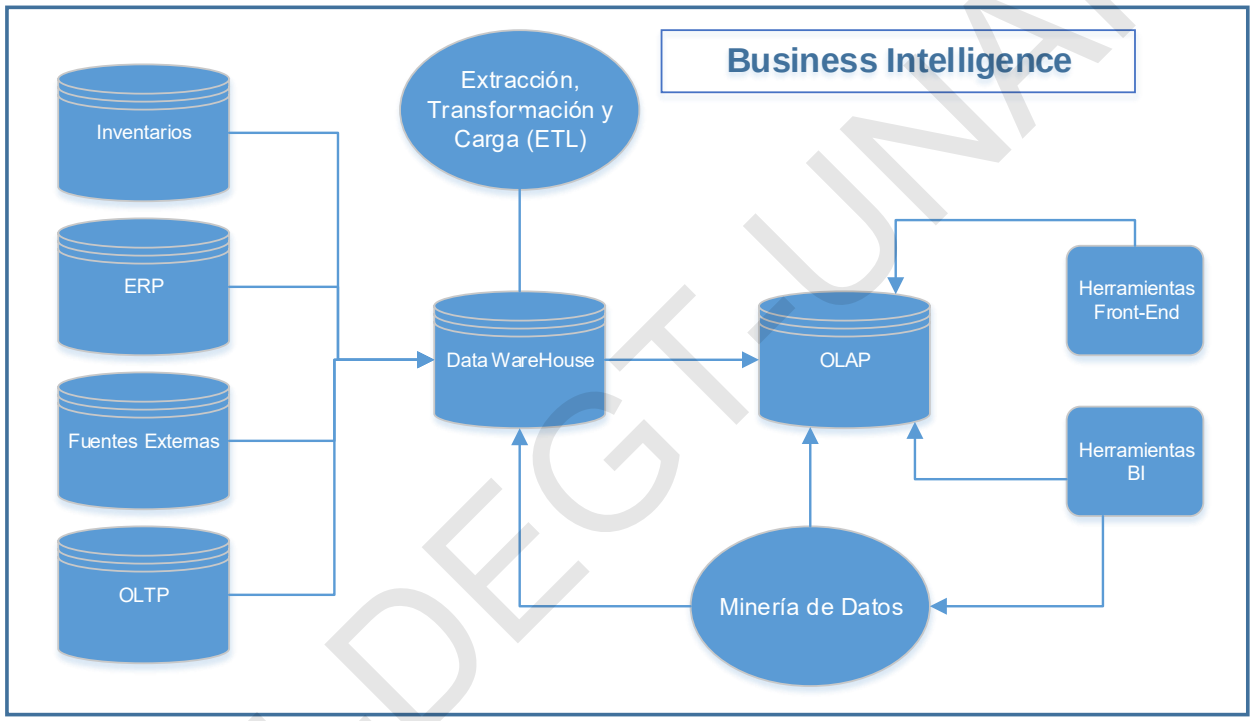
3.1.6 Minería de Datos

La minería de datos es un proceso por medio del cual se obtiene conocimiento a partir de grandes conjuntos de datos almacenados de distintas formas y que antes eran desconocidos, normalmente desde un sistema OLAP. A partir del análisis de esta información se obtienen patrones entre los datos, e incluso patrones que podrían estar ocultos. La minería de datos además permite realizar predicciones basados en los patrones encontrados entre un conjunto de datos histórico.

La minería de datos puede proporcionar según (Chavez & Bavera, 2013) un pronóstico de comportamiento a futuro, esto a través de los patrones encontrados en los datos, para esto existen algoritmos que se ajustan a las necesidades y la información histórica almacenada.

Antes de continuar analizando la minería de datos, se hará un resumen de lo citado hasta el momento y se establecerá una relación entre los temas estudiados, como se muestra en la siguiente figura

Figura 7: Relación entre componentes de Business Intelligence



Fuente de elaboración: Propia.

Como se puede apreciar en la figura anterior, todo lo estudiado como OLTP, OLAP, Data WareHouse son parte del proceso de Inteligencia de negocios o Business Intelligence, como ya se había citado anteriormente.

Pero ¿de qué forma se integra la minería de datos en este diagrama?, la minería de datos y las herramientas o algoritmos utilizados en la misma, son aplicados a la data o información contenida en los sistemas o bases de datos OLAP, o bien directamente sobre el Data warehouse, considerando que al aplicar los algoritmos de minería de datos sobre el data warehouse, el trabajo a realizar es mayor y las probabilidad de falla también lo son; esto se debe a que los datos no se

encuentran completamente procesados, ni se han creado las vistas necesarias o cubos para procesar la información según las necesidades.

Finalmente sobre la minería de datos se aplican lo que son herramientas BI, para mostrar la información ya procesada o el resultado de la aplicación de la minería a los usuarios finales.

3.1.6.1 Origen de la minería de datos.

La minería de datos tiene sus orígenes desde los años 60, en donde ya se hablaba como cita (PAUTSCH, 2009) de términos como Data Mining o Data Archaeology, los cuales pretendían encontrar patrones o relaciones en la base de datos sin una hipótesis previa.

Los algoritmos utilizados en la minería de datos tales como redes neuronales, arboles de decisión y la regresión son utilizados desde los años 40 y 60, sin embargo fue hasta los 80 cuando se empezaron a definir de forma más clara los conceptos de Minería de Datos, y en el área de tecnologías de la información, algoritmia y desarrollo de software la minería de datos se ha insertado desde los años 90 según (Flores & Julca, 2010).

El volumen y variedad de información ha crecido espectacularmente en la última década y gran parte de esta información es histórica, y es aquí que nace la minería de datos ya que esta información almacenada además de ser "memoria de la organización" es útil para predecir la información futura.

Para (Orallo, 2010) la minería de datos tiene sus orígenes en tres ámbitos o conceptos importantes:

- Estadística Clásica: Ya que a través de la estadística se aplican lo que son las herramientas o técnicas de minería de datos, como el análisis de regresión, desviación estándar, varianza, análisis de clustering, intervalos de confianza, entre otros.

- Inteligencia Artificial: La inteligencia artificial se sumó a la minería de datos para aumentar el poder de procesamiento y en conjunto con la estadística dar mejores resultados.

- Aprendizaje Automático: A través del uso de patrones se implementan las técnicas de aprendizaje automático para poder interpretar y aprender de grandes volumes de datos.

3.1.6.2 Características de la minería de datos.

La minería de datos es un proceso de mucha utilidad en las empresas y como es de saber el éxito de muchos negocios depende de la habilidad de ver nuevas tendencias, y para esto la minería de datos es de mucha utilidad, si las empresas las utilizan en sus procesos, la toma de decisiones debería der mucho más fácil.

Una de las principales características de la minería de datos es esta, la facilidad y ventajas que se le proporcionan a la empresa, la toma de decisiones sobre grandes volúmenes de datos y de fuentes muy diversas se hace mucho más fácil, con el uso de minería de datos, sobre todo si no se cuenta con usuarios expertos en estadística, y además se produce un ahorro en el tiempo de procesamiento de todos estos datos.

Otra característica importante que define (Vallejos, 2006) es que la minería de datos permite explorar los datos que se encuentran en las profundidades de los almacenes de datos, y que pueden contener información de varios años.

La minería de datos utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente como cita (Vallejos, 2006) estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos.

Otra característica muy importante es el uso de algoritmos para encontrar patrones y realizar predicciones, para esto debe generarse lo que es un modelo de minería de datos, el cual como menciona (Vallejos, 2006)s incluye desde la formulación de preguntas acerca de los datos, la creación de un modelo para responder a estas preguntas y finalmente la implementación del modelo.

3.1.6.3 Áreas de Aplicación.

Como ya sabemos la minería de datos nos permite detectar tendencias o patrones que existen en los datos, y permite sacar una ventaja competitiva en la empresa al tomar decisiones con base a estas tendencias, pero la minería de datos no es usada únicamente para encontrar o realizar predicciones de ventas, puede utilizarse en diferentes áreas y no únicamente en ventas. Por ejemplo se pueden utilizar los patrones para encontrar nuevos clientes, o definir una tendencia de consumo, lo cual nos permite aplicar minería de datos en áreas como bancos, seguros, transporte.

En educación puede utilizarse para predecir la matricula del siguiente año, la matricula por carrera, la deserción estudiantil, los índices de reprobación, entre otros. Por ejemplo en las universidades puede servir para determinar la oferta o carga académica dependiendo del movimiento de los alumnos, las clases aprobadas y reprobadas.

Entre otras cosas la minería de datos puede ser utilizada para identificar patrones, realizar asociaciones, predecir respuestas, analizar resultados, identificar reglas, determinar gastos, determinar consumos, realizar planificaciones, asociaciones y estimaciones. Por lo tanto la minería puede aplicarse en áreas como ventas, comercio, marketing, educación, transporte, medicina y diferentes áreas en donde las empresas almacenen la información histórica de sus procesos.

La cantidad de aplicaciones que podemos darle a la minería de datos es inmensa, y solo necesitamos información, que en la actualidad en la que vivimos esta por todos lados, solo debemos saber utilizarla y aplicando estas técnicas obtener los mayores beneficios.

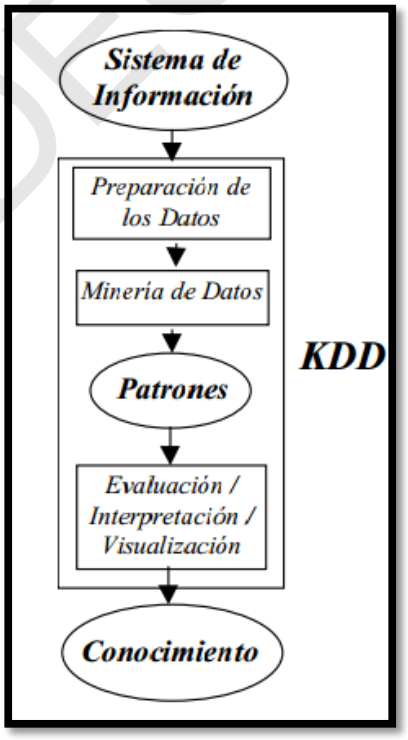
3.1.6.4 Descubrimiento de Conocimiento en Bases de Datos (KDD)

La minería de datos se encuentra dentro de un proceso más grande conocido como KDD (Knowledge Discovery from Databases), este proceso es el que se encarga de extraer la información de las bases de datos a través de los diferentes algoritmos. Sin embargo muchos autores como (Chavez & Bavera, 2013) creen que en la actualidad el KDD es el remplazo de la minería de datos.

Para comprender mejor de que se trata el KDD se en la siguiente figura (Orallo, 2010) definió las 5 fases de KDD:

- Integración y recopilación de datos: Esta es la fase que en la figura se puede ver como Sistemas de información, es de aquí de donde se recolectan los datos, que finalmente deben llegar al almacén de datos o Data WareHouse para posteriormente ser analizados.
- Selección, limpieza y transformación: Esta es la fase de preparación de los datos, y es donde se procede a seleccionar los datos y transformarlos para que lleguen a los sistemas OLAP.
- Minería de datos: Es aquí donde se encuentran los patrones ocultos dentro de los datos, a través del uso de los diferentes algoritmos.
- Evaluación e interpretación: Una vez aplicados los algoritmos, se procede a evaluar los mismos, a encontrar los patrones, realizar las predicciones y analizar el resultado de las mismas.
- Difusión y uso: Finalmente estos datos son presentados a los usuarios finales, esto puede hacerse a través de diferentes aplicaciones de business inteligence ya comercializadas en el mercado o un sistema desarrollado a la medida.

Figura 8: Fases del KDD.



Fuente (Orallo, 2010)

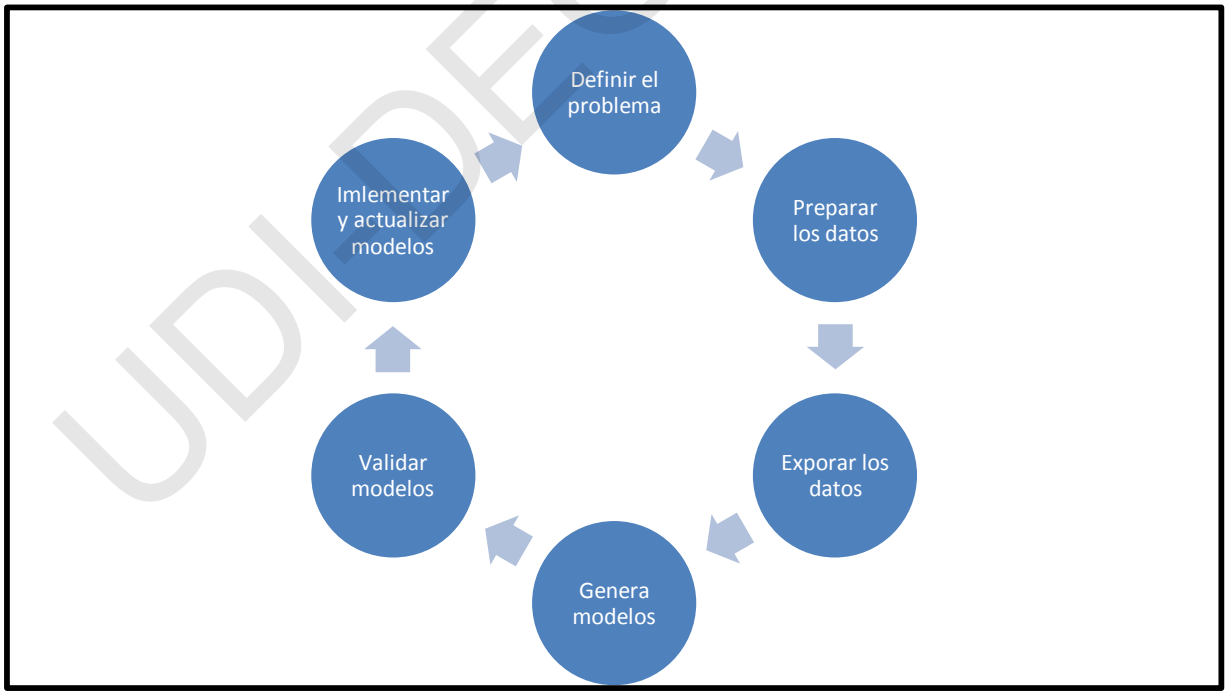
3.1.6.5 Proceso de la minería de datos.

Como se citó anteriormente algunos autores creen que el proceso de minería de datos es muy parecido a lo que es el KDD, sin embargo en la minería de datos debe definirse lo que es un modelo, este modelo se construye mediante la aplicación de los algoritmos de minería de datos, y según (Microsoft, 2014) es un proceso que consta de seis pasos básicos, que se muestran en la siguiente figura.

Este proceso es dinámico e iterativo, es decir que podemos regresar a cualquiera de las fases en distintos momentos, tal y como se muestra en la figura; por ejemplo si estando en la fase de crear el modelo de minería nos damos cuenta que no tenemos los datos suficientes, podemos regresar a buscar datos. Y de esta forma podemos repetir los pasos hasta crear el modelo adecuado.

Se definirán de una forma general cada uno de los pasos para implementar el modelo de minería de datos, de forma que más adelante se puedan detallar los principales algoritmos que puede utilizarse para definir este modelo.

Figura 9: Proceso de la minería de datos.



Fuente de elaboración: Propia.

- Definir el problema:

  Este paso es muy importante ya que a partir de aquí se realizaran las demás fases. Debemos tener claro cuál es el problema que queremos resolver o investigar y definir los objetivos de nuestro proyecto de minería de datos.

  (Microsoft, 2014) Recomienda en esta fase resolver preguntas como:

  - ¿Qué está buscando? ¿Qué tipos de relaciones intenta buscar?
  - ¿Refleja el problema que está intentando resolver las directivas o procesos de la empresa?
  - ¿Desea realizar predicciones a partir del modelo de minería de datos o solamente buscar asociaciones y patrones interesantes?
  - ¿Qué resultado o atributo desea predecir?
  - ¿Qué tipo de datos tiene y qué tipo de información hay en cada columna? En caso de que haya varias tablas, ¿cómo se relacionan? ¿Necesita limpiar, agregar o procesar los datos antes de poder usarlos?

  Es importante poder resolver estas preguntas de forma que se tenga claro cuál es el problema que se pretende resolver, es decir porque nuestra empresa necesita aplicar técnicas de minería de datos.

- Preparar los datos:

  Una vez que se tiene definido el problema que se investigara en esta fase se deben consolidar y limpiar los datos que pueden estar en distintas formas, y en distintos lugares. Dentro de la limpieza de los datos se deben eliminar los datos inválidos, interpretar valores faltantes, determinar los orígenes más precisos y fiables de los datos.

  Es importante que esta fase quede completa antes de proceder a explorar los datos, ya que la exploración no será consistente con datos que no estén limpios o que no sean reales y consistentes.

- Explorar los datos

  Después de que los datos han sido preparados se deben explorar los mismos para tomar las decisiones adecuadas antes de crear el modelo de minería de datos.

Al explorar los datos nos daremos cuenta si son suficientes para lo que necesitamos o si tienen defectos que puedan afectar nuestros resultados, por ejemplo al determinar el valor máximo de ingresos en el mes nos damos cuenta que no es el valor real, debemos de revisar cual es el problema, probablemente algún error en la etapa de preparación de los datos y debamos retroceder a este paso.

- Generar modelos

  En este paso se crea la estructura de minería de datos que se vinculará con el origen de datos ya preparados y explorados, los algoritmos o técnicas de minería de datos que se pueden utilizar son descritos más adelante.

- Explorar y validar los modelos

  Una vez construido el modelo antes de realizar las predicciones en base a este, debe de validarse que los resultados sean correctos y de esta forma asegúranos que el modelo funcione correctamente. Esto se puede hacer creando varias configuraciones para ver cuál es el mejor resultado para el problema que nos planteamos en el primer paso.

- Implementar y actualizar los modelos

  El último paso en el proceso de minería de datos es implementar el modelo en un entorno de producción. Con estos modelos se puede crear consultas para generar estadísticas realizar predicciones, crear informes, etc. Es importante crear un proceso que este actualizando el modelo a medida se van obteniendo nuevos datos o los mismos van cambiando para seguir presentado predicciones o estadísticas reales.

3.1.6.6 Algoritmos de Minería de Datos

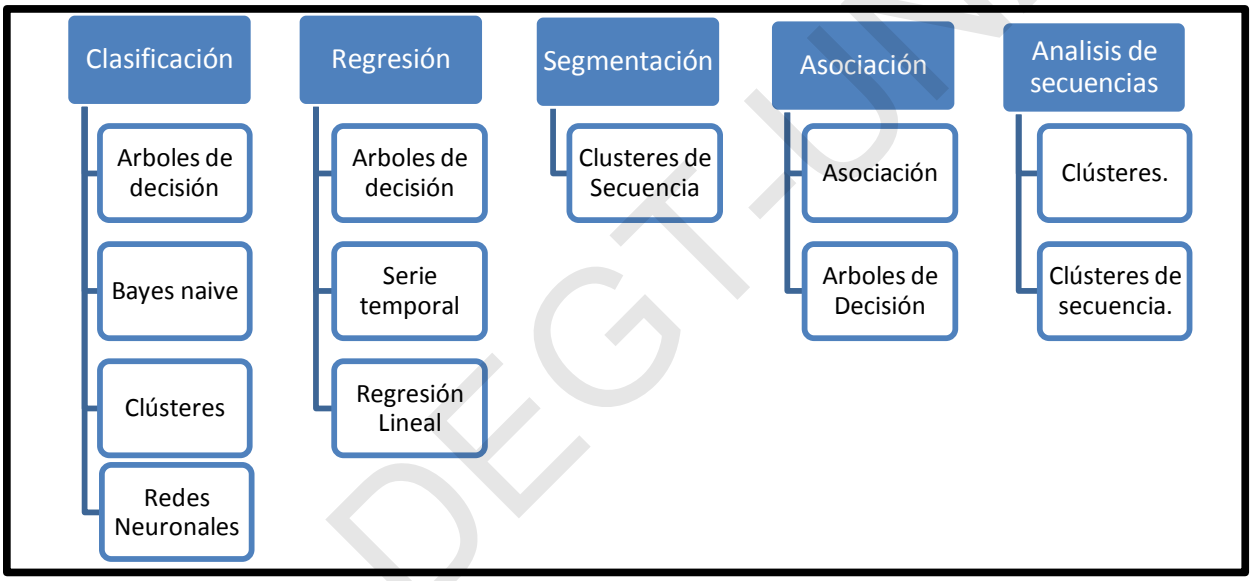
Un algoritmo de minería de datos es el mecanismo empleado para crear un modelo de minería de datos, el algoritmo analiza datos, busca patrones y permite realizar predicciones al aplicar el modelo de minería de datos.

Existen diferentes clasificaciones de algoritmos, los más conocidos citados por (Cerón Reyes & Gómez Díaz, 2010) son los de Clasificación, Regresión, Segmentación, Descripción, Asociación y

exploratorios. (Microsoft, 2014) Cita todas estas clasificaciones y utiliza además los algoritmos de análisis de secuencia, esta clasificación a su vez tiene un conjunto de algoritmos que pueden aplicar una o más características de las diferentes clasificaciones.

En la siguiente figura se muestran la clasificación de los algoritmos de minería de datos, y los nombres de los principales algoritmos en cada categoría.
A continuación se explica brevemente cada una de las clasificaciones y después se explicaran cada uno de los principales algoritmos.

Figura 10: Clasificación Algoritmos de Minería de Datos



Fuente de elaboración: Propia.

3.1.6.6.1 Clasificación de Algoritmos

- Algoritmos de Clasificación

  Estos tipos de algoritmos predicen una o más variables basándose en un conjunto de datos, según (Cerón Reyes & Gómez Díaz, 2010) la meta de estos algoritmos es inducir un modelo para poder predecir una clase dados los valores de los atributos.

- Algoritmos de Regresión

  En este caso (Cerón Reyes & Gómez Díaz, 2010) indica que la meta de estos algoritmos es inducir un modelo para poder predecir el valor de la clase dados los valores de los atributos, estos algoritmos realizan la predicción de una o más variables continuas, como las pérdidas o los beneficios, basándose en otros atributos del conjunto de datos.

- Algoritmos de Segmentación

  Estos algoritmos dividen los datos en grupos de elementos que tengan características similares, según (Rodríguez Suárez & Díaz Amador, 2011) estos concentran los 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 de las otras clases.

- Algoritmos de Asociación

  Estos tipos de algoritmos según cita (Molina López & García Herrero) permiten establecer las posibles relaciones o correlaciones entre distintas acciones o sucesos aparentemente independientes; pudiendo reconocer como la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros

- Algoritmos de Análisis de Secuencias

  Según (Microsoft, 2014) estos algoritmos resumen secuencias o episodios frecuentes en los datos, como un flujo de rutas web, es decir que muestran la secuencia del comportamiento de los datos, pudiendo por ejemplo predecir la navegación de un cliente en una página.

3.1.6.6.2 Principales Algoritmos

- Algoritmo de árboles de decisión

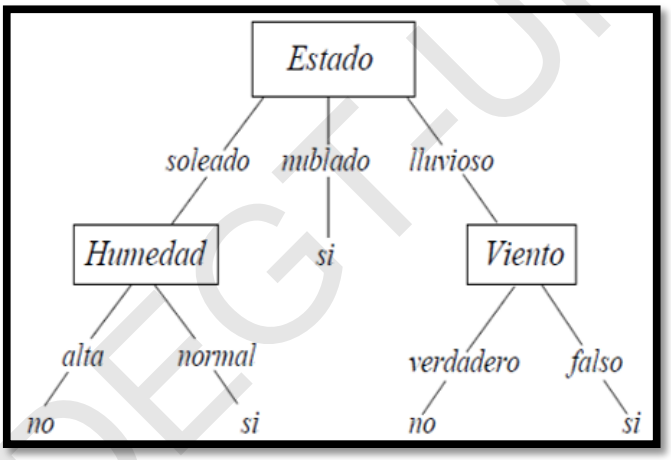
  Este es un algoritmo tanto de regresión como clasificación, según (Cerón Reyes & Gómez Díaz, 2010) los arboles de decisión representan reglas donde atributos independientes determinan los valores finales. En estos árboles cada nodo representa una propiedad que puede tomar diversos valores, cada uno de los cuales genera una rama. Los nodos hojas representan las clasificaciones finales.

En la siguiente figura, se muestra un árbol de decisión, en donde, el nodo Estado, puede presentar tres valores: soleado, nublado y lluvioso. Estos valores a su vez generan una rama con otros posibles valores, finalmente esos valores son las hojas de esta rama; en el caso de la Humedad, las hojas son Alta y Normal.

Por ejemplo el árbol de la figura se utiliza para determinar si se puede jugar al Tenis:

- Si (Estado = Soleado) ^ (Humedad = Alta)
  Entonces Jugar Tenis = No
- Si (Estado = Soleado) ^ (Humedad = Normal)
  Entonces Jugar Tenis = Si

Figura 11: Ejemplo Árbol de Decisión



Fuente de elaboración: (Cerón Reyes & Gómez Díaz, 2010).

(Microsoft, 2014) Hace una diferencia entre los arboles usados para la clasificación y la regresión, en donde, al realizar el algoritmo de clasificación se hacen predicciones basándose en las relaciones entre las columnas de entrada de un conjunto de datos. Utiliza los valores, conocidos como estados, de estas columnas para predecir los estados de una columna que se designa como elemento de predicción. Para la regresión, el algoritmo usa la regresión lineal para determinar dónde se divide un árbol de decisión.

Por ejemplo, en un escenario para predecir qué estudiantes van a ingresar a x carrera, si nueve de diez estudiantes son varones, pero solo uno es mujer, el algoritmo infiere que el sexo es un buen elemento de predicción en la matrícula de x carrera.

- Algoritmo de Bayes Naive

  El algoritmo Bayes naive es un algoritmo de clasificación basado en los teoremas de Bayes. Según cita (Microsoft, 2014) el algoritmo es poco complejo y genera rápidamente modelos de minería de datos que detectan las relaciones entre las columnas de entrada y las columnas de predicción.

  Se puede utilizar este algoritmo para realizar la exploración inicial de los datos y, más adelante, aplicar los resultados para crear modelos de minería de datos adicionales con otros algoritmos más complejos y precisos.

  Según (Cerón Reyes & Gómez Díaz, 2010) cuando no tenemos muy claro qué atributo se puede predecir en función de otros, se utiliza este algoritmo tratando de predecir el valor de todos los atributos en función de todos los atributos (un "todos contra todos").

- Algoritmo de Clústeres

  Según (Microsoft, 2014) este algoritmo utiliza técnicas iterativas para agrupar los casos de un conjunto de datos dentro de clústeres que contienen características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones.

  Los modelos de agrupación en clústeres identifican las relaciones en un conjunto de datos que no se podrían derivar lógicamente a través de la observación casual.

  Por ejemplo, puede discernir lógicamente que las personas desempleadas en Tegucigalpa no son originarias de Tegucigalpa, el algoritmo puede encontrar otras características que no son evidentes acerca de los egresados desempleados que viven en Tegucigalpa.

- Algoritmo de Redes Neuronales

  Citando a (Microsoft, 2014) este algoritmo combina cada posible estado del atributo de entrada con cada posible estado del atributo de predicción, y usa los datos de entrenamiento

para calcular las probabilidades. Posteriormente, puede usar estas probabilidades para la clasificación o la regresión, así como para predecir un resultado del atributo de predicción basándose en los atributos de entrada.

Este algoritmo según (Cerón Reyes & Gómez Díaz, 2010) puede ser adecuado para detectar patrones no lineales, difícilmente descriptibles por medio de reglas y se usa como alternativa al algoritmo de árboles de decisión

- Algoritmo de Serie Temporal

  Este algoritmo puede predecir tendencias basadas únicamente en el conjunto de datos original, sin necesidad de buscar relaciones entre los datos.

  Según (Microsoft, 2014) muchos algoritmos requieren columnas adicionales de nueva información como entrada para predecir una tendencia, sin embrago los modelos de serie temporal no las necesitan.

  Por ejemplo se puede predecir el incremento o disminución de los índices de desempleo con respecto al pasar de los años, solo con tener las estadísticas de los años anteriores.

- Algoritmo de Regresión Lineal

  Este algoritmo según (Microsoft, 2014) es una variación del algoritmo de árboles de decisión que ayuda a calcular una relación lineal entre una variable independiente y otra dependiente y, a continuación, utilizar esa relación para la predicción.

- Algoritmos de Clústeres de Secuencia

  Citando a (Microsoft, 2014) el algoritmo de clústeres de secuencia es un algoritmo que puede utilizarse para explorar los datos que contienen eventos que pueden vincularse mediante rutas o secuencias. El algoritmo encuentra las secuencias más comunes mediante la agrupación, o agrupación en clústeres, de las secuencias que son idénticas.

  Por ejemplo:

  - Registros de transacciones que describen el orden en el que un cliente agrega elementos a una cesta de la compra de un comerciante en línea.

- Registros que siguen las interacciones del cliente (o paciente) a lo largo del tiempo, para predecir cancelaciones del servicio u otros malos resultados.
- Registros de la forma en que un alumno ingresa sus clases en un periodo académico.

- Algoritmo de Asociación

  Los modelos de asociación se generan basándose en conjuntos de datos que contienen identificadores para casos individuales y para los elementos que contienen los casos según (Microsoft, 2014), es así como un grupo de elementos de un caso se denomina un conjunto de elementos. Un modelo de asociación se compone de una serie de conjuntos de elementos y de las reglas que describen cómo estos elementos se agrupan dentro de los casos. Este algoritmo por lo tanto detecta eventos que se producen de manera simultánea.

  Las reglas que el algoritmo identifica pueden utilizarse para predecir las probables clases de un alumno en el futuro, basándose en la carga académica actual de un estudiante.

3.1.6.7 Herramientas de Minería de Datos

Para aplicar los diferentes algoritmos de minería de datos antes mencionados, existen diferentes herramientas en el mercado que hacen esto mucho más fácil, algunas de estas herramientas son nativas de los gestores de bases de datos y otras son independientes, en ambos casos existen herramientas libres y licenciadas.

3.1.6.7.1 Herramientas Nativas del Gestor de Base de Datos

La inteligencia de negocios y la minería de datos se han vuelto una herramienta de suma importancia en las empresas y que aportan gran valor a la misma, es por esta razón que en los últimos años diferentes empresas que distribuyen los gestores de bases de datos han incorporado a estos herramientas para el análisis de datos, facilitando de esta forma la toma de decisiones para la empresa.

Entre algunos de los gestores más importantes que incluyen la utilización de este tipo de herramientas se pueden enumerar los siguientes:

- Oracle
- SQL Server
- DB2

- Minería de Datos con Oracle

  Según (Haberstroh, 2008) Oracle Data Mining permite a las empresas desarrollar aplicaciones de inteligencia de negocio avanzadas que exploten las bases de datos corporativas, descubran nuevos conocimientos e integren esa información en aplicaciones comerciales.

  Oracle tiene diversas funcionalidades de minería de datos, según (Robles Aranda & Sotolongo, 2013) las más importantes son las siguientes:

  - Puede realizar algoritmos de agrupamiento (k-means, O-Cluster).
  - Aplicación de algoritmos de árboles de decisión.
  - Algoritmo Clasificador bayesiano (naive bayes).
  - Máquinas de soporte vectorial (support vector machines).
  - Reglas de asociación (APRIORI).

  Las implementaciones de cada uno de estos algoritmos están incluidos dentro del motor de base de datos, Oracle Data Mining incluye Oracle Data Miner, esta es una interfaz gráfica para los usuarios que permite el análisis de datos con el objetivo de aplicar técnicas de minería de datos. Según (Oracle, 2007) Oracle Data Miner guía al analista de datos a través del proceso data mining con total flexibilidad y presenta los resultados en formatos gráficos y tabulares. Oracle Data Miner puede generar el código PL/SQL asociado con una Actividad de Recuperación de los Datos.

- Minería de Datos con SQL Server

  SQL Server ofrece varias características para crear complejas soluciones de minería de datos, además es un entorno escalable y está basado en un modelo de autoservicio, y los usuarios pueden utilizarlo con facilidad.

  Según (Robles Aranda & Sotolongo, 2013) algunas de las funcionalidades que ofrece son las siguientes:

  - Permite aplicar soluciones de minería de datos utilizando Microsoft Excel.
  - Entender cómo, cuándo y dónde aplicar los algoritmos que se incluyen en el servidor de SQL.
  - Realizar la extracción de datos de procesamiento analítico en línea (OLAP).

- Utilizar SQL Server Management Studio para acceder y proteger los objetos de minería de datos.
- Utilizar SQL Server Business Intelligence Development Studio para crear y gestionar proyectos de minería de datos.
- Los algoritmos implementados por Microsoft son: Árboles de decisión, Bayes naive, Clústeres, Redes neuronales, Serie temporal, Regresión lineal, Clústeres de secuencia, Asociación.

Entre las ventajas de la minería de datos de Microsoft Según (Robles Aranda & Sotolongo, 2013) esta la integración estrecha con la plataforma de base de datos de clase mundial SQL Server, ya que aprovecha el desempeño, la seguridad y las características de optimización de SQL Server; la extensibilidad, ya que se puede extender la minería de datos de Microsoft para implementar algoritmos que no vienen incluidos en el producto.

- Minería de Datos con DB2

  IBM DB2 en la versión Developer Edition proporciona prestaciones avanzadas de gestión de datos e inteligencia de negocios, entre las cuales se incluyen optimización del rendimiento y el almacenamiento, gestión de carga de trabajo, servicios de cubo OLAP, minería de datos para análisis predictivo y de descubrimiento y funcionalidad de textos para analizar contenido no estructurado.

  La minería de datos en DB2 soporta los siguientes modelos de minería de datos:

  - Modelos de asociación
  - Modelos de secuencias
  - Modelos de clasificación
  - Modelos de clústeres
  - Modelos de regresión

3.1.6.7.1 Herramientas Independientes del Gestor de Base de Datos

Debido a la importancia de la inteligencia de negocios en las empresas, distintas compañías han creado muchas herramientas para aplicar la minería de datos y business intelligence, estas empresas venden estas herramientas de forma separada a un gestor de bases de datos, con la facilidad de

acoplarse a diferentes bases de datos, o a una única fuente de información. Algunas de estas herramientas importantes de mencionar, con las siguientes:

- Microstrategy

  Microstrategy es una herramienta de Business Intelligence que ha acoplado el DataMining en su plataforma y es capaz de presentar los resultados de DataMining en los dashboards. Esta herramienta incluye según (Microstrategy, 2015) la creación de modelos y la distribución de los resultados a los usuarios a través del visor de modelos previsibles, que presenta unas características e información gráfica diferente según el tipo de análisis que se realice.

  La minería de datos en DB2 soporta los siguientes modelos de minería de datos:

  - Regresión lineal
  - Agrupación (Clustering)
  - Árbol de Decisiones
  - Series Temporales
  - Asociación.

- Weka

  Weka (Waikato Environment for Knowledge Analysis) es una de las herramientas de minería de datos más conocida y de las más populares, esta herramienta es de código abierto que utiliza la licencia GNU (General Public License).

  Según (Waikato, 2015) Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos y los mismos pueden aplicarse directamente a un conjunto de datos o a través de código Java.

  La minería de datos con Weka soporta los siguientes modelos de minería de datos:

  - Algoritmos de clasificación
  - Algoritmos de regresión
  - Algoritmos de cluster
  - Algoritmos de asociación

- Rapid Miner

Rapid Miner es otra herramienta de código abierto y actualmente es uno de los líderes en minería de datos según (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006), RapidMiner es un entorno para el aprendizaje automático y permite el diseño de cadenas de operadores complejos anidados para un gran número de problemas de aprendizaje.

Rapid Miner soporta los siguientes algoritmos de datos:

- Algoritmos de clasificación
- Algoritmos de regresión
- Arboles de decisión.
- Algoritmos de cluster
- Algoritmos de asociación

3.1.7 Desempleo

Entiéndase como desempleo la situación de aquellas personas que pudiendo trabajar carecen de empleo, para referirse al número de desempleados de la población se utiliza la tasa de desempleo por región o país.

No todas las personas que se encuentran sin empleo son considerados desempleados, según (Zelaya, 2013) podemos clasificar a las personas que no tiene empleo en tres tipos:

1. Inactivos: Esta categoría comprende el porcentaje de la población adulta que está estudiando, realiza tareas domésticas, está jubilada, está demasiado enferma para trabajar o simplemente no está buscando trabajo.

2. Desempleados: Se dice que una persona está desempleada, si no está trabajando y ha realizado anteriormente esfuerzos para encontrar empleo en el último mes o 2 semanas.
   Es importante destacar que se considera desempleados solo si esta persona ha realizado esfuerzos por conseguir un empleo.

3. Desocupados: Incluye a los cesantes y a los trabajadores nuevos que no trabajaron, pero están disponibles para hacerlo y realizaron acciones de búsqueda de empleo.
   A su vez el desempleo se puede dividir en dos subtipos:

- Desempleo Visible: Son las personas que trabajan menos de 44 horas a la semana, y no pueden encontrar otra fuente de empleo para suplir el resto de horas.

- Desempleo Invisible: Son las personas que trabajan 44 horas o más a la semana pero que no ganan el salario mínimo.

Según (Blanchard, 2006) **La tasa de desempleo** se calcula como el número de desempleados dividido por la población activa, y se expresa en forma de porcentaje. Es decir, no es una proporción entre el total de la gente desempleada y el total de la población, sino el de aquélla que se denomina "económicamente activa" que son los desempleados definidos anteriormente más las personas empleadas.

3.1.8 Subempleo

Entiéndase por subempleo a la situación de aquellas personas que poseen un trabajo, pero que necesitan trabajar más, ya sea porque su salario es inferior al mínimo o porque trabajan menor cantidad de horas.

(Ramírez Rojas & Guevara Fletcher, 2006) Definen tres criterios para identificar, entre las personas ocupadas, a las visiblemente subempleadas:

- Trabajan menos de la duración normal.
- Lo hacen de forma involuntaria.
- Desean trabajo adicional y están disponibles durante el período de referencia.

El subempleo puede clasificarse en dos tipos:

- Subempleo visible: aquellas personas que trabajan menos de 36 horas a la semana.
- Subempleo invisible: aquellas personas que tienen ingresos menores a un salario mínimo, y trabajan más de 36 horas.

3.1.8.1 Subempleo Profesional o Sobreducación

El subempleo profesional se refiere a todas aquellas personas con títulos universitarios que están trabajando en una profesión o actividad distinta a su carrera universitaria.

Se produce sobreducación cuando el esfuerzo educativo no recibe suficientes compensaciones económicas ni sociales en el mercado laboral. Tal situación, en principio, es consecuencia del aumento del nivel educativo de la población demandante de empleo.

Existen dos tipos de subempleo profesional según (Gobernado Arribas, 2007):

- Subempleo profesional generalizado: aquel que afecta a la mayoría, por ejemplo todos aquellos con un título universitario.

- Subempleo profesional relativo: aquel que solo afecta a una minoría, ya que su nivel educativo es superior, por ejemplo personas con doctorados.

3.1.9 Desempleo en Honduras

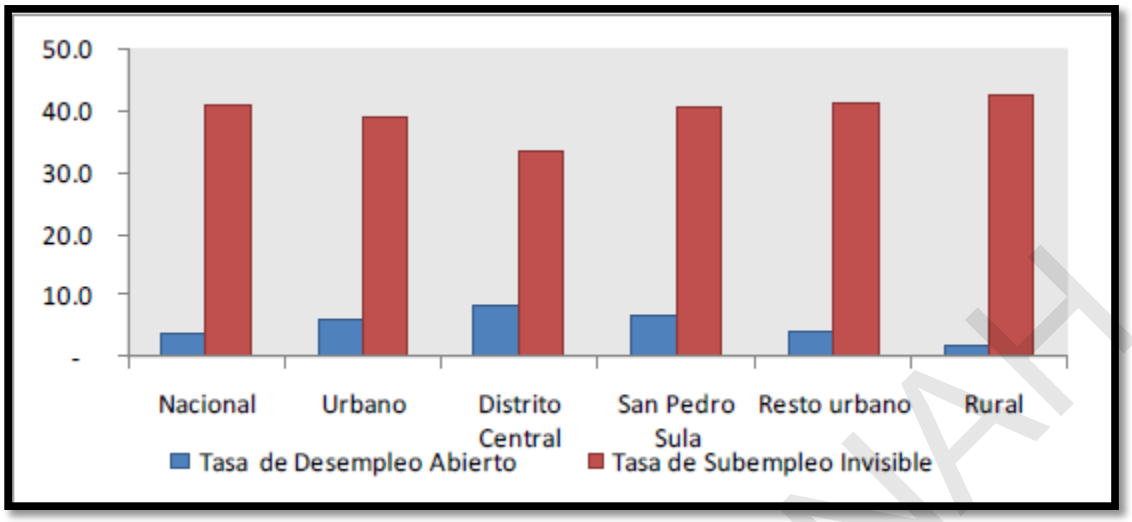
El desempleo en general es un problema de carácter personal y social, debido a que puede ocasionar problemas familiares por la reducción de ingresos, y esto a la vez puede afectar a otras personas, como la dueña de la pulpería donde se compran los alimentos, las tiendas donde compramos la ropa, los negocios a donde antes acudíamos, y sumando todos estos afecta a todo el país.

Mucha de la población hondureña se encuentra desempleada o desocupada, e inclusive aquí cabe una nueva clasificación que son los desalentados, que son aquellas personas que creen que ya no existe oportunidad de que consigan un empleo, y esto los hace pasarse a las personas inactivas.

Según las cifras del Observatorio de Mercado Laboral (OML) Honduras cerró el año 2013 con una tasa de desempleo abierto de 3.9 por ciento, lo que significa que son 141,724 personas desempleadas, lo cual significa un aumento de 1.376% con el año 2012 lo que representaba 13,940 personas sin trabajo; tal como lo muestra el grafico 1.

El problema del desempleo es mayormente urbano, potenciado probablemente por la migración constante de personas del campo a la ciudad y la poca capacidad del mercado laboral para absorber esta fuerza de trabajo. Mientras la TDA urbana se estima en 6.0%, la rural es de 2.0%; el Distrito Central tiene la mayor tasa de desempleo 8.6%.

Gráfico 1: Tasa de Subempleo Invisible y Tasa de Desempleo Abierto



Fuente (INE, 2013)

Existen también una elevada parte de la población económicamente activa trabaja menos de la jornada reglamentaria a la semana y desean trabajar más; y otras que trabajando más de la jornada reglamentaria tienen ingresos inferiores a un salario mínimo.

Según el (INE, 2013) el desempleo se concentra en la población joven; del total de 141,724 desempleados del país, de los cuales (47.8%) son jóvenes menores de 24 años.

3.1.9.1 Desempleo de Egresados Universitarios

Como se pudo observar anteriormente la tasa de desempleo en Honduras es alta y registra aumentos todos los años, es una situación preocupante para todos los hondureños sobre todo los que se encuentran sin trabajo.

Esta es una situación todavía más preocupante para los hondureños, ya que muchos logran graduarse de la universidad con grandes sacrificios de sus padres o de ellos. Las principales razones por las cuales las personas estudian una carrera universitaria son:

- Superación Personal.
- Conseguir un Empleo.
- Conseguir un mejor Salario.

Sin embargo existen muchas personas que están desempleadas, por lo cual las 2 últimas razones no se cumplen, y sumándole a esto el esfuerzo que muchas personas realizan para poder estudiar nos encontraremos con un gran porcentaje que serán desempleados desalentados por no haber logrado su objetivo.

Según estadísticas realizadas por el (INE, 2013) existen en el país 408,875 personas con Subempleo Visible, y 1,422,210 con subempleo Invisible. Esto se refleja en una Tasa de Subempleo Invisible (TSI) nacional del 40.8%; como se muestra en la siguiente tabla:

Tabla 2: Tasa de subempleo en Honduras.

Categoría	Subempleo		Potencialmente Activos	Desalentados
	Visible	Invisible		
Nacional	408,875	1422,210	57,915	152,529
urbano	192,268	643,586	17,211	67,340
Rural	216,607	778,624	40,705	85,189

Fuente (INE, 2013)

Dentro del grupo de personas con desempleo invisible, se encuentran profesionales egresados de las distintas universidades que al no conseguir un empleo conforme a su carrera, trabajan en empleos no relacionados con lo que estudiaron o donde no ejercen su profesión, por las necesidades económicas que se presentan.

Lo anterior demuestra, que el principal problema del mercado laboral no es el desempleo, sino el subempleo invisible, que se asocia a bajos ingresos con empleos de baja productividad.

¿Pero a qué se debe que personas capacitadas salgan de la universidad y no encuentren trabajo, será que la situación económica de Honduras es el único motivo?

Según Carlos Madero Erazo, subdirector de empleo de la Secretaría de Trabajo, en entrevista con Diario La Tribuna en diciembre del 2012 manifestó que en varias investigaciones se concluyó que la oferta de los profesionales universitarios no está acorde al mercado de trabajo. "Se están formando a ciudadanos en carreras que probablemente ya las posibilidades de vacantes están muy limitadas y no existe un sistema que permita acomodar las estructuras de trabajo para que la oferta laboral que salga de las universidades atienda a la demanda de trabajo".

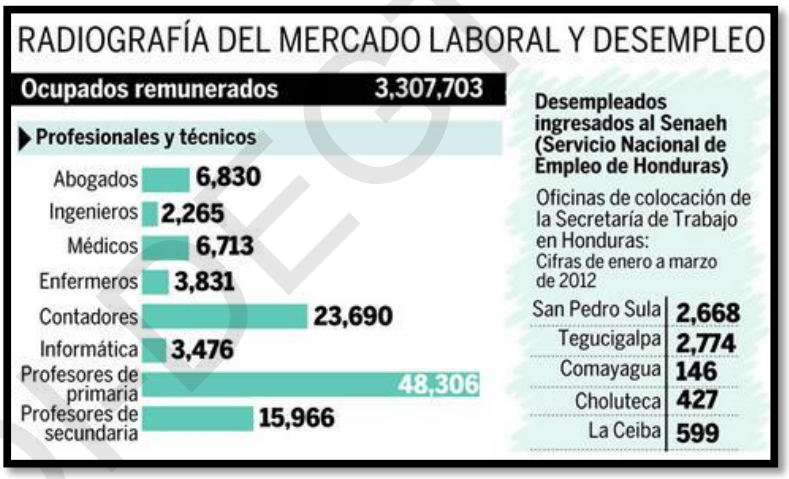
3.1.10 Factores que influyen al desempleo y subempleo profesional de egresados universitarios.

Además de lo mencionado por el subdirector de empleo de la Secretaría de Trabajo en el 2012, otras de las principales razones por las que existen altos índices de desempleo de egresados universitarios según estudios realizados por Manpowergroup (Hondurasq, 2013) son las siguientes:

- La falta de competencias técnicas
- Falta de experiencia
- Falta de competencias en el lugar de trabajo
- Búsqueda de mayor sueldo y candidatos no dispuestos a trabajar en puestos de tiempo parcial o eventuales

Según estudios realizados por (diariowebcentroamerica, 2012) muestra en la siguiente figura las profesiones con mayores problemas para conseguir empleos en Honduras:

Figura 12: Mercado Laboral y Desempleo.



Fuente diario web Centroamérica.

Además de las carreras mencionadas en la figura anterior, una de las carreras universitarias que tienen mayor problema de empleo es Administración de Empresas, porque la tasa de desempleo abierta es de 9.5 por ciento. O sea que es la carrera que tiene mayores graduados universitarios y que se encuentran desempleados.

Los meses buscando trabajo para desempleados con educación secundaria según el (INE, 2013) son 2.2. Para los universitarios son 3.1 meses. Esto hace pensar que las personas con educación universitaria tienen más problemas para conseguir empleo.

En estos meses muchas personas no encuentran empleo, por esta razón aceptan cualquier empleo aunque no esté relacionado con su profesión universitaria y esta es la principal causa del subempleo profesional.

Otro factor que influye en el desempleo, es el sector donde las personas quieren trabajar, por ejemplo los egresados de medicina prefieren trabajar en ciudades como Tegucigalpa o San Pedro Sula donde esta área se encuentra saturada, cuando en el interior del país probablemente hay oportunidad laboral.

Este factor también produce subempleo profesional, ya que las personas evitan ir a trabajar al interior del país o a la zona rural y aceptan otro tipo de empleo.

Sin duda otro factor muy importante es el salario que se le paga una persona recién egresada, el promedio salarial en los hogares cuyos jefes poseen educación superior, llega a los Lps.8,410.00 según (INE, 2013), lo cual es bastante bajo para la preparación que han tenido, para el esfuerzo que se va a realizar y considerando que normalmente las empresas solicitan maestrías o experiencia antes de trabajar, lo que reduce el grupo de profesionales aptos para el puesto.

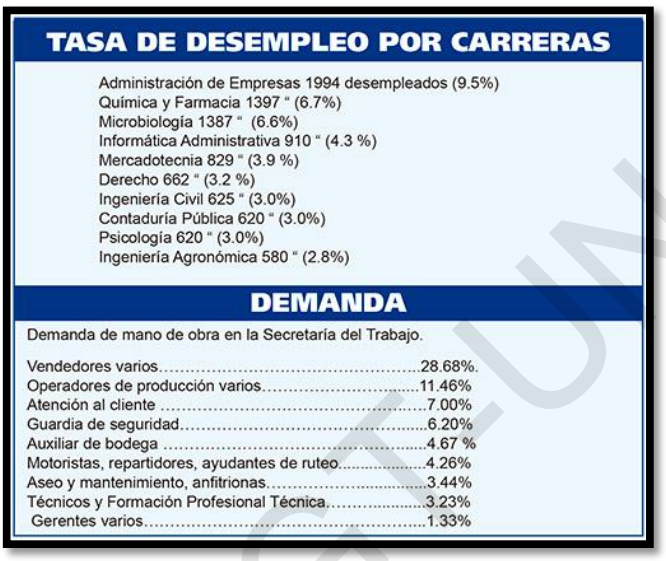
Para mayo de 2013 el 64.5% de los hogares hondureños se encuentran en condiciones de pobreza, ya que sus ingresos se encuentran por debajo del costo de una canasta básica de consumo que incluye alimentos y otros bienes y servicios. Por esta razón los profesionales universitarios aceptan cualquier trabajo sin importar que se relacione son su título universitario.

También existen carreras como ingeniería en sistemas en donde el porcentaje de desempleados es de 1%, lo cual es bastante bajo en comparación con otras carreras, podemos ver un resumen de tasa de desempleo por carreras para el año 2012 según la Secretaría de Trabajo en la siguiente figura.

A pesar de que las autoridades universitarias recomiendan que los estudiantes se matriculen en carreras que tienen tasas de desempleo inferiores, como las áreas de productividad, tecnología y

telecomunicaciones, las medidas que se deben aplicar deberían de ser mayores, por ejemplo establecer un límite de personas en las carreras, para evitar la sobrepoblación de las mismas.
Esto no solo generara menor cantidad de egresados en carreras que tienen grandes índices de desempleo si no que permitirá a la universidad prestar un mejor servicio a sus estudiantes.

Figura 13: Tasa de Desempleo por Carrera.



Fuente Secretaria de Trabajo, 2010

También es necesario que las autoridades tomen en cuenta el comportamiento del mercado para cambiar los planes de estudio según las exigencias del mercado actual, crear distintos problemas que fortalezcan al estudiante y que le permitan adquirir una mayor experiencia que lo ayude a competir al momento de egresar y salir a buscar un empleo.

El gobierno y la empresa privada también deben crear iniciativas en donde le den prioridad al profesional universitario para evitar el subempleo, ya que otra causa de este es que las empresas prefieren contratar recursos humanos más baratos y menos calificados y no profesionales universitarios. En algunos casos se contratan egresados pero con un salario inferior o igual al de una persona sin título.

Según el (INE, 2013) en aquellos hogares cuyos jefes poseen educación superior, el ingreso del hogar llega a los Lps.8,410.00. Para mayo de 2013 el 64.5% de los hogares hondureños se

encuentran en condiciones de pobreza, ya que sus ingresos se encuentran por debajo del costo de una canasta básica de consumo que incluye alimentos y otros bienes y servicios. Por esta razón los profesionales universitarios aceptan cualquier trabajo sin importar que se relacione son su título universitario.

CAPITULO IV. ENFOQUE Y TIPO DE INVESTIGACIÓN

4.1 Enfoque de investigación

La investigación se desarrollara bajo un conjunto de procesos secuenciales y probatorios, esto según lo planteado por (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010) es una investigación cuantitativa, en donde cada proceso será realizado por medio de etapas, en donde una etapa precede a la otra.

4.2Tipo de Investigación

Esta investigación basados en es de tipo descriptiva, ya que se definirán los principales conceptos de la minería de datos aplicada al desempleo y subempleo profesional, además se establecerá una relación entre las causas del el subempleo profesional y el desempleo de profesionales universitarios.

Adicionalmente en esta investigación se diseñara una parte práctica en donde se aplicara la minería de datos para determinar los índices de subempleo y desempleo en profesionales universitarios.

CAPITULO V: VARIABLES

5.1 Variables

Tabla 3: Definición de Variables,

Variable	Definición Conceptual	Indicador
Desempleo en profesionales universitarios	Es la parte de la población que estando en edad, condiciones y disposición de trabajar (población activa) carece de un puesto de trabajo. (Cfr. Samuelson, 2006)	• Cantidad de desempleados con título universitario. • Tiempo transcurrido de desempleo. • Cantidad de entrevistas de trabajo.
Subempleo Profesional	Profesionales que desempeña una ocupación que no tiene relación con los estudios que cursaron. (Moreno & Burga, 2011)	• Cantidad de empleados con título universitario y sin trabajar en su carrera. • Horas Trabajadas. • Salario
Causas de desempleo de profesionales universitarios	Un factor es lo que contribuye a que se obtengan determinados resultados al caer sobre él la responsabilidad de la variación o de los cambios.	• Competencias técnicas. • Experiencia. • Salarios Bajos • Crisis económica • Sobrepoblación de la carrera. • Tiempo de búsqueda de trabajo.
Causas de subempleo profesional	Causas por las que se emplea a alguien en un cargo o puesto inferior al que su capacidad le permitiría desempeñar. (Diccionario de la Lengua Española, 2012)	• Salario. • Tiempo de búsqueda de un trabajo. • Competencias técnicas. • Experiencia. • Ingresos familiares. • Crisis económica. • Sobrepoblación de la carrera.

Fuente: Elaboración propia

5.2 Operacionalizacion de las variables

Tabla 4: Definición Operacional de las variables,

Variable	Definición Operacional
Desempleo en profesionales universitarios	Profesionales con títulos universitarios que se encuentren sin trabajo.
Subempleo Profesional	Profesionales con título universitario que trabajan en algo distinto a su profesión universitaria.
Causas de desempleo de profesionales universitarios	Razones por las que existen altos índices de desempleo de los egresados de las universidades.
Causas de subempleo profesional	Razones por las que existen altos índices de subempleo profesional de los egresados de las universidades.

Fuente: Elaboración Propia

CAPITULO VI: ESTRATEGIA METODOLOGICA

6.1 Diseño de la Investigación

Esta investigación es de tipo transeccional descriptiva, ya que se tiene como objetivo indagar la incidencia de más de una variable en la población y después presentar la descripción de la misma. En esta investigación las variables a estudiar y describir son el desempleo, subempleo profesional en Honduras y las causas de ambos.

Debido a que esta investigación es de tipo descriptiva no se presentan hipótesis en la misma, pero se realizara una parte práctica en la cual se aplicaran técnicas de minería de datos para poder estimar las causas del desempleo y subempleo profesional en Honduras.

6.2 Población, Muestra y Muestreo

6.2.1 Delimitación de la población

Tomando en cuenta los objetivos de esta investigación la población que se estudiará, comprende todas aquellas personas con títulos universitarios que se encuentren actualmente desempleados y aquellos que se encuentren trabajando en algo distinto a su carrera, es decir con un subempleo profesional.

6.2.2 Tamaño de la Muestra

Según el (INE, 2013) la cantidad de desempleados en Honduras son 141,724, pero los desempleados con educación superior corresponden a 19,180.

La cantidad de personas con educación superior con subempleo tanto visible como invisible corresponden a 78,610.

Para calcular la muestra se utilizó el programa PHSTAT de Pearson Education, con los siguientes valores:

$$\text{Probabilidad de Ocurencia} = 50\%$$

$$\text{Error de muestreo} = 5\%$$

$$\text{Nivel de Confianza} = 95\%$$

$$\text{Población Desempleada} = 19180$$

$$\text{Población con Subempleo} = 78610$$

$$\text{Población Total} = 97790$$

En la siguiente tabla se muestra el detalle del cálculo realizado y se obtiene que la muestra a utilizar es de 383 personas.

Tabla 5: Calculo de la muestra

Data	
Population Standard Deviation	**0,5**
Sampling Error	**0,05**
Confidence Level	**95%**
Intemediate Calculations	
Z Value	-1,9600
Calculated Sample Size	384,1459
Result	
Sample Size Needed	**385,0000**
Finite Populations	
Population Size	**97790**
Sample Size with FPC	382,6467
Sample Size Needed	**383**

Fuente: PHSTAT

6.2.3 Tipo de muestreo

En esta investigación se utilizara una muestra probabilística simple, ya que todos los individuos de la población de desempleados y subempleados profesionales tienen la misma probabilidad de ser escogido.

Como no fue posible ubicar por universidad o geográficamente a toda la población de profesionales desempleados o subempleados, no se aplicó una fórmula para conocer las personas exactas a las cuales se debía aplicar el cuestionario. Por lo tanto la selección de esta muestra de personas a investigar se realizó al azar, y se obtuvieron únicamente indagando si cumplían la característica

base de esta investigación, que son personas con títulos universitarios desempleadas o subempleadas, no se hizo una selección por universidad, carrera o género, por lo cual la muestra es completamente probabilística.

6.3 Recolección de Datos

6.3.1 Instrumento de Investigación.

Para obtener la información necesaria para desarrollar esta investigación, se utilizara el Cuestionario como principal técnica de recolección de información, este será aplicado a personas con títulos universitario con problemas de desempleo o subempleo profesional a la muestra definida anteriormente.

Para eso se definen a continuación las secciones a abordar en el instrumento, los objetivos del mismo y las variables a utilizar.

Esta encuesta está dividida en tres secciones:

1. Información general (IG)
2. Información de los desempleados (ID)
3. Información de subempleados profesionales (ISP)

Objetivos:

1. Determinar las principales causas del desempleo de profesionales universitarios.
2. Determinar las principales causas del subempleo de profesionales universitarios.

Este instrumento fue aplicado al azar entre profesionales de la ciudad de Tegucigalpa, cabe destacar que muchas de las personas potenciales a encuestar no pudieron encuestarse, ya que no todas las personas tenían conocimiento acerca de su situación laboral en cuanto al subempleo, y asumían que su trabajo estaba relacionado con su carrera.

Muchas personas además presentan resistencia a llenar encuestas, a pesar de que los datos son anónimos, otras de las personas no contestaron adecuadamente las preguntas, el porcentaje de estas fue mínimo, pero las encuestas debieron descartarse.

Finalmente con la información recolectada y procesada por medio del instrumento de investigación de realiza el análisis descrito en capítulos más adelante y adicionalmente se aplican algoritmos de minería de datos para cumplir con los objetivos de la investigación.

6.3.2 Validez del Instrumento

La validación del instrumento de investigación se realizó a través del juicio de expertos. El instrumento fue validado por tres expertos, un experto en el área de minería de datos (E1), otro especialista en investigación (E2) y otro del área de ciencias sociales (E3).

También se validó el orden del contenido y la redacción del mismo.

Además se realizó una prueba con 2 personas para determinar si el instrumento era lo suficientemente entendible para ser aplicado.

Después del juicio de los expertos, se realizó el cálculo del índice de la validez del instrumento según Lawshe de la siguiente forma:

$$IV = \frac{P}{N}$$

$Donde$: $P = resultado\ de\ validez$

$N = numero\ de\ expertos$

Las tablas 6 y 7 muestran la validación aplicada por los expertos y el cálculo del índice de validez por sección del instrumento. Según este cálculo, se determina que el índice de validación (IV) menor es de un 67%, en ese caso solo se tuvo un rechazo por experto, y las demás preguntas fueron aceptadas en un 100%

En base a todo esto se realizaron modificaciones y el instrumento final se encuentra en el anexo 1 de este documento.

Tabla 6: Juicio de expertos para la sección de subempleo profesional

Sección de subempleo profesional				
Ítem	**E1**	**E2**	**E3**	**IV**
1	1	0	1	0.67
2	1	0	1	0.67
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	0	0.67
7	1	1	1	1
8	1	1	1	1
9	0	1	1	0.67
10	0	1	1	0.67
11	1	1	1	1
12	1	1	1	1
13	1	1	1	1
14	1	1	1	1
15	1	1	1	1
16	1	1	1	1
17	1	1	1	1

Fuente: Elaboración Propia

Tabla 7: Juicio de expertos para la sección de desempleo profesional

Sección de desempleo				
Ítem	**E1**	**E2**	**E3**	**IV**
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	1	0	1	0.67
6	1	1	1	1
7	1	1	1	1
8	1	1	1	1
9	1	1	1	1
10	1	1	1	1
11	1	1	1	1
12	1	1	1	1
13	1	1	1	1
14	1	1	1	1
15	1	1	1	1
16	1	1	1	1

Fuente: Elaboración Propia

6.3.3 Confiabilidad del instrumento

Para determinar la confiabilidad del instrumento se aplicó una prueba piloto a un grupo de 23 personas, dentro de los cuales se encontraban personas con graduadas de la universidad desempleadas y con empleos.

La confiabilidad se determina con el coeficiente alfa de Cronbach, de la siguiente forma:

$$\alpha = \frac{K}{K-1}\left[1-\frac{\sum S_i^{\,2}}{S_T^{\,2}}\right]$$

Donde: $K : El\ número\ de\ items$

$$\sum S_I^2 : Sumatoria\ de\ la\ varianza\ de\ los\ items$$

$$S_T^2 : Varianza\ de\ la\ suma\ de\ los\ items$$

$$\propto : Coeficiente\ Alfa\ de\ Cronbach$$

Con la tabulación de estas encuestas se obtuvo lo siguiente para cada una de las variables:

6.3.3.1 Desempleo

$$\propto = \frac{8}{8-1}\left[1-\frac{1.45}{9.74}\right] = 0.973 \cong \mathbf{0.97}$$

Este resultado indica que el instrumento es altamente confiable para la variable de desempleo, ya que el valor se aproxima a 1.

6.3.3.2 Subempleo

$$\propto = \frac{10}{10-1}\left[1-\frac{2.85}{9.88}\right] = 0.7905 \cong \mathbf{0.79}$$

Este resultado indica que el instrumento es altamente confiable para la variable de subempleo, ya que el valor se aproxima a 1.

6.3.3.3 Causas del Desempleo

$$\propto = \frac{19}{19-1}\left[1-\frac{8.14}{29.91}\right] = 0.7682 \cong \mathbf{0.77}$$

Este resultado indica que el instrumento es altamente confiable para la variable de causas de desempleo, ya que el valor se aproxima a 1.

6.3.3.4 Causas del Subempleo

$$\propto = \frac{14}{14-1}\left[1 - \frac{11.92}{91.74}\right] = 0.9369 \cong \mathbf{0.94}$$

Este resultado indica que el instrumento es altamente confiable para la variable de causas de subempleo, ya que el valor se aproxima a 1.

Finalmente después de estas validaciones, se puede concluir que el instrumento es válido para ser aplicado y recolectar la información necesaria.

6.3.4 Análisis del Instrumento por Variable

A continuación se definen las variables a medir y en que sección del instrumento serán medidas.

Tabla 8: Variables en el Instrumento.

Variable	Indicador	Sección del cuestionario	Numero de Ítem del cuestionario
Desempleo	Cantidad de desempleados	IG ISP ID	7 5, 16 1, 3, 4, 5, 16
Subempleo profesional	Cantidad de personas con subempleo.	ISP ID	2, 11, 13, 15 6, 10, 15
	Salario	ISP ID	7 12
	Horas Trabajadas	ISP	8
Causas del desempleo	Experiencia laboral	ID ISP	2, 9 17
	Competencias técnicas	ID ISP	2, 8 17
	Salarios	ID ISP	2, 8, 12, 14 4

Variable	Indicador	Sección del cuestionario	Numero de Ítem del cuestionario
	Ofertas en las empresas	ID ISP	2, 4 4
	Crisis económica	ID	2
	Sobrepoblación de la carrera	ID ISP	2 4
	Tiempo de búsqueda de trabajo	ID	3
Causas del subempleo	Salarios	ISP ID	4, 12, 14, 17 7, 11
	Experiencia laboral	ISP ID	12, 14, 17 7, 11, 4
	Competencias Técnicas	ISP ID	12, 14, 17 7, 4
	Crisis económica	ISP	12, 14, 17
	Sobrepoblación de la Carrera	ISP	17
	Tiempo de búsqueda de trabajo	ISP ID	3, 5, 12, 17 4
	Ingresos Familiares	ISP ID	12, 14 7, 11, 4

Fuente: elaboración Propia.

CAPITULO VII: PLAN DE ANALISIS

El análisis de los datos obtenidos para alcanzar los objetivos planteados se desarrolla bajo el siguiente plan:

7.1 Selección de la Herramienta de Minería de Datos a Utilizar

Para poder desarrollar y aplicar la minería de datos, previamente se realizó una elección entre los diferentes gestores de bases de datos que cuentan con herramientas de minería de datos, esta elección se basó en lo estudiado en el marco teórico de esta investigación.

Después de analizar entre las diferentes herramientas, se optó por seleccionar SQL Server como gestor de base de datos ya que este es básicamente el líder en el mercado en el manejo de grandes volúmenes de datos.

Adicionalmente SQL Server cuenta con la ventaja de poseer una herramienta nativa para la minería de datos, Analysis Services, a diferencia de otros gestores que están limitados a las herramientas de minería de datos , ya que lo que ofrecen nativamente son soluciones para el manejo de OLTP pero de una forma muy manual. Microsoft se encuentra además entre los líderes en el mercado en Inteligencia de Negocios según el cuadrante de Gartner (Gartner Inc, 2015), lo cual hace sumamente confiable y efectivo a SQL Server y Analysis Services.

7.2 Tabulación de los Datos

La tabulación de los datos obtenidos para esta investigación se divide en dos partes, ya que los datos son obtenidos de dos fuentes distintas:

7.2.1 Encuestas Aplicadas:

En el caso de los datos obtenidos al aplicar el instrumento de investigación, estos primero son tabulados a Excel para determinar si la encuesta fue completada correctamente, y después de esto proceden a cargarse a la BD que fue definida para desarrollar la minería de datos.

La encuesta fue aplicada en distintos lugares de la ciudad de Tegucigalpa, y se escogieron personas al azar, que presentaran problemas laborales de desempleo y subempleo.

7.2.2 Datos Históricos del INE

Los datos históricos obtenidos por medio del INE, son únicamente resúmenes de los datos totales y estos corresponden a 13 años, por lo cual la carga de los mismos se realiza en la base de datos directamente.

7.3 Aplicación de la metodología de minería de datos

En esta sección se desarrollaran los pasos necesarios para poder aplicar la metodología de minería de datos, orientada a cumplir con los objetivos de esta investigación y con base en el marco teórico desarrollado.

Para eso definen un conjunto de pasos, que debería comprender una metodología de minería de datos, estos son los siguientes:

a) Análisis de la problemática.
b) Diseño y creación de la base de datos OLTP
c) Diseño y creación de la base de datos OLAP
d) Diseño y creación del proceso de ETL.
e) Definición de los algoritmos de minería de datos a utilizar.
f) Aplicación de los algoritmos de minería de datos.
g) Selección del algoritmo de minería de datos.
h) Análisis de Resultados.

Sin embargo por la limitante de la cantidad de datos con los que se cuenta, definida al inicio de esta investigación, no se realizaran los pasos c y d. La base de datos OLAP no será desarrollada debido a que no estará en constante movimiento porque únicamente será utilizada para esta investigación y será alimentada únicamente por datos recolectados en la tabulación de los mismos. Así mismo al no contar con una base de datos OLAP, no es necesario desarrollar el proceso del ETL.

Al finalizar esta sección se podrá obtener un análisis de la aplicación de las técnicas de minería de datos utilizadas y realizar posibles predicciones de la problemática laboral planteada en esta

investigación, con base en estos resultados se fundamentaran y analizaran las conclusiones de este estudio.

7.4 Análisis de los Datos de las encuestas

Una vez que los datos recolectados por medio del instrumento de investigación hayan sido tabulados y almacenados en la base de datos desarrollada en la metodología de minería de datos, se procederá a desarrollar un análisis de las características comunes encontradas en los datos para generar información relevante para esta investigación.

Es así como los datos serán evaluados por categoría y variables, mostrando los resultados relevantes en gráficos, para poder analizar cada uno de ellos y con base a los resultados reflejados fundamentar las conclusiones en cada uno de ellos.

7.5 Análisis final de los resultados

Después de haber realizado el análisis de los datos recolectados en las encuestas y el resultado de la aplicación de la metodología de minería de datos, se procederá a realizar un análisis final de la relación entre los resultados de las encuestas y la minería de datos.

Finalmente se desarrollara un análisis para determinar si se alcanzaron o no los objetivos de la investigación, y hacer un resumen de los resultados finales obtenidos en la investigación.

CAPITULO VIII: METODOLOGIA DE MINERIA DE DATOS

Para poder aplicar las técnicas de minería de datos y con base en lo estudiado en esta investigación se definieron los pasos a seguir como una metodología para aplicar las técnicas de minería de datos, a su vez estos pasos fueron ajustados a las necesidades de la investigación, esta metodología y sus resultados se definen a continuación:

a) Análisis de la problemática.
b) Diseño y creación de la base de datos OLTP
c) Definición de los algoritmos de minería de datos a utilizar.
d) Aplicación de los algoritmos de minería de datos.
e) Selección del algoritmo de minería de datos.
f) Análisis de Resultados.

En el caso de esta investigación no se desarrolló una base de datos OLAP, debido a que la información con la que se cuenta es poca, además de esto la BD no estará en constante movimiento porque únicamente será utilizada para esta investigación y fue alimentada por las encuestas aplicadas y por los datos estadísticos proporcionados por el INE.

Para el caso de la investigación únicamente se utilizara la BD OLTP y la minería se realizó directamente de estos objetos. Al no tener una BD OLAP no fue necesario crear el proceso ETL y por lo tanto la metodología definida no incluye el análisis de estos puntos.

Si bien es cierto el proceso de ETL es fundamental en el desarrollo de una minería de datos, ya que los datos deben formatearse y transformarse antes de ser utilizados por los algoritmos, en el caso de esta investigación este paso se puede omitir, debido a que en la tabulación de los datos a la base de datos OLTP, los mismos fueron clasificados para su correcto uso en los algoritmos.

A continuación se desarrollan los pasos definidos en la metodología de minería de datos a aplicar en esta investigación:

8.1 Análisis de la Problemática

Según el último informe del instituto nacional de estadística (INE, 2013) en mayo del 2013, la tasa de desempleo abierto a nivel nacional se estima en un 3.9% de la Población Económicamente Activa (PEA); lo cual equivale a 141,724 personas, se estima que de esta población desempleada

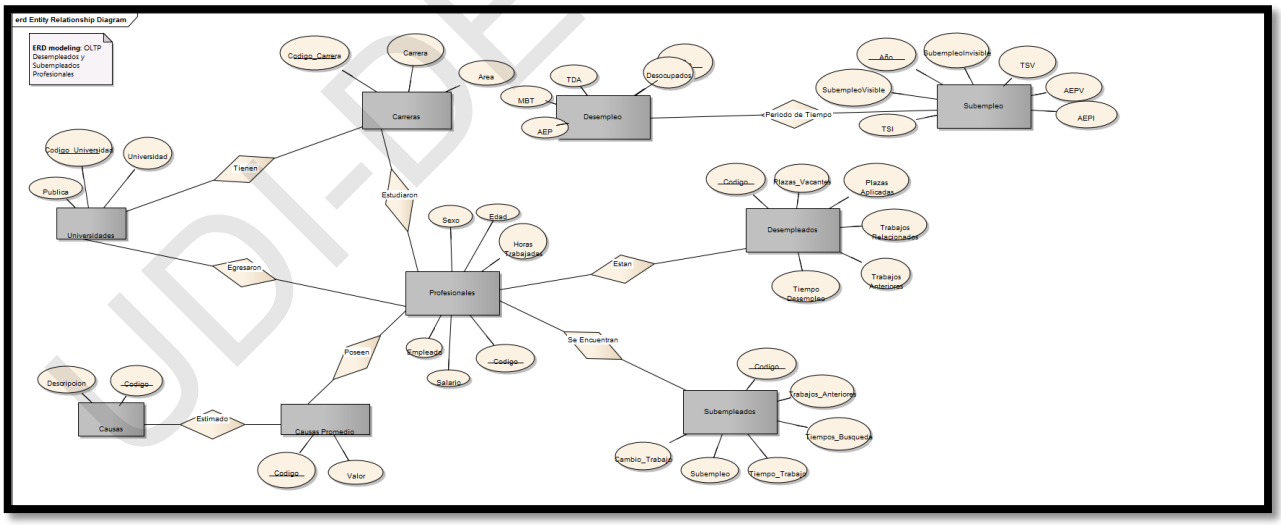
el 13.5% son personas con nivel de educación superior, es decir que ya egresaron de la universidad. A su vez de la población que se encuentra laborando el 4.29% corresponden a subempleados profesionales.

Debido a que esta población es un grupo representativo de los egresados de las distintas universidades a nivel nacional, es necesario poder estimar las principales causas que provocan esta situación, y conocer de qué forma este porcentaje se verá incrementado en la medida en la que no se tomen acciones para reducir tanto el desempleo como subempleo profesional. A través del estudio y la estimación de las causas y tasas de desempleo y subempleo profesional, se permitirá brindar un escenario futuro para que las autoridades a nivel nacional y de cada universidad puedan emprender las medidas necesarias para realizar cambios y ajustes en la situación académica y laboral.

8.2 Diseño y creación de la base de datos OLTP

8.2.1 Diagrama Entidad Relación

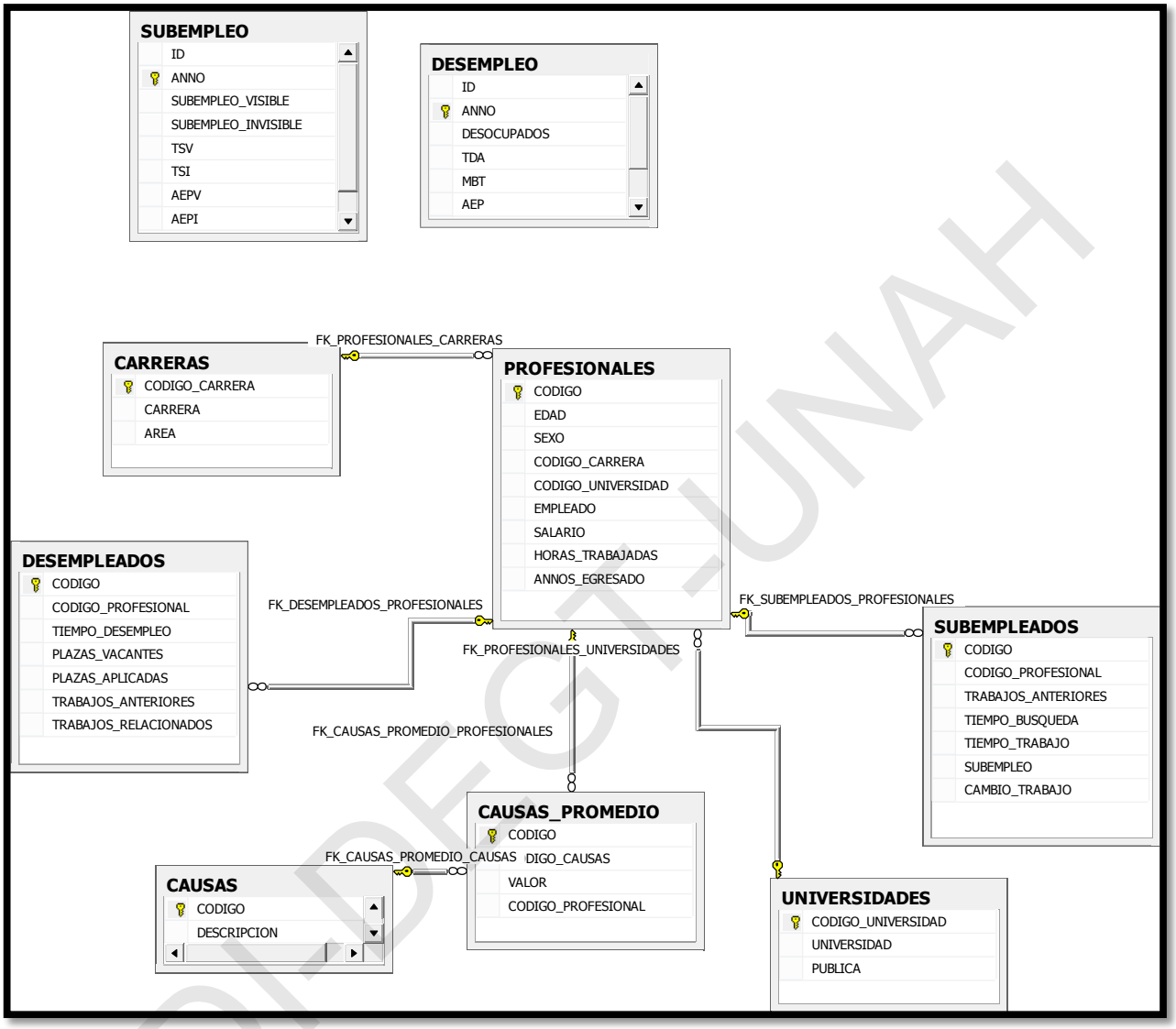
Figura 14: Diagrama Entidad-Relación OLTP.



Fuente Elaboración propia

8.2.2 Diagrama de BD

Figura 15: Diagrama de BD OLTP.



Fuente: Elaboración Propia.

8.3 Definición de los algoritmos de minería de datos a utilizar

Según lo estudiado en capítulos anteriores, se han seleccionado los siguientes algoritmos para aplicar la minería de datos

- *Arboles de decisión:*

Se seleccionó este algoritmo de clasificación y regresión, ya que permite tratar atributos discretos y continuos de forma más fácil, y crear relaciones en un conjunto de datos para predecir posibles estados. Específicamente, el algoritmo identifica las columnas de entrada que se correlacionan con la columna de predicción.

- *Algoritmos de Clústeres:*

  Este algoritmo se eligió porque permite la agrupación de un conjunto de casos con características similares y de esta forma crear predicciones de forma más fácil que mediante la simple observación de forma casual.

- *Algoritmos de serie temporal:*

  Este algoritmo es ideal para predecir tendencias basadas únicamente en el conjunto de datos original, adicional a esto es necesario que dentro del conjunto de datos se cuente con una columna de datos de tiempo, a través del cual se realizara la predicción.

- *Algoritmo Bayes Naive:*

  Este es otro algoritmo elegido por la simplicidad y rapidez de su aplicación, además permite realizar una exploración y asociación de los datos para poder realizar las predicciones de forma más fácil.

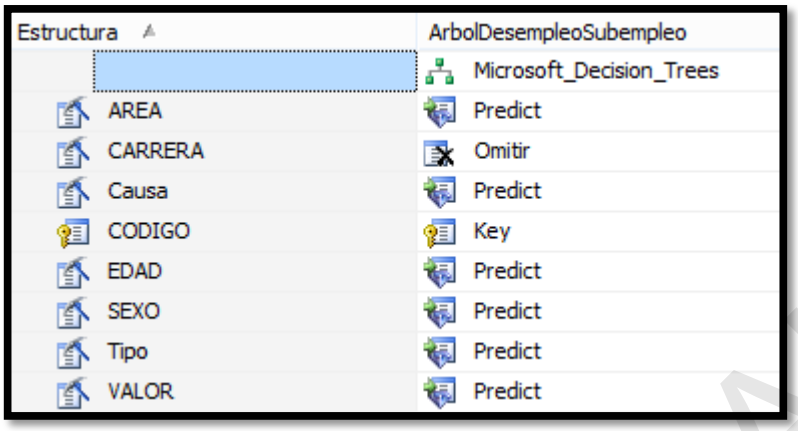
8.4 Aplicación de los algoritmos de minería de datos

A continuación se describen los resultados de la aplicación de cada uno de los algoritmos descritos anteriormente, y utilizando los datos obtenidos y tabulados por medio de la encuesta aplicada.

8.4.1 Arboles de Arboles de decisión:

Para este caso se utilizó la vista **CausasProblemasLaborales** y se definieron los siguientes atributos en la estructura:

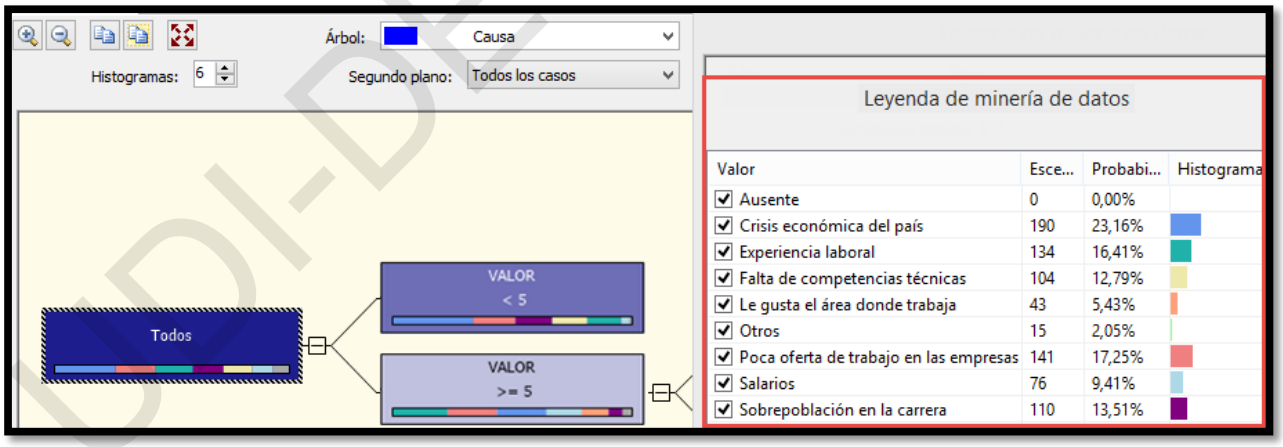
Figura 16: Estructura Árbol de Decisión.



Fuente Elaboración Propia.

En este caso todas las columnas son valores discretos que se usan tanto de entrada como de predicción, con el objetivo de poder crear relaciones y predicciones entre los diferentes datos, en el caso de la presente investigación, este es el resultado obtenido para predecir las posibles causas de desempleo y subempleo profesional:

Figura 17: Resultado Algoritmo Árbol de Decisión.



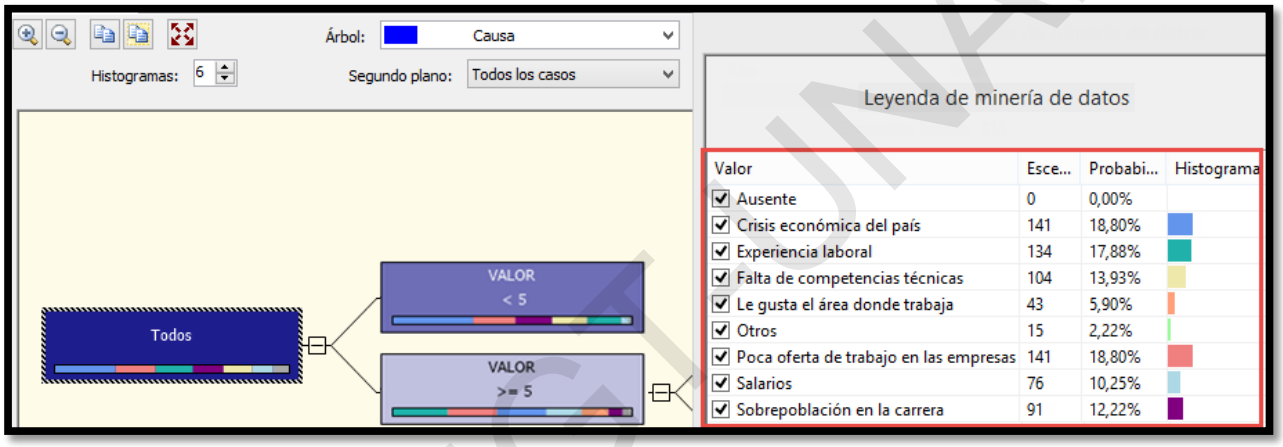
Fuente Elaboración Propia.

La leyenda mostrada a la derecha de la imagen en el recuadro rojo es la probabilidad de aparición de las diferentes causas, como se puede ver la probabilidad de aparición de la causa “Crisis

económica en el país" es del 23.16%, "Falta de competencias técnicas" 12.79% y así sucesivamente puede observarse la probabilidad para cada causa.

Al lado izquierdo de la imagen se muestra el gráfico de los resultados, y las diferentes asociaciones, en este caso se muestran la relación entre las causas y el valor que los encuestados atribuían a cada causa, el cual se definió entre los valores de 1 a 5, considerándose 1 menor importancia y 5 mayor importancia. Con esto se puede también obtener la siguiente relación:

Figura 18: Algoritmo de Árbol de Decisión. Análisis de causas con mayor prioridad

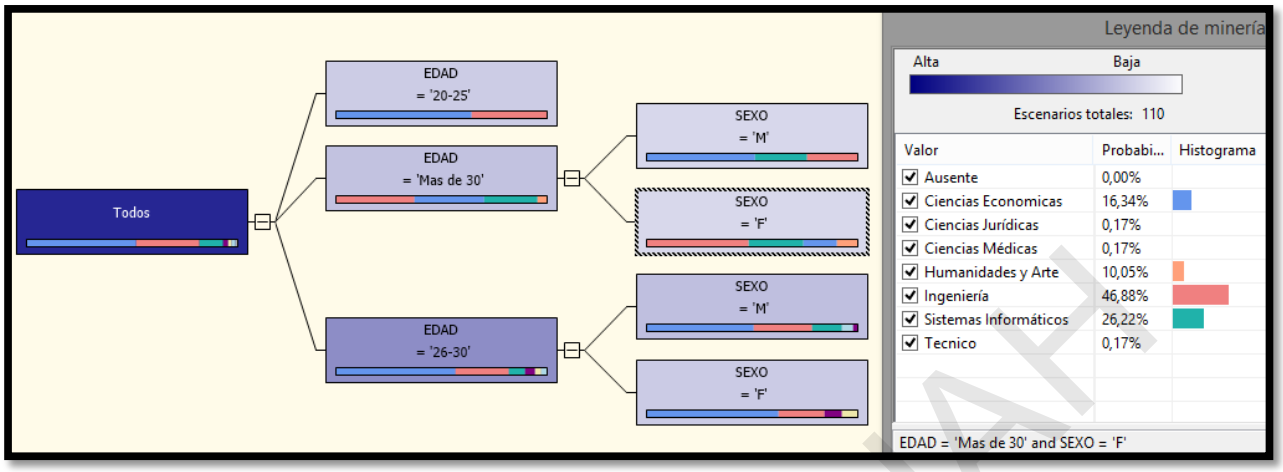


Fuente Elaboración Propia.

En este análisis se puede observar que la Crisis económica del país tiene una probabilidad del 18.80% según la importancia que los encuestados le dieron a la misma.

Otra facilidad de este algoritmo es realizar predicciones de las relaciones de los datos, esto se muestra en la siguiente figura, por ejemplo existe un 16.34% de probabilidad de que los profesionales universitarios con problemas laborales de más de 30 años y de sexo femenino sean del área de ciencias económicas.

Figura 19: Algoritmo de Árbol de Decisión. Predicciones con relaciones.



Fuente Elaboración Propia.

Este algoritmo nos permite visualizar la relación entre los diferentes atributos, para que posteriormente se puedan realizar operaciones para realizar predicciones, por ejemplo si deseamos obtener la probabilidad del valor mínimo de cada una de las causas, podemos realizar la siguiente consulta mediante funciones de predicción:

Figura 20: Algoritmo de Árbol de Decisión. Función de Predicción.

```
SELECT
  [ArbolDesempleoSubempleo].[VALOR],
  PredictHistogram([ArbolDesempleoSubempleo].[Causa]) as Results
From
  [ArbolDesempleoSubempleo]
NATURAL PREDICTION JOIN
(SELECT 1 AS [VALOR]) as t
```

Results

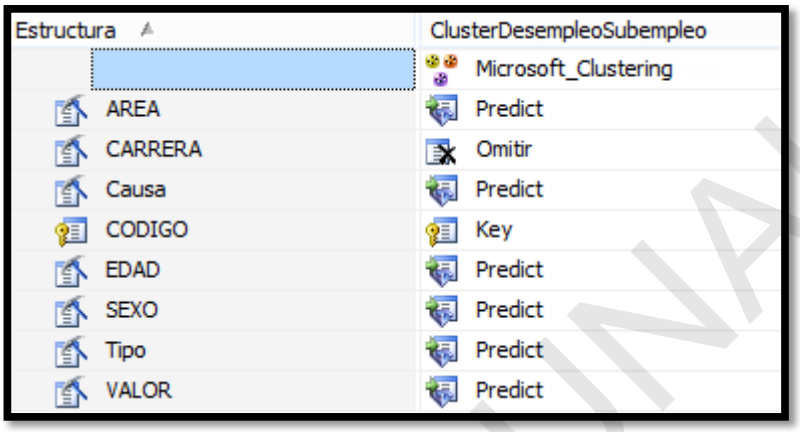
Causa	$SUPPORT	$PROBABILITY
Poca oferta de trabajo en las empresas	31	0,308047493403694
Experiencia laboral	22	0,218997361477573
Falta de competencias técnicas	22	0,218997361477573
Salarios	14	0,139841688654354
Le gusta el área donde trabaja	9	0,0903693931398417
Otros	2	0,0211081794195251
Sobrepoblación en la carrera	0	0,00131926121372032
Crisis económica del país	0	0,00131926121372032

Fuente Elaboración Propia.

8.4.2 Algoritmos de Clústeres:

Para este algoritmo se definió la siguiente estructura:

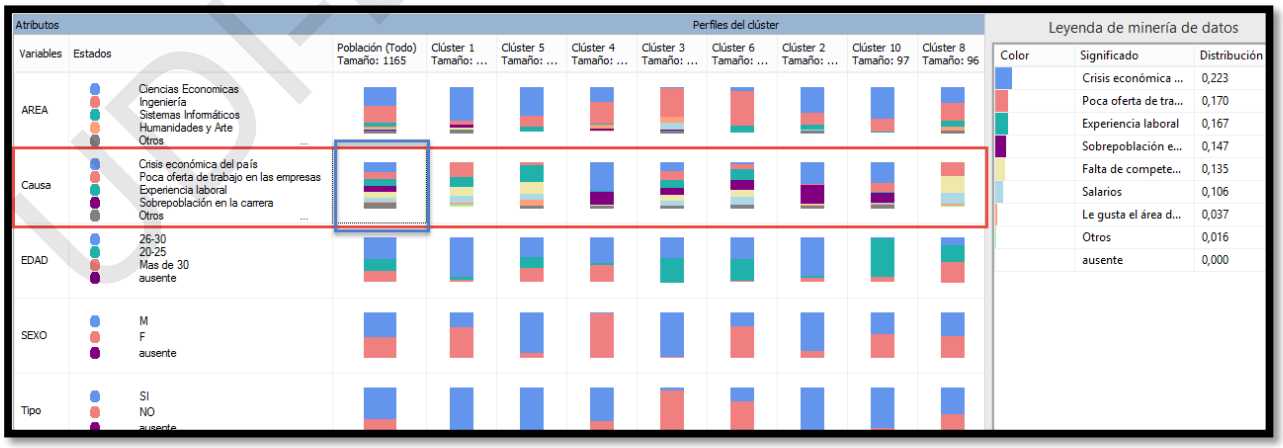
Figura 21: Estructura algoritmo de Clústeres de Microsoft.



Fuente Elaboración Propia.

Una particularidad de este algoritmo es que no se necesitan definir columnas de predicción, ya que se crea una asociación y predicción en cada uno de los nodos o clúster que se generan. Por medio de este algoritmo podemos visualizar la predicción de las causas en cada uno de los clúster, como se muestra en la siguiente figura:

Figura 22: Predicción por algoritmo de Clúster.



Fuente Elaboración Propia.

En el clúster marcado en azul, se puede apreciar la predicción total de las causas con la leyenda a la derecha, en donde la "crisis económica del país" tiene una probabilidad de 22.3%, a la vez se pueden analizar cada uno de los diferentes clúster que se presentan y observar las predicciones en cada uno de estos, por ejemplo si analizamos el clúster 6, marcado en rojo en la siguiente figura, podemos observar a los desempleados del área de Ingeniería, con edades entre los 20 y 25 años, en donde la probabilidad de que la causa de su desempleo sea "Poca oferta de trabajo en las empresas es de 12.3%.

Figura 23: Algoritmo de Clúster. Predicciones por Clúster.

Fuente Elaboración Propia.

Mediante funciones de predicción también podemos realizar consultas en este algoritmo, por ejemplo si deseamos saber cuál es la causa más común para que los profesionales de Ciencias Jurídicas estén desempleados, podemos usar la consulta que se muestra en la siguiente figura, la cual nos mostrara que la probabilidad es del 20%

Figura 24: Algoritmo de Clúster. Función de Predicción.

```
SELECT
  [ClusterDesempleoSubempleo].[Causa],
  PredictProbability([Causa]) as Results
From
  [ClusterDesempleoSubempleo]
NATURAL PREDICTION JOIN
(SELECT 'NO' AS [Tipo],
'Ciencias Jurídicas' AS [Area]) as t
```

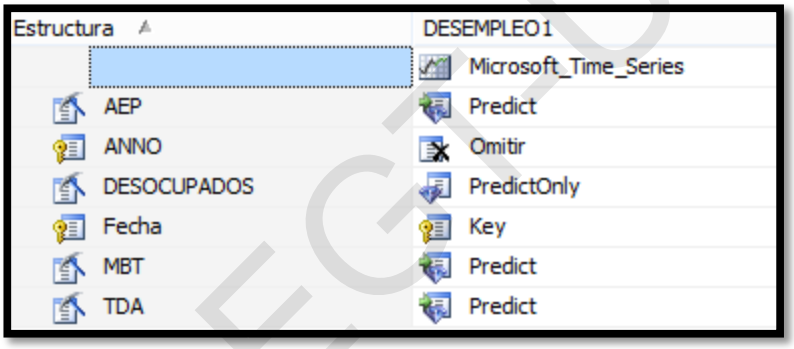
Causa	Results
Crisis económica del país	0,2025035228...

Fuente Elaboración Propia.

8.4.3 Algoritmos de Serie Temporal:

Para este algoritmo es necesario contar con una clave de tiempo, que es a partir de la cual se realizara la predicción, para esto se definió la siguiente estructura:

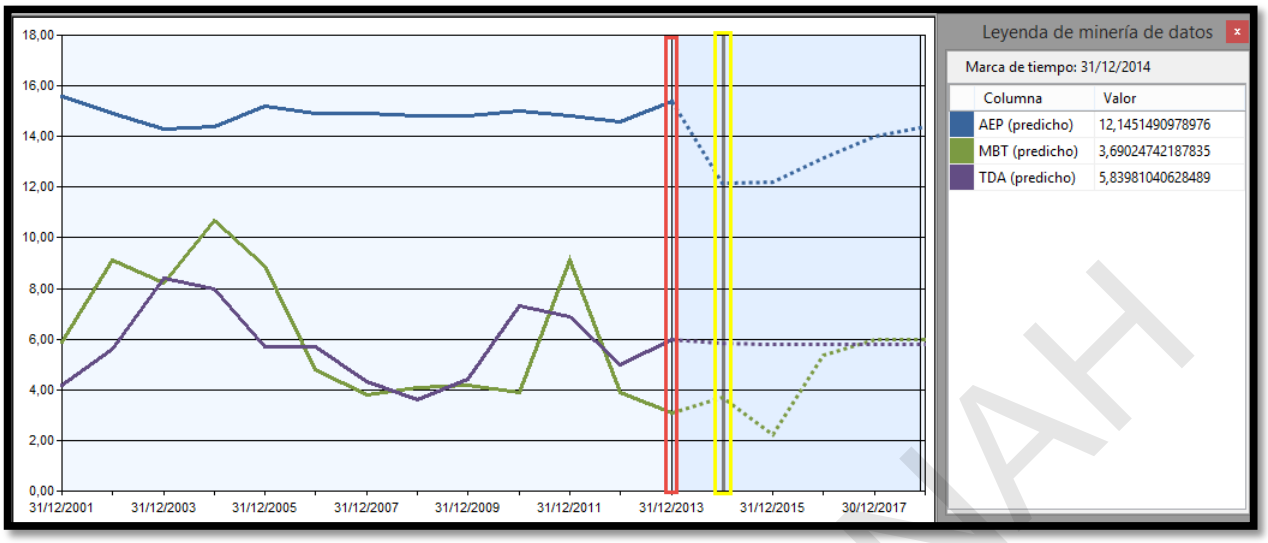
Figura 25: Estructura Serie Temporal



Fuente Elaboración Propia.

La columna de tiempo son diferencias dadas en año, que es la información tabulada y brindada por el INE según las encuestas de hogares realizadas anualmente, a partir de eso se realiza la predicción de la cantidad de profesionales desempleados, la TDA (Tasa de desempleo abierto), AEP (años de estudio promedio) y MBT (meses de búsqueda de trabajo), con lo cual se obtiene lo siguiente:

Gráfico 2: Predicción Serie Temporal

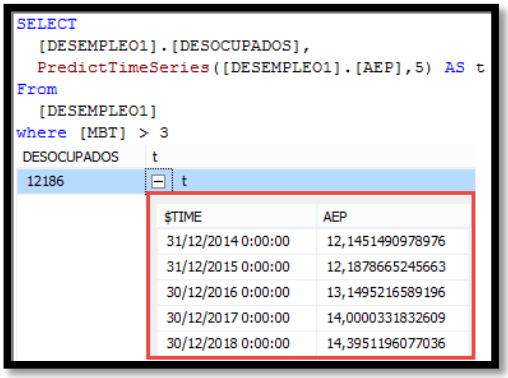


Fuente Elaboración Propia.

En la figura anterior podemos visualizar que de la línea roja hacia la izquierda se nos muestra la información real de los datos del año 2013 y anteriores. El lado derecho de la línea roja es la predicción del algoritmo con el paso de los años. La línea marcada en amarillo es lo que nos interesa conocer, en este caso es la predicción al año 2014, según la leyenda para el 2014 se estima que existe una TDA de 5.84%.

Si bien este algoritmo nos presenta la predicción sin necesidad de tener muchas entradas y relaciones adicionales al periodo de tiempo, también es posible establecer predicciones basadas en relaciones por medio de funciones, como se muestra en la siguiente figura:

Figura 26: Serie Temporal. Función de Predicción.



```
SELECT
  [DESEMPLEO1].[DESOCUPADOS],
  PredictTimeSeries([DESEMPLEO1].[AEP],5) AS t
From
  [DESEMPLEO1]
where [MBT] > 3
```

DESOCUPADOS	t
12186	t

$TIME	AEP
31/12/2014 0:00:00	12,1451490978976
31/12/2015 0:00:00	12,1878665245663
30/12/2016 0:00:00	13,1495216589196
30/12/2017 0:00:00	14,0000331832609
30/12/2018 0:00:00	14,3951196077036

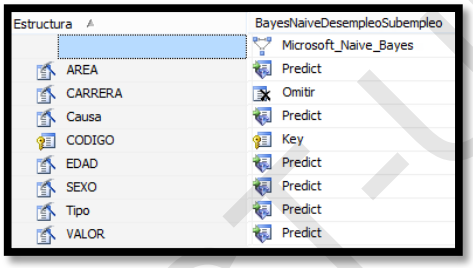
Fuente Elaboración Propia.

En la función se intenta determinar el estimado de años de estudio promedio, basados en la cantidad de desocupados y en donde los meses de búsqueda de trabajo no sean mayores a 3, así se obtiene que para el 2015 se estima que sean 12.19 años de estudio promedio, el último dato proporcionado por el INE en el 2013 son 15.4 años.

8.4.4 Algoritmo Bayes Naive:

Para aplica este algoritmo de relación se definió la estructura mostrada en la siguiente figura, en donde se intenta predecir las causas y áreas donde se produce el desempleo y subempleo profesional.

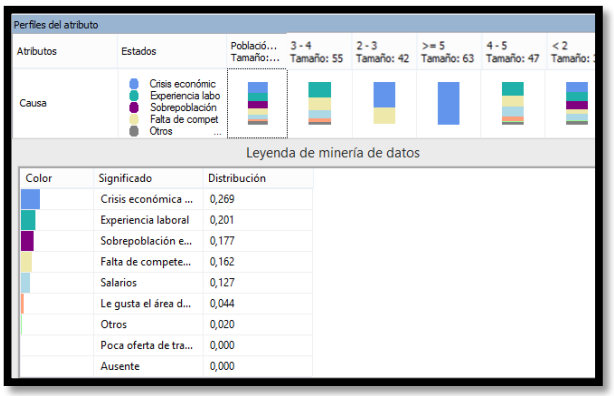
Figura 27: Estructura Algoritmo Bayes Naive.



Fuente Elaboración Propia.

Al aplicar este algoritmo se obtiene la distribución de la siguiente figura para predecir las causas del desempleo y subempleo profesional, aquí se observa que la causa más probable es "Crisis económica del país", además de esto se puede visualizar dependiendo del valor que los encuestados le dan a cada causa en los diferentes nodos a la derecha de la figura.

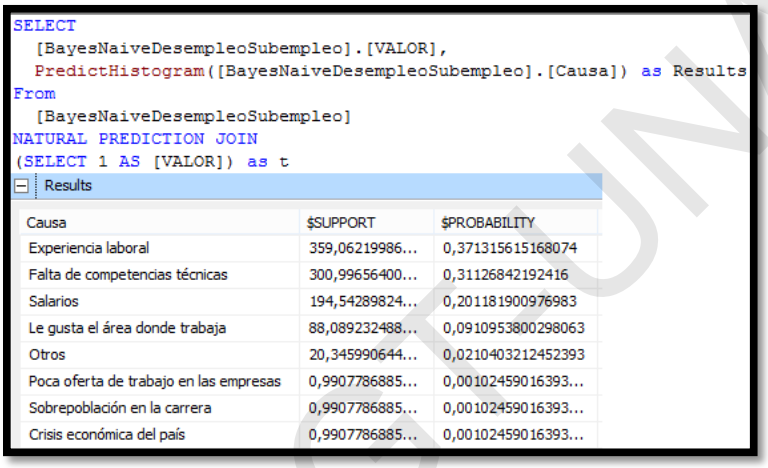
Figura 28: Predicción Bayes Naive.



Fuente Elaboración Propia.

Con este algoritmo también podemos aplicar funciones de predicción como se muestra en la siguiente figura, en la figura se está obteniendo la probabilidad de ocurrencia de cada una de las causas tanto del desempleo como subempleo. Como se puede apreciar en la figura la causa con mayor probabilidad de ocurrencia en la problemática laboral es la crisis económica del país, seguido por la falta de experiencia laboral; estos resultados coinciden con los obtenidos al aplicar el algoritmo directamente, sin necesidad de utilizar funciones de predicción de forma manual.

Figura 29: Bayes Naive. Función de Predicción.



```
SELECT
  [BayesNaiveDesempleoSubempleo].[VALOR],
  PredictHistogram([BayesNaiveDesempleoSubempleo].[Causa]) as Results
From
  [BayesNaiveDesempleoSubempleo]
NATURAL PREDICTION JOIN
(SELECT 1 AS [VALOR]) as t
```

Results

Causa	$SUPPORT	$PROBABILITY
Experiencia laboral	359,06219986...	0,371315615168074
Falta de competencias técnicas	300,99656400...	0,31126842192416
Salarios	194,54289824...	0,201181900976983
Le gusta el área donde trabaja	88,089232488...	0,0910953800298063
Otros	20,345990644...	0,0210403212452393
Poca oferta de trabajo en las empresas	0,9907786885...	0,00102459016393...
Sobrepoblación en la carrera	0,9907786885...	0,00102459016393...
Crisis económica del país	0,9907786885...	0,00102459016393...

Fuente Elaboración Propia.

8.5 Selección del Algoritmo de Minería de Datos

Una vez finalizada la ejecución y aplicación de los algoritmos de árboles de decisión, algoritmo de clústeres, serie temporal y bayes naive, se eligieron 2 que permitirán resolver la problemática planteada, un algoritmo para predecir el desempleo y subempleo profesional en años futuros; y otro algoritmo para determinar las principales causas del desempleo y subempleo.

a) Para la poder predecir el estimado de desempleo y subempleo profesional, se utilizara el algoritmo de Serie Temporal o Serie de Tiempo, debido a que este algoritmo como su nombre lo indica permite realizar las predicciones basándose en un periodo de tiempo; este periodo de tiempo serán los años de información brindados por el INE y el dato a predecir será la tasa de desempleo y subempleo.

De los algoritmos aplicados este es el que más se ajusta a las necesidades, para predecir una tasa o un número determinado usando la variable tiempo, debido a que ese optimiza la previsión en el tiempo de valores continuos, como ser la tasa de desempleo o subempleo.

b) Para estimar las principales causas de desempleo y subempleo profesional, se pueden utilizar cualquiera de los otros tres algoritmos estudiados, omitiendo el de serie temporal, ya que la información recolectada para obtener las causas fue obtenida en un solo periodo de tiempo, por lo cual no podríamos utilizar este algoritmo.

Como se aplicaron los tres algoritmos sin problemas, y con los tres fue posible realizar predicciones, cualquiera de los tres puede ser elegido para estimar las principales causas de desempleo y subempleo profesional. Para elegir cuál de los tres algoritmos es el ideal para realizar la estimación se utilizara el gráfico de precisión de minería de datos proporcionado por Analysis Services, el cual se muestra a continuación.

En este gráfico se puede apreciar la línea color verde, que es el modelo ideal del algoritmo, la línea amarilla representa el algoritmo de cluster de Microsoft, la línea morada es el algoritmo de Bayes Naive, finalmente la línea rosada es el algoritmo de árboles de decisión.

La columna “Población correcta” nos indica el porcentaje de población que se puede predecir con cada uno de los algoritmos. Y la columna “Probabilidad de Predicción” nos indica el porcentaje de población que se puede predecir con cada algoritmo.

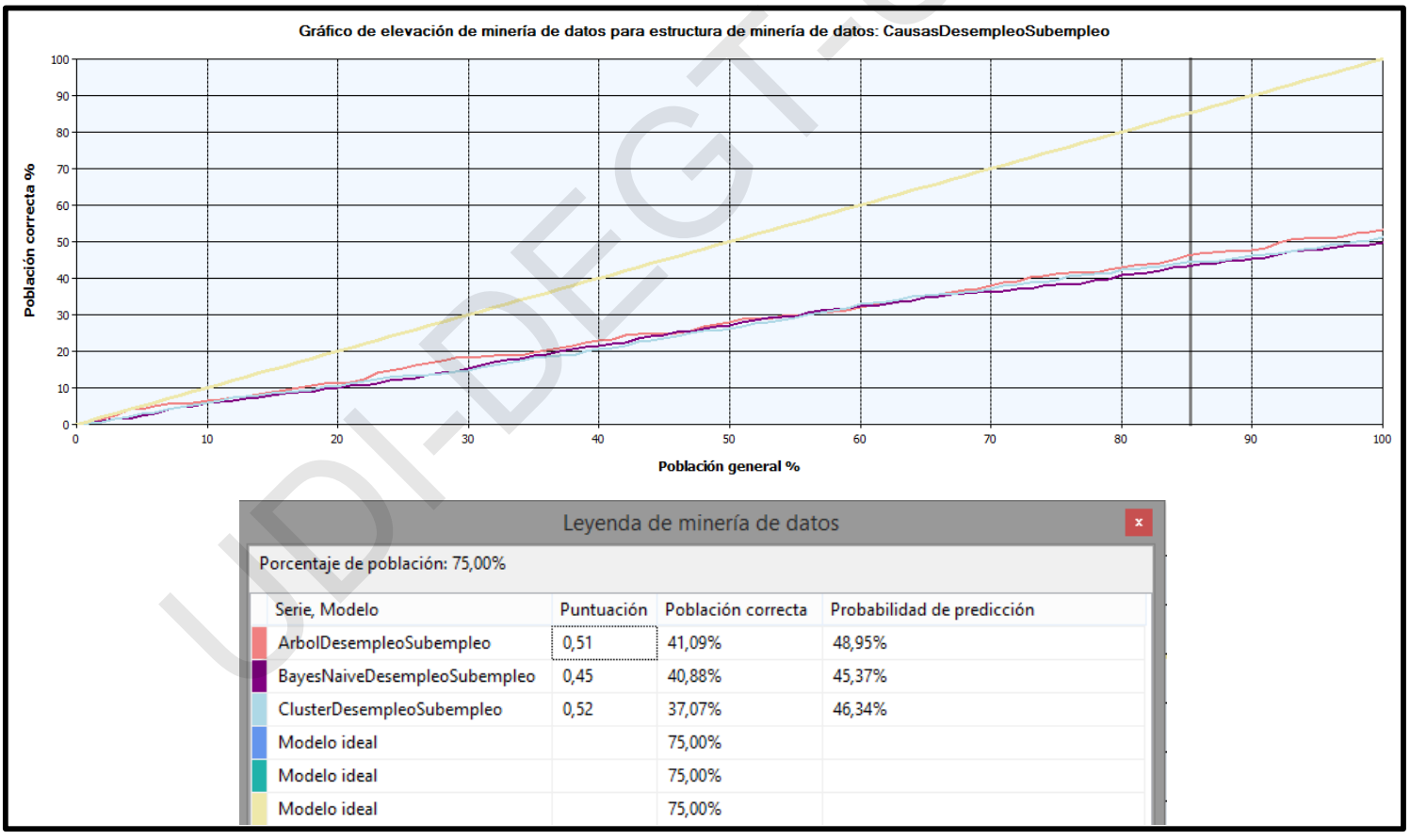
Finalmente el algoritmo que más se aproxime al modelo ideal del algoritmo (la línea color verde) será el algoritmo seleccionado para realizar la predicción, ya que es el que contara con el mayor porcentaje de población que podrá predecirse.

En el gráfico se puede observar que tanto la población, como el puntaje y la probabilidad son mayores para el algoritmo de árbol de decisión, con un 41.09%, 0.51 y 48.95% respectivamente. Con segunda mejor probabilidad de predicción se encuentra el algoritmo de Clúster con 46.34%, y por ultimo con una probabilidad de predicción de 46.34% el algoritmo de Naive Bayes. Con base a esta información y lo que se visualiza en la gráfica, el algoritmo a elegir es el de Arboles de Decisión.

Con respecto a la población que cada algoritmo es capaz de predecir, en el caso del algoritmo de Arboles de decisión que es el que hemos elegido, podemos predecir el 41% de la población; eso quiere decir que se podrá predecir cuales son las causas de desempleo y subempleo profesional del 41% de la población, si bien esto es inferior al 50%, es debido a la limitante de datos con la que se cuenta, a medida se cuente con una población mayor, el índice de la población capaz de predecir debe aumentar.

También se puede observar que la probabilidad de predicción de los tres modelos es inferior al 50%, esto nos da una idea de la limitante que se planteaba al inicio de esta investigación, ya que los datos con los que se cuentan son muy pocos, y la confiabilidad de las predicciones aumenta a medida se cuente con una muestra mucho más grande.

Gráfico 3: Gráfico de Elevación de Minería de Datos: Causas Desempleo y Subempleo.



Leyenda de minería de datos

Porcentaje de población: 75,00%

Serie, Modelo	Puntuación	Población correcta	Probabilidad de predicción
ArbolDesempleoSubempleo	0,51	41,09%	48,95%
BayesNaiveDesempleoSubempleo	0,45	40,88%	45,37%
ClusterDesempleoSubempleo	0,52	37,07%	46,34%
Modelo ideal		75,00%	
Modelo ideal		75,00%	
Modelo ideal		75,00%	

Fuente Elaboración Propia.

8.6 Análisis de Resultados

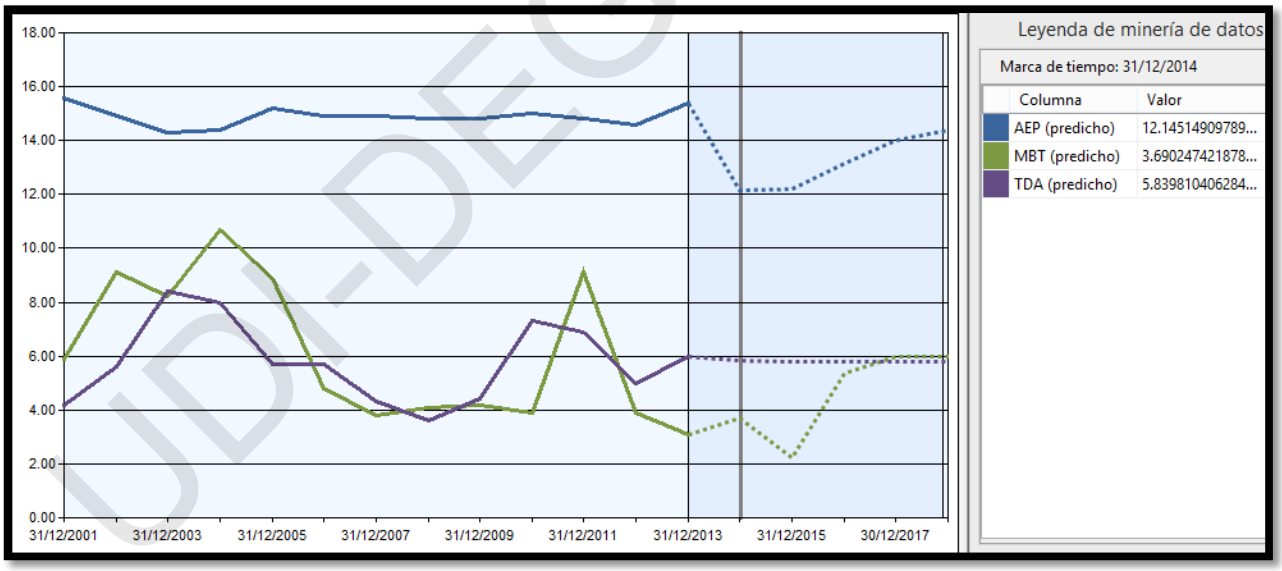
Según la aplicación de los algoritmos elegidos, se han obtenido los siguientes resultados:

8.6.1 Estimación de la Tasa de Desempleo y Subempleo Profesional:

Como se muestra en el siguiente gráfico, la tasa estimada de desempleo para el 2014 es de 5.84%, así mismo los meses de búsqueda de trabajo son 3.7 y los años de estudio promedio de los desempleados son 12.15 meses.

Como se puede observar en la gráfica la tasa de desempleo no presenta una reducción con el paso de los años, en cambio los meses de búsqueda de trabajo, presenta un cambio destacado, tal como se ha venido observando en los últimos años. Por ejemplo en el año 2011 los meses de búsqueda de trabajo aumentaron drásticamente, para el año 2015 en la predicción, la tasa de desempleo disminuye, y los meses de búsqueda de trabajo también disminuyen, estas dos variables se observan que están muy relacionadas.

Gráfico 4: Tasa estimada de Desempleo Profesional 2014.



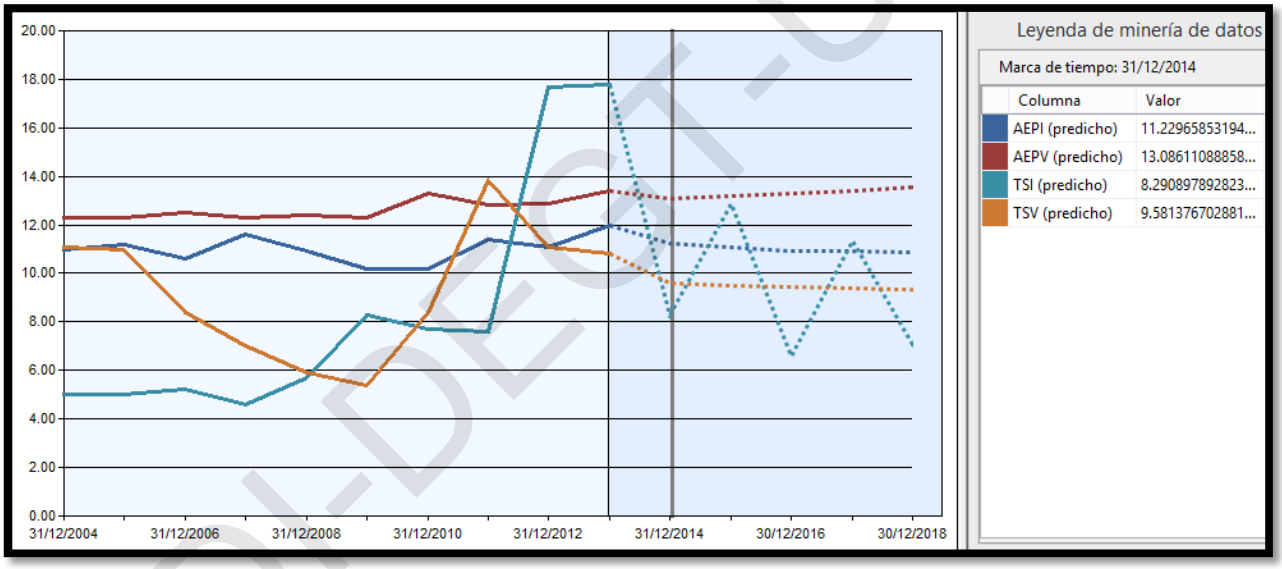
Fuente Elaboración Propia.

Para el caso de subempleo profesional, este se divide en subempleo visible e invisible, como se muestra en el gráfico a continuación, la tasa de subempleo profesional visible para el 2014 es de

9.58% y del subempleo profesional invisible es de 8.29%. Los años de estudio promedio son de 11.23 y 13.09 para el subempleo profesional invisible y visible respectivamente.

En la gráfica a continuación se puede observar que la tasa de subempleo visible es mayor a la invisible, esto puede deberse a que los profesionales no consiguen un trabajo con muchas horas de trabajo, y por lo tanto cuentan con un sueldo menor y no logran cubrir sus necesidades y las de su familia, lo cual se relaciona con la principal causas del subempleo y desempleo profesional que es la crisis económica del país.

Gráfico 5: Tasa estimada de Subempleo Profesional 2014.



Fuente Elaboración Propia.

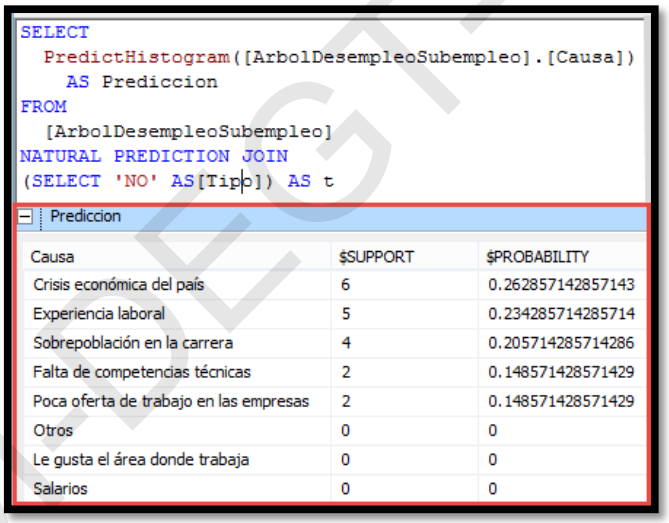
8.6.2 Estimación de las Principales Causas de Desempleo y Subempleo Profesional:

Como se muestra en la siguiente figura, la principal causa de desempleo profesional es la crisis económica del país, con un 26% de probabilidad, la segunda causa más importante es la experiencia laboral con la que cuentan los profesionales, seguida de la sobrepoblación de la carrera con un 21%, las causa de menor importancia según el modelo son la falta de competencias técnicas y la poca oferta de trabajo en las empresa, ambas con un 14.8% de probabilidad.

Las causas de menor importancia pueden estar relacionadas entre sí al tener la misma probabilidad de aparición, y esto puede deberse a que bien las empresas están exigiendo contratar profesionales con diferentes competencias técnicas a las que estos poseen y esto puede llevar a los profesionales a juzgar como pocas ofertad de trabajo en las empresas. Otro punto importante es que muchas empresas se dedican a contratar a personas sin contar con un título universitario para no tener que pagar sueldos muy altos.

Las principales causas del subempleo profesional según el algoritmo de predicción son compartidas con el desempleo, estas causas son la crisis económica del país con un 28% y la experiencia laboral con un 18%. La crisis económica del país, la pobreza, la necesidad de ayudar con los ingresos de la familia, son la principal causa por la cual los profesionales aceptan un trabajo sin relación con su profesión y por la cual prefieren continuar con su trabajo antes que estar desempleados.

Figura 30: Causas Estimadas de Desempleo Profesional.



```
SELECT
  PredictHistogram([ArbolDesempleoSubempleo].[Causa])
    AS Prediccion
FROM
  [ArbolDesempleoSubempleo]
NATURAL PREDICTION JOIN
(SELECT 'NO' AS[Tipo]) AS t
```

Prediccion

Causa	$SUPPORT	$PROBABILITY
Crisis económica del país	6	0.262857142857143
Experiencia laboral	5	0.234285714285714
Sobrepoblación en la carrera	4	0.205714285714286
Falta de competencias técnicas	2	0.148571428571429
Poca oferta de trabajo en las empresas	2	0.148571428571429
Otros	0	0
Le gusta el área donde trabaja	0	0
Salarios	0	0

Fuente Elaboración Propia.

La causa de menor importancia son los salarios que reciben los profesionales, y la comodidad o preferencia que sienten con su área de trabajo, el detalle de estas causas de muestra en la figura a continuación. Esto lleva a pensar que los profesionales aceptan un trabajo sin relación son su profesión incluso con sueldos bajos.

Figura 31: Causas estimadas de Subempleo Profesional.

```
SELECT
  PredictHistogram([ArbolDesempleoSubempleo].[Causa])
    AS Prediccion
FROM
  [ArbolDesempleoSubempleo]
NATURAL PREDICTION JOIN
(SELECT 'SI' AS[Tipo]) AS t
```

Prediccion

Causa	$SUPPORT	$PROBABILITY
Crisis económica del país	79	0.283216783216783
Experiencia laboral	50	0.181818181818182
Poca oferta de trabajo en las empresas	46	0.167832167832168
Sobrepoblación en la carrera	34	0.125874125874126
Falta de competencias técnicas	31	0.115384615384615
Le gusta el área donde trabaja	19	0.0734265734265734
Salarios	9	0.0384615384615385
Otros	2	0.013986013986014

Fuente Elaboración Propia.

CAPITULO IX: ANALISIS DE RESULTADOS

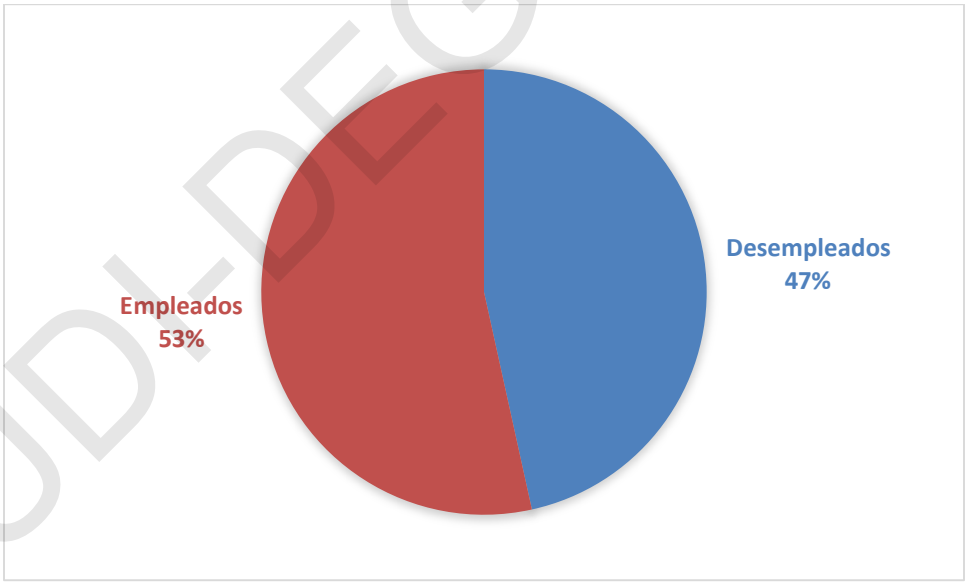
9.1 Análisis de los Datos

A continuación se presentan los resultados obtenidos en función de los objetivos de la investigación, derivados del procesamiento de datos a través del procesamiento de los cubos y de la minería de datos, el detalle de estos resultados se presentan mediante gráficos divididos entre el desempleo y subempleo profesional:

Según la muestra de la población analizada, el 53% de los encuestados corresponden a personas con subempleo profesional y el 47% son desempleados universitarios, como se muestra en el gráfico 6.

En la muestra de desempleados el 60% corresponden a hombres y el 40% a mujeres, estos fueron seleccionados al azar, de igual forma que la muestra de profesionales con problemas de subempleo en donde el 43% son mujeres y el 57% hombres.

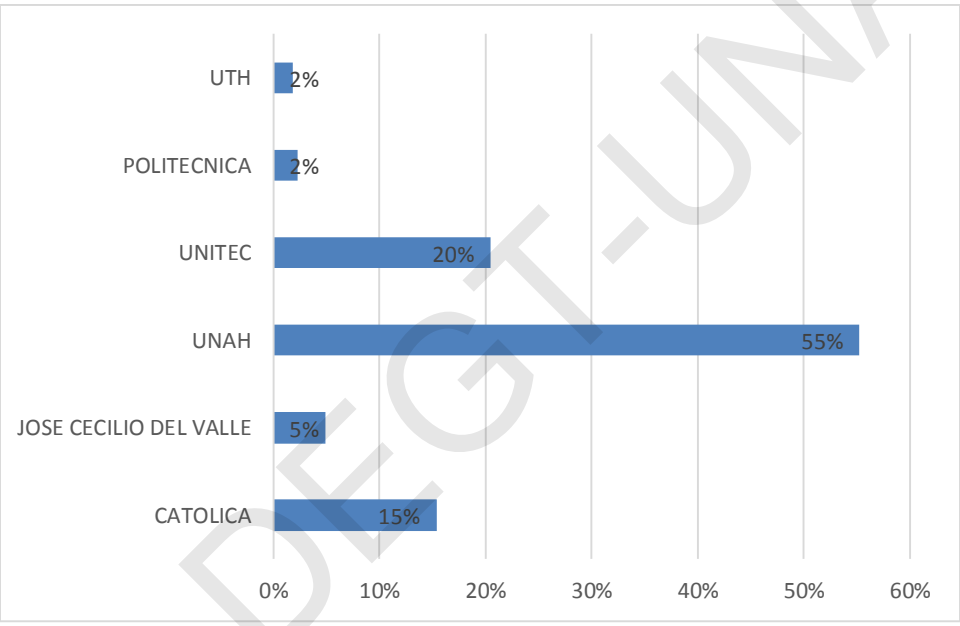
Gráfico 6: Situación Laboral.



Fuente Elaboración Propia.

De esta muestra analizada, más del 50% corresponden a profesionales egresados de la UNAH y el 49% son profesionales de universidades privadas, tal como se muestra en el gráfico a continuación, a pesar de que la muestra fue seleccionada al azar, la mayoría de las personas son egresadas de la UNAH, esto se debe a que según educación superior, la mayor parte de la población universitaria estudian en la UNAH, y el mayor porcentaje de egresados universitarios son de la UNAH.

Gráfico 7: Muestra por Universidad



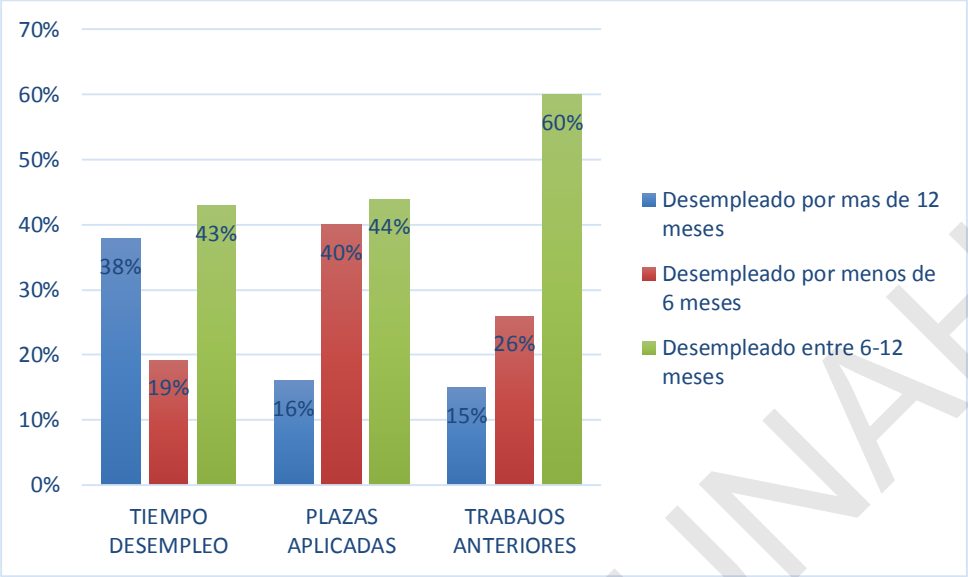
Fuente Elaboración Propia.

9.1.1 Desempleo

De las personas desempleadas únicamente el 16% de ellos no ha aplicado a ninguna plaza u oportunidad de trabajo en los últimos 12 meses, el otro 84% ha aplicado a lo más a 5 plazas en el último año.

El 26% de los desempleados nunca han tenido un trabajo, de estos el 60% son personas que obtuvieron su título universitario hace más de 6 meses, y únicamente el 15% de las personas que no han tenido un trabajo son desempleados desde hace más de 12 meses.

Gráfico 8: Información Desempleados



Fuente Elaboración Propia.

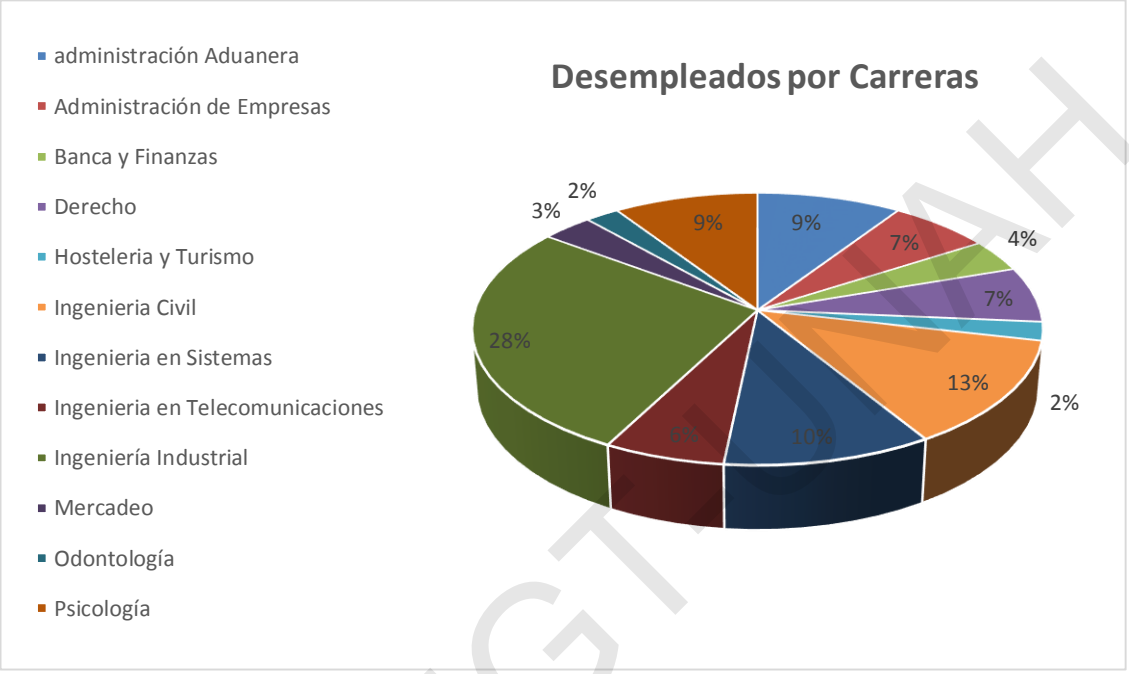
El 16% de las personas que no ha aplicado a ninguna plaza son desempleados por más de 12 meses, por lo cual se podría pensar que son personas que han caído en la situación de desalentados, ya que no han aplicado a ninguna plaza y no han encontrado trabajo, al contrario de las personas desempleadas por menos de 6 meses, que se encuentran aplicando a varias plazas.

La mayor cantidad de personas desempleadas (43%) tienen al menos 6 meses de no encontrarse laborando, comparado con el 19% de personas desempleadas que no están laborando hace menos de 6 meses, es decir que la mayor cantidad de desempleados tienen mucho tiempo de no estar trabajando. El 24% de las personas desempleadas que han tenido trabajos anteriormente, no tenían un trabajo relacionado con su profesión universitaria, es decir que estas personas eran subempleados profesionales.

El 28% de los desempleados egresaron de la carrera de Ingeniería Industrial, como se muestra en el siguiente gráfico, y la mayoría de ellos tiene entre 26 y 30 años además todos tienen menos de 30 años, el 88%. La menor cantidad de desempleados corresponden a personas egresadas de la

carrera de Odontología, y todos ellos son hombres y únicamente el 25% de ellos no ha tenido trabajos anteriormente y no tienen más de un año de estar desempleados.

Gráfico 9: Desempleados por Carreras

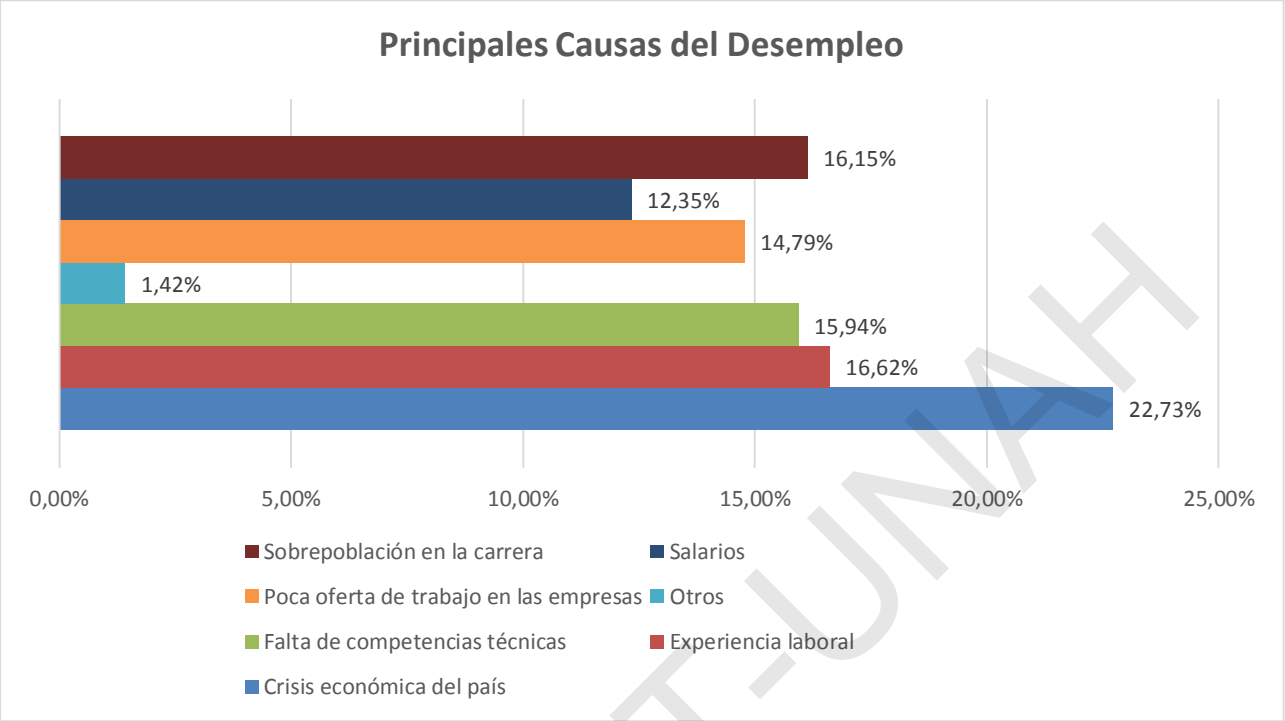


Fuente Elaboración Propia.

Según los encuestados, la principal causa por las cual se encuentran desempleados es por la crisis económica del país con 22.73% como se muestra en la gráfica, esto incluye los problemas económicos familiares que pudieron tener por la crisis económica del país.

La falta de experiencia laboral y la sobrepoblación de la carrera son otras dos causas importantes que según los profesionales evitan que puedan conseguir trabajo, esto se ve reflejado en que el 75% de los desempleados cuentan con experiencia laboral, y el 33% de estos encuestados tienen un promedio de 1-3 años de experiencia. Los salarios bajos son la causa de menor importancia para los desempleados, eso quiere decir que ellos aceptarían tener un trabajo con salarios bajos.

Gráfico 10: Principales Causas del Desempleo de Profesionales Universitarios.



Fuente Elaboración Propia.

El 68% de los desempleados afirmo que aceptaría trabajar en algo sin relación con su carrera, y la principal causa de realizar esta afirmación, es la necesidad económica con la que cuentan los desempleados en conjunto con la crisis económica del país, esto corresponde a 24%, esto está muy ligado a la principal causa de desempleo.

La segunda causa por la cual los desempleados aceptarían un trabajo sin relación con su carrera es la falta de experiencia laboral con un 16% al igual que la sobrepoblación de la carrera, los desempleados afirman que existen demasiadas personas laborando en su misma profesión, lo cual no les permite encontrar un trabajo relacionado con su carrera, y están dispuestos a aplicar a otro tipo de trabajo. De la misma forma están dispuestos a laborar en un trabajo sin relación con su carrera para adquirir experiencia laboral y que esto les permita aplicar a un mejor trabajo, aunque no sea necesariamente relacionado con su carrera.

La causa con menor ocurrencia son los salarios, los desempleados aceptarían un trabajo sin relación a su carrera por obtener un mayor salario, y muchos de ellos afirman que se obtienen mejores salarios laborando en un trabajo sin relación con su carrera.

Gráfico 11: Causas para aceptar un subempleo profesional.

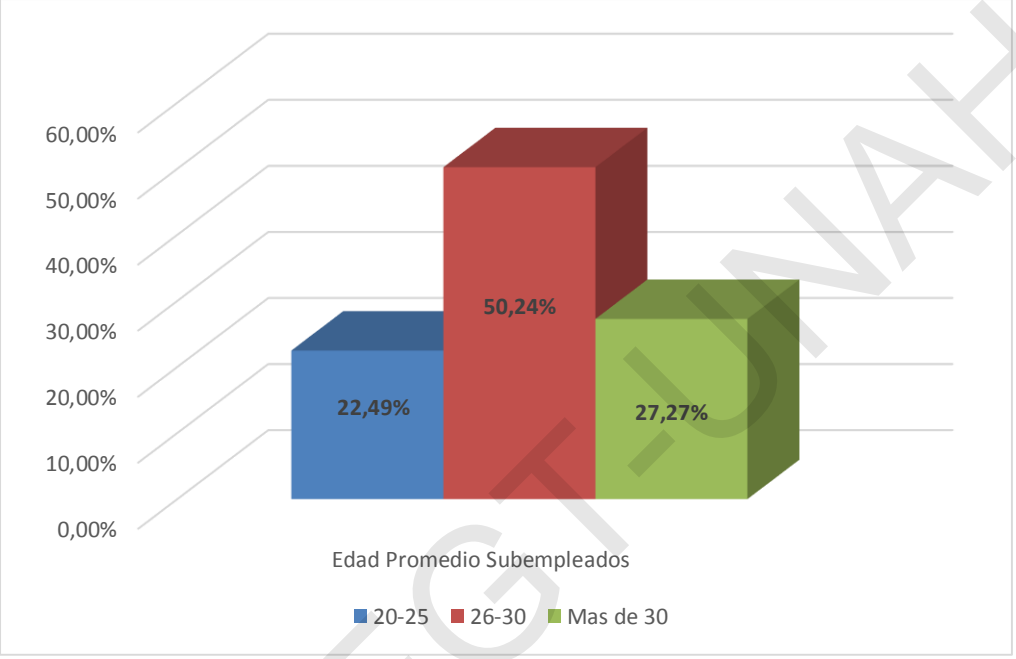
Fuente Elaboración Propia.

9.1.2 Subempleo Profesional

El 51% de los subempleados profesionales encuestados reciben un salario entre L.15,000 y L.25,000, el cual es superior al mínimo y únicamente el 3% recibe un salario inferior al salario mínimo, sin embargo ese 3% que reciben un salario tan bajo laboran más de 36 horas a la semana, por lo cual se puede deducir que este 3% son personas que son explotadas laboralmente, ya que no cuentan con un salario según la ley y además trabajan muchas horas en el día.

La muestra analizada de subempleados corresponden en su mayoría a personas entre los 26 y 30 años, esto representa el 50.24%, el 27.27% son subempleados mayores a 30 años y el resto son personas menores a 26 años, esto puede apreciarse en el gráfico a continuación.

Gráfico 12: Edad Promedio de los Subempleados Profesionales.

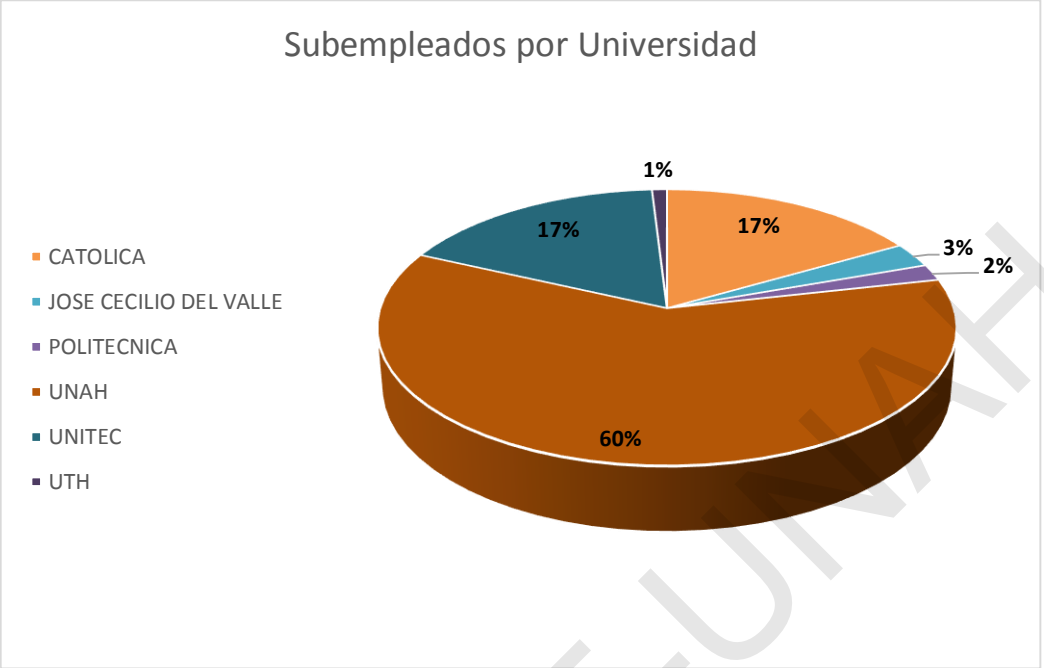


Fuente Elaboración Propia.

Únicamente un 2% de los subempleados trabajan menos de 36 horas, y ellos reciben un salario entre L.15,000 y L.25,000, esta dato es de mucho interés, ya que estas personas reciben un salario mayor al 3% de las personas que reciben menos de un salario mínimo.

Los subempleados que laboran menos de 36 horas a la semana, además egresaron de la universidad hace no más de 3 años, y son una muestra egresada equitativamente de la UNAH y Unitec (50% y 50%). El 60% de las personas encuestadas corresponden a subempleados profesionales egresados de la UNAH, la segunda universidad con mayor cantidad de subempleados según el muestreo es Unitec y la universidad Católica con un 17% cada una de ellas, esto puede verse de forma más clara en el gráfico a continuación.

Gráfico 13: Subempleados Profesionales por Universidad.

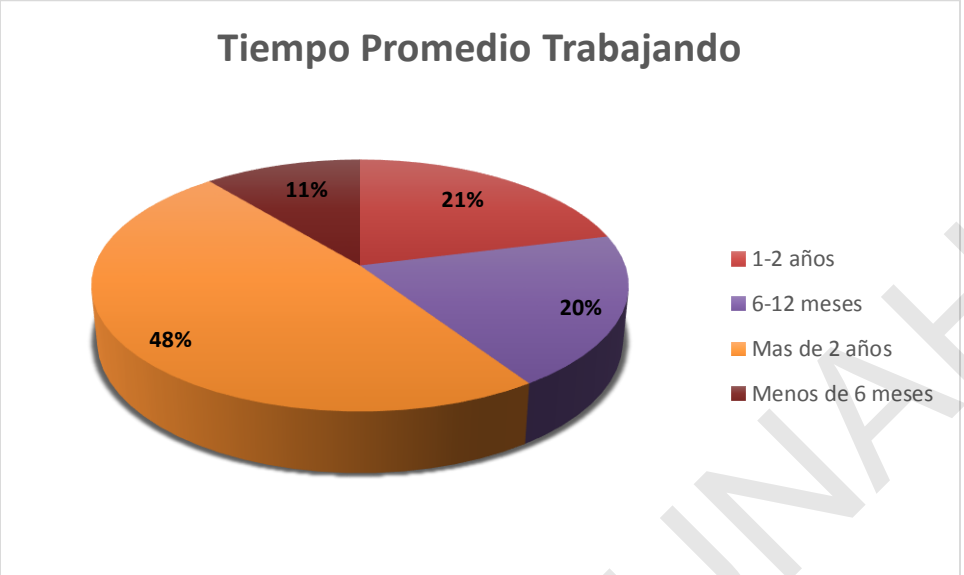


Fuente Elaboración Propia.

El 48% de los profesionales encuestados tienen más de 2 años de ejercer un trabajo distinto al de su profesión, y únicamente el 11% tiene menos de 6 meses laborando; como puede verse en el gráfico 14 a mayor detalle. En el caso de las personas que tienen más de dos años laborando son personas que ya se acostumbraron a un trabajo sin relación a su profesión y que probablemente continúen con el mismo durante más tiempo, o si cambian de trabajo sea algo relacionado con su trabajo actual.

Cuando se les consulto a los subempleados por qué tenían ese trabajo sin relación con su profesión, el 32% estableció que se debía a que estaban recibiendo un buen salario. Esto está muy relacionado con lo analizado anteriormente en la sección de desempleo, donde un 24% de los desempleados aceptarían un subempleo profesional un buen salario y únicamente un 16% lo hace por ganar experiencia laboral. Esto lleva a pensar que una vez que los profesionales están establecidos con un trabajo, y con un buen salario, no realizan esfuerzos por ejercer su profesión o buscar un nuevo trabajo.

Gráfico 14: Tiempo Promedio Trabajando de los subempleados profesionales.

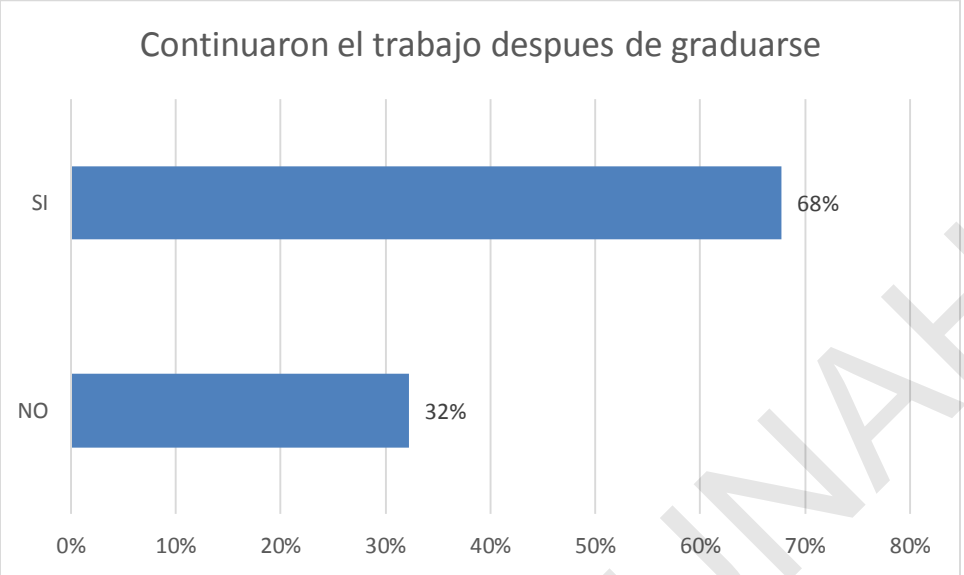


Fuente Elaboración Propia.

El 57% de los subempleados tenían un trabajo antes de graduarse de la universidad, de este 57% únicamente el 15% tenían un trabajo que tenía relación con la carrera que estaban cursando en ese momento (su profesión actualmente) y el 85% tenían un trabajo que no estaba relacionado con su carrera.

Como se muestra en el gráfico 15, únicamente el 32% de los encuestados que tenían trabajo antes de graduarse cambiaron de trabajo después de que se graduaron. El 68% no cambiaron de trabajo, y continúan laborando en un trabajo sin relación con su carrera. El 32% que se cambió de trabajo, actualmente se encuentran laborando en algo no relacionado con su profesión, lo que lleva a pensar que estas personas no encontraron un empleo relacionado con su profesión, o que se encontraban mejor establecidos en un área distinta a la de su trabajo.

Gráfico 15: Profesionales Universitarios que continuaron trabajando después de graduarse.

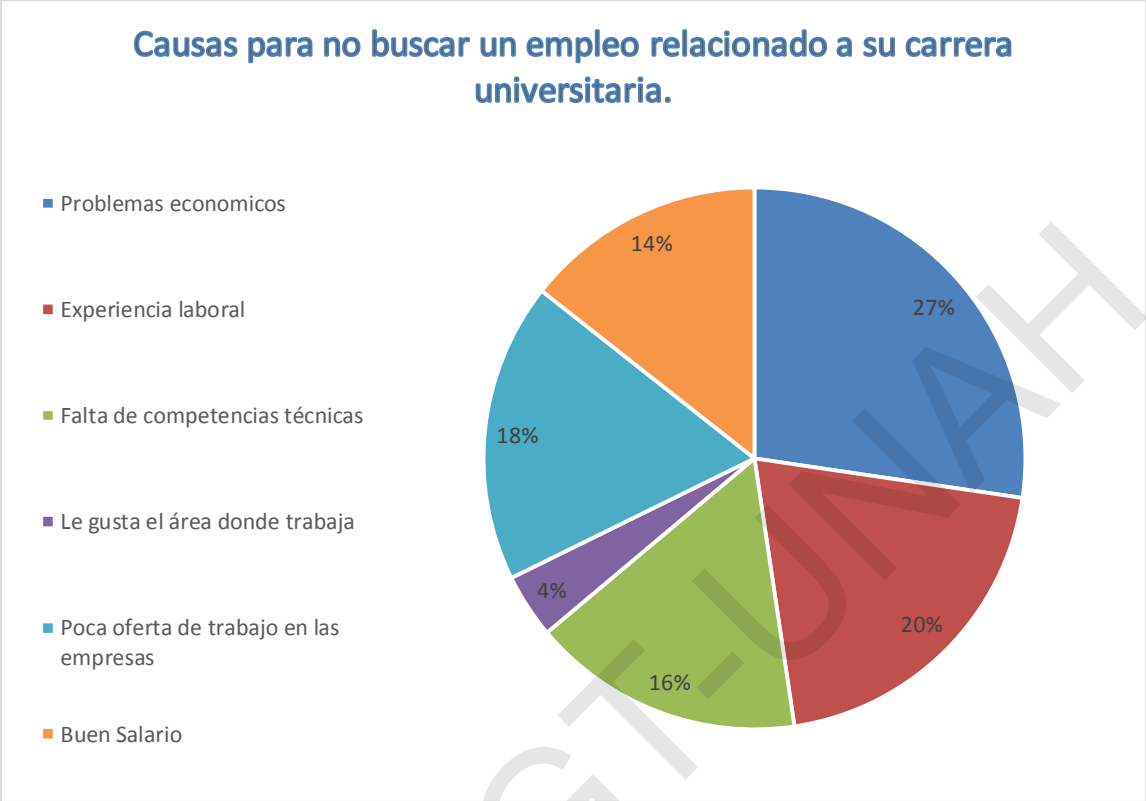


Fuente Elaboración Propia.

Las principales razones por las que ese 32% de los profesionales no cambiaron de trabajo son las que se muestran en el gráfico a continuación. La causa más común es por problemas económicos (27%), los profesionales necesitan una fuente de ingresos para hacer frente a las necesidades económicas. La segunda causa más importante con un 20% es la experiencia laboral, los profesionales continuaron con el trabajo sin relación con su carrera, ya que creen que de esta forma pueden ganar experiencial labora, aunque sea en un área distinta, y que después puede servirles para adquirir un mejor trabajo.

La causa de menor importancia por la que los profesionales continúan con un trabajo distinto al de su carrera, es porque les gusta el área donde trabajan, esta causa únicamente tiene un 4% de ocurrencia, lo cual lleva a pensar que los profesionales no están muy a gusto con su trabajo, pero este les permite satisfacer sus necesidades básicas y por este motivo continúan con el mismo.

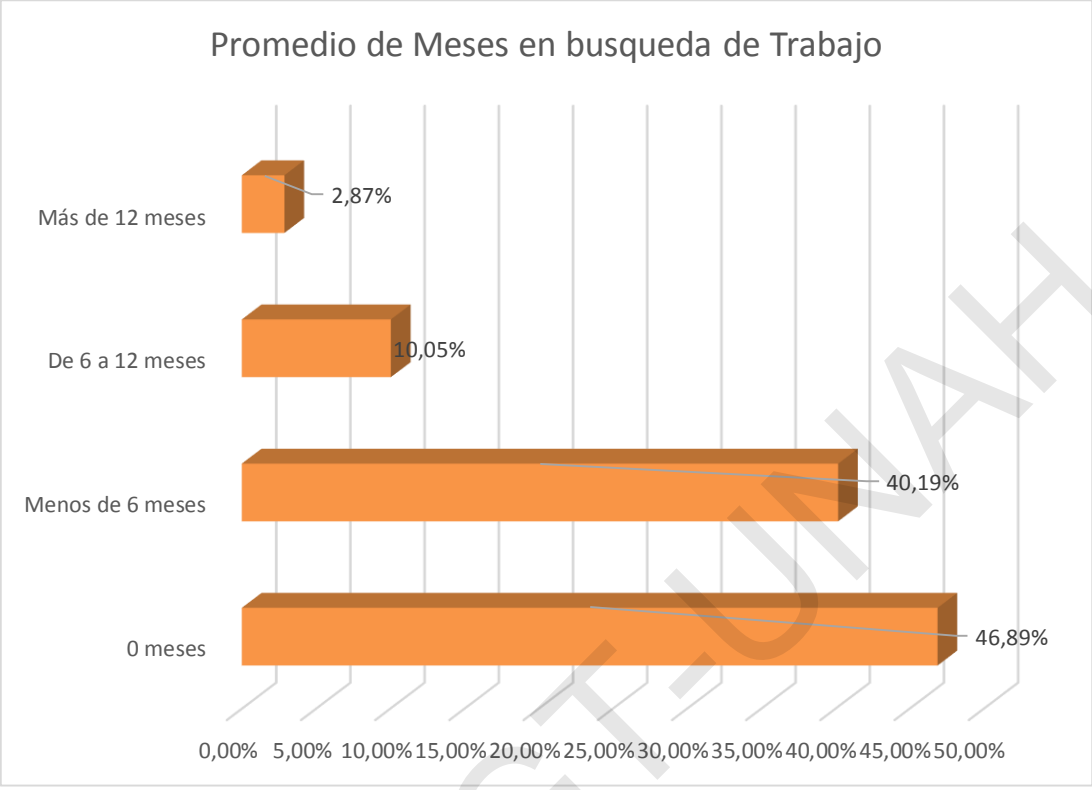
Gráfico 16: Causas para no buscar un empleo relacionado a la carrera universitaria.



Fuente Elaboración Propia.

El promedio de meses de búsqueda de empleo de los subempleados profesionales encuestados es en su mayoría de 0 meses, esto incluyendo a las personas que conservaron su trabajo después de graduarse de la universidad. Además el 40% de los profesionales se demoraron menos de 6 meses en encontrar un empleo; esto puede verse en el gráfico a continuación. Con esta situación se puede afirmar que los profesionales cuentan con estos trabajos, porque no realizaron una búsqueda más exhaustiva por miedo a no encontrar un trabajo, o que aceptaron el primer trabajo que se ajustara a sus necesidades.

Gráfico 17: Promedio de Meses buscando trabajo de los Subempleados Profesionales.



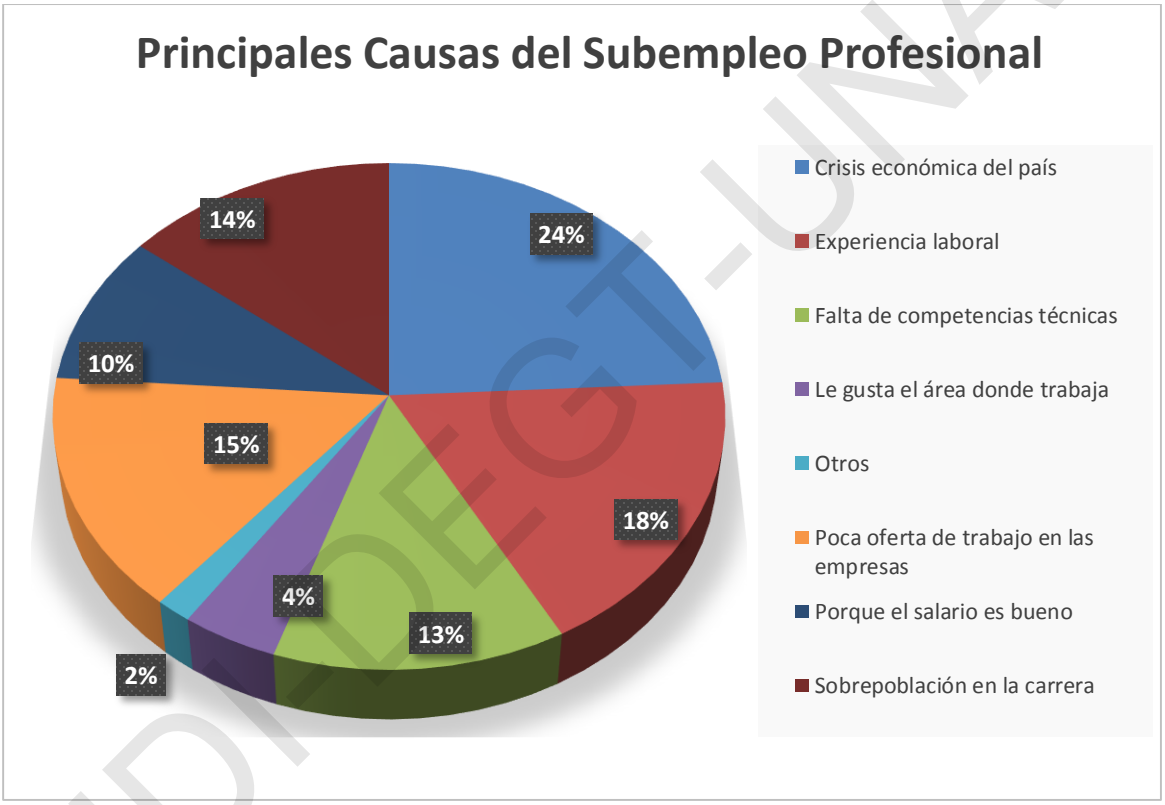
Fuente Elaboración Propia.

Finalmente las principales causas del subempleo profesional según la muestra encuestada son las que se muestran en el siguiente gráfico, aquí se puede observar que la mayoría (24%) acepta trabajos sin relación con su carrera universitaria por la crisis económica del país, ya sea para solventar sus problemas económicos o para apoyar con los ingresos de sus familias, lo anteriormente descrito está estrechamente relacionado al porque algunos profesionales conservaron su trabajo después de egresar de la universidad; a pesar de que este no tenga relación con su carrera.

La segunda causa más importante es ganar experiencia laboral, la cual es la segunda causa por la cual los desempleados aceptarían un trabajo sin relación a su profesión universitaria, por lo cual se puede concluir que para los universitarios con problemas de empleo (desempleo o subempleo) la experiencia es un factor de suma importancia y por lo cual están en la necesidad de aceptar cualquier tipo de empleo.

Otra de las principales causas por la cual los profesionales universitarios aceptan un trabajo sin relación con su carrera o permanecen en su trabajo, es porque les agrada el área donde trabajan (4%), estas personas no buscan otro empleo o algo relacionado con su carrera porque se sienten cómodos con el trabajo que ejercen o porque simplemente les gusta más que su carrera universitaria, y probablemente obtengan un salario más alto que ejerciendo su profesión; por lo cual no se ven en la necesidad de buscar otro empleo.

Gráfico 18: Principales Causas del Subempleo Profesional.



Fuente Elaboración Propia.

9.2 Discusión de los Resultados

A través de la metodología de minería de datos definida en el capítulo ocho, se pudo realizar un análisis apropiado que permitió determinar y predecir las principales causas de desempleo y subempleo profesional, a la vez se encontraron relaciones entre los datos, y a través de las mismas se pueden realizar diferentes predicciones de la problemática laboral en los profesionales universitarios.

Adicionalmente con la aplicación de esta metodología se realizó un análisis de diferentes algoritmos para determinar cuál permitía realizar una predicción acertada del estimado de desempleo y subempleo profesional, y de las principales causas que lo originan. El análisis se realizó partiendo de la información recolectada a través de las encuestas aplicadas a una muestra de profesionales, determinando de esta forma cuales son los principales factores que influyen en el desempleo y subempleo profesional.

Con la aplicación del algoritmo de serie temporal se determinó el índice o tasa estimada de desempleo del 2014, (el cual aún no se proporciona por el INE), sin embargo en las gráficas mostradas, se permite observar el estimado para varios años en curso. La limitante de este y los demás algoritmos consiste en que la predicción es más aproximada a medida se cuente con más datos, sin embargo como se planteó al inicio de la investigación esa es una de las limitantes del estudio, ya que no se cuenta con la suficiente información.

Sin embargo al tener una metodología definida para poder aplicar estas técnicas y al contar con mayor información se pueden realizar predicciones con la confiablidad de que las mismas sean acertadas. A su vez al contar con mayor información y si la misma se encuentra en constante actualización, en la metodología se pueden agregar dos ítems más antes de aplicar los algoritmos de datos. Estos pasos consistirían en:

1. El análisis, diseño y creación de una base de datos OLAP.
2. Creación y mantenimiento del proceso de ETL.

Ambos procesos son abarcados en el marco teórico de esta investigación y no serán difíciles de implementar si se realiza un análisis apropiado de la problemática a resolver, ya que de la misma forma se desarrollaron el resto de pasos de la metodología definida en el capítulo anterior.

Así mismo esta investigación puede extenderse al contar con mayor información con el paso del tiempo y de esta forma permitir la alimentación continua del proceso para poder brindar estas predicciones y estimaciones a las universidades, observatorios de problemas laborales, INE, secretaria de trabajo, y a cualquier otra entidad que le concierne.

Además se pueden realizar las encuestas cada cierto periodo de tiempo, y con esto realizar predicciones en cuanto al cambio de los factores que producen el desempleo y subempleo profesional con el paso del tiempo.

Si bien esta investigación es de tipo descriptiva porque no se establece una relación entre variables, dentro de la metodología se pueden definir hipótesis a resolver a través de la aplicación de los algoritmos de minería de datos, como por ejemplo:

- La causa estimada de desempleo profesional para el 2018 es del 5.8%, esto con base en los datos obtenidos según la predicción por medio del algoritmo de serie de tiempo.
- La principal causa de desempleo profesional es la crisis económica de país, esto según la predicción realizada por el algoritmo de árboles de decisión.

Según los resultados obtenidos de las principales causas del desempleo y subempleo profesional, se puede determinar que la crisis económica que atraviesa el país afecta en la búsqueda y conservación de un empleo, en el subempleo porque los profesionales aceptan un trabajo sin relación con su profesión porque necesitan los ingresos que el mismo les produce, y en el caso de los desempleados, las empresas pueden realizar recortes de personal para tratar de ajustar su situación a la crisis económica global que atraviesa el país.

Dentro de las causas generadas con la aplicación de los algoritmos se puede observar que muchas son compartidas entre el desempleo y subempleo, esto puede deberse a que los profesionales desempleados no encuentran empleo por las mismas razones que los subempleados aceptan trabajos sin relación con su carrera, por ejemplo, la sobrepoblación de la carrera, no se encuentran plazas disponibles o no se encuentran plazas de su profesión.

La experiencia laboral es otra de las causas compartidas, en donde los desempleados no encuentran trabajo debido a su falta de experiencia, y los subempleados aceptan otros trabajos en donde no necesitan tener experiencia.

9.3 Análisis del Cumplimiento de los Objetivos

En esta investigación se definieron cuatro objetivos, de los cuales se concluye a continuación:

1. *Aplicación de técnicas de minería de datos que permita estimar las principales causas de desempleo y subempleo profesional de egresados universitarios.*

   En este objetivo se planteaba la necesidad de desarrollar una metodología de minería de datos para estimar, la cual fue desarrollada y puede observarse en el capítulo VIII de esta investigación, y mediante la cual se pudieron cumplir los siguientes objetivos.

2. *Determinar que modelos o técnicas de minería de datos se adaptan mejor al análisis del desempleo y subempleo profesional de egresados universitarios.*

   Este objetivo pudo cumplirse mediante la aplicación de la metodología de minería de datos, en donde se desarrolló una comparativa y análisis de diferentes algoritmos de minería de datos para determinar cuál era el adecuado para realizar la predicción más acertada.

3. *Determinar los principales factores que influyen en el desempleo y subempleo profesional de egresados universitarios.*

   Las principales causas o factores fueron obtenidas mediante la tabulación y análisis de las encuestas, y apoyados con el uso de la minería de datos, también pudo realizarse una predicción de las mismas.

4. *Estimar el nivel de desempleo y* subempleo profesional del año 2014 a través de la aplicación de técnicas de minería de datos.

   Este objetivo pudo cumplirse gracias a la predicción desarrollada con los datos proporcionados por INE y con la aplicación de la metodología de minería de datos definida en el primer objetivo.

CONCLUSIONES

1. Al analizar tanto los datos reales proporcionados por las encuestas, como los datos obtenidos por medio de la aplicación de los algoritmos de minería de datos, se obtuvieron resultados muy similares, indicando que la crisis económica del país es la principal causa del desempleo y subempleo profesional en el país, y dado que esto es un problema social en el país, las medidas y acciones que deben tomarse para reducir estos índices deben ser una labor coordinada de todo el gobierno.

2. Con base en los resultados obtenidos, se puede concluir que las causas de desempleo y subempleo profesional son muy similares, por lo cual se puede afirmar que los profesionales con trabajos diferentes a su profesión, se encuentran laborando en el mismo para no estar desempleados.

3. Las predicciones de minería de datos, permiten tomar decisiones, antes de contar con los resultados reales, y ya que son muy acertadas pueden ayudar a las universidades a elaborar planes de acción a seguir para apoyar a sus egresados a combatir los problemas de desempleo y subempleo.

4. La metodología de minería de datos desarrollada en esta investigación, es solo una guía a seguir para la aplicación de diferentes algoritmos, pero la misma puede cambiar conforme a las necesidades de la problemática a analizar.

5. La minería de datos no es una solución a los problemas laborales que se vive en el país, pero puede ayudar a crear un mecanismo que permita realizar predicciones y análisis de los problemas que están por venir, y tomar medidas para enfrentar o disminuir los mismos.

6. La aplicación de la minería de datos no se limita a la problemática laboral, la misma puede utilizarse para predicciones de la deserción estudiantil, del crecimiento de las matriculas en las distintas carreras, y otro tipo de problemas o situaciones que necesiten ser analizadas, bien para realizar acciones en el presente o en el futuro.

RECOMENDACIONES PARA ESTUDIOS FUTUROS

Mencionando parcialmente las limitaciones, en el caso de futuros estudios, se podrían considerar una muestra de egresados únicamente de la UNAH, con la cual se podrían estudiar los desempleados y subempleados específicamente por carreras; lo cual podrá servir para la toma de decisiones o emprender futuras acciones a cada una de las facultades de la universidad.

Así mismo la recolección de datos con las encuestas u otros instrumentos de investigación, podrán ser almacenados periódicamente para poder contar con un histórico de información y de esta forma construir un Data Warehouse, al contar con este almacén de datos, se debe aplicar un proceso de ETL de forma periódica y además se deberá construir una base de datos OLAP, en la cual se crearan cubos para poder obtener la información de una forma más rápida y eficiente. Finalmente se podrán aplicar las técnicas de minería de datos para que los resultados sean más aproximados a medida se cuente con un mayor volumen de información.

Se podrán realizar estudios utilizando la minería de datos, pero con un enfoque diferente, dentro de la universidad podría aplicarse para predecir o determinar la carga académica del siguiente año, para determinar la cantidad de alumnos que ingresaran por periodo o año académico. Otra aplicación podría ser la deserción estudiantil por cada una de las carreras, o para determinar las principales razones de la sobrepoblación de algunas carreras.

Otra forma de aplicar la minería de datos es en el INE (Instituto Nacional de Estadística) ya que ellos cuentan con la información recopilada por la Encuesta permanente de hogares de propósitos múltiples, esta encuesta cuenta con información de muchas áreas de interés en el país, y con todos los años en que la encuesta ya ha sido aplicada, podría crearse un almacén de datos, y un sistema para el análisis de esta información, no únicamente aplicando minería de datos, si no también todo lo que es inteligencia de negocios y que ya se ha definido en esta investigación. Aplicar la inteligencia de negocios y la minería de datos a la información con la que cuenta el INE permitiría a las autoridades nacionales contar con información detallada de la situación del país y de esta forma apoyarlos en la toma de decisiones.

BIBLIOGRAFIA

Azoumana, K. (2013). *Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos. Pensamiento Americano, 41-51*.

Blanchard, O. (2006). *Macroeconomía.* Madrid: Pearson/Prentice Hall, 4a ed.m.

Cerón Reyes, M. d., & Gómez Díaz, H. (2010). *Mineria de Datos.*

Cfr. Samuelson, N. (2006). *Economía.* Madrid: McGraw-Hill.

Chavez, N., & Bavera, C. (2013). *Prototipo de Sistema de Inteligencia de Negocios utilizando Mineria de Datos sobre Software Libre.* Asunción, Paraguay: Universidad del Cono Sur de las Americas.

Cibertec. (2010). *Inteligencia de Negocios.* Lima, Peru: Cibertec.

Corrales, E., & Rodiguez, B. (2003). *La transicion del sistema educativo al mercado laboral. Analisis de los factores determinantes del primer empleo.* Universidad Rovira i Virgili: Actas de la V Jornadas de Economia de la Educación.

Date, C. (2001). *Introducción a los Sistemas de Base de Datos.*

diariowebcentroamerica. (30 de Abril de 2012). *diariowebcentroamerica.* Obtenido de http://www.diariowebcentroamerica.com/economia-y-turismo/mas-de-un-millon-gana-abajo-del-salario-minimo-en-honduras/

Diccionario de la Lengua Española. (2012). *Diccionario de la Real Academia Española-Vigesima segunda edición.*

Flores, J., & Julca, D. (29 de Marzo de 2010). *slideshare.net.* Obtenido de http://www.slideshare.net/janettejf/mineria-de-datos-3582262

Gartner. (Enero de 2006). *Glosario de Gartner.* Obtenido de www.gartner.com

Gartner Inc. (23 de 02 de 2015). *http://www.gartner.com/*. Obtenido de http://www.gartner.com/technology/reprints.do?id=1-2ACLP1P&ct=150220&st=sb

Garzon Perez, M. T. (2010). Sistemas Gestores de Bases de Datos. *Innovación y Experiencias Educativas*.

Gobernado Arribas, R. (2007). *La sobreeducación en España: estudio descriptivo y revisión crítica del concepto.* Málaga: Universidad de Malaga.

Haberstroh, R. (2008). *Oracle (r) data mining tutorial for Oracle Data Mining 11g Release 1*. Oracle.

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la Investigación* (5ta ed.). Perú: Mc Graw Hill.

Hondurasq, U. (29 de Mayo de 2013). *universia.hn*. Obtenido de http://noticias.universia.hn/en-portada/noticia/2013/05/29/1027084/10-profesiones-mas-demandadas-mercado-laboral.html

INE. (2013). *INE- Instituto Nacional de Estadisticas.* Obtenido de http://www.ine.gob.hn/index.php

INE. (2013). *XLIV Encuesta permanente de hogares de propositos multiples.* Tegucigalpa.

Lapuente, M. J. (2013). *Bases de Datos. Tesis doctoral.* Madrid: Universidad Complutense de Madrid.

Lopez, R. (2012). Empleo y desempleo en Honduras. *SOB - HONDURAS*.

Microsoft. (2010). Business Intelligence in SharePoint Server 2010.

Microsoft, D. N. (Marzo de 2014). *http://msdn.microsoft.com/.* Obtenido de http://msdn.microsoft.com/es-es/library/ms174949.aspx

Microstrategy. (Julio de 2015). *microstrategy.com*. Obtenido de http://www.microstrategy.com/es/soluciones

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). *RapidMiner: Rapid Prototyping for Complex Data Mining Tasks.* (2006) Yale (now: RapidMiner):: Yale.

Molina López, J., & García Herrero, J. (s.f.). *Técnicas de análisis de datos.* Madrid, España: Universidad Carlos III.

Moreno, M., & Burga, C. (2011). *¿Existe subempleo profesional en el Perú urbano?* Lima: Consorcio de Investigación Económica y Social.

Oracle. (2007). *Oracle Database 11g para Data Warehousing e Inteligencia de Negocios.*

Orallo, J. H. (2010). *Minería de Datos.* Valencia, España: Universidad Politecnica de Valencia.

Osorio, E. J. (2011). *Metodologia para el desarrollo de un sistema de interligencia de negocios basada en el proceso unificado.* Bogota, Colombia: Universidad Nacional de Colombia.

Parkin, M., & Esquivel, G. (2008). *Economía versión para Latinoamérica*. Mexico: Pearson.

PAUTSCH, J. G. (2009). *Minería de Datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación.* Posadas, Argentina: Universidad Nacional de Misiones.

Perversi, I. (2007). *Aplicación de la mineria de datos para la exploracion y deteccion de patrones delictivos en Argentina.* Buenos Aires: Instituto Tecnologico de Buenos Aires.

Ramírez Rojas, M. Á., & Guevara Fletcher, D. A. (2006). Mercado de trabajo, subempleo, informalidad y precarización del empleo: los efectos de la globalización. *Economía y Desarrollo*.

Robles Aranda, Y., & Sotolongo, A. R. (2013). *Integración de los algoritmos de minería de datos 1R, PRISM e ID3 a PostgreSQL.* La Habana, Cuba: Universidad de las Ciencias Informáticas.

Rodríguez Suárez, Y., & Díaz Amador, A. (2011). Herramientas de minería de datos. . *Revista Cubana de Ciencias Informáticas.*, 3-4. Obtenido de http://rcci.uci.cu/index.php/rcci/article/view/78

Silberschartz, K., & Sudarshan. (2005). *Fundamentos de Bases de Datos.*

Superior, D. d. (2007). Universidades Hondureñas.

Vallejos, S. J. (2006). *Mineria de Datos.* Corrientes, Argentina: Universidad Nacional del Norde.

Villamandos, N. C., Ocerin, J. M., & Castro, M. A. (2007). *El Perfil del Egresado Universitario Desempleado.* Cordoba: Universidad de Cordoba.

W.H., I. (1992). *Building Data Warehouse.* Technical Publishing Group.

Waikato, U. o. (Julio de 2015). *Weka*. Obtenido de http://www.cs.waikato.ac.nz/ml/weka/

Zelaya, H. (30 de Abril de 2013). Obtenido de http://hondurasdesempleo.blogspot.com/

ANEXOS

Anexo 1: Instrumento de Medición

Encuesta dirigida a Egresados Universitarios

El objetivo de la siguiente encuesta es conocer su experiencia laboral después de haber egresado de la universidad.

Indicaciones:
Siga las instrucciones que se le presentan a lo largo del cuestionario, para cada una de las proposiciones encierre la opción que corresponde a su caso, las preguntas marcadas con * son obligatorias de contestar.

1. ¿Cuál es su edad? *
   a) 20-25
   b) 26-30
   c) Más de 30

2. Seleccione su sexo *
   a) Femenino
   b) Masculino

3. ¿Hace cuantos años se graduó de la universidad?
   a) Menos de 1 año
   b) 1-3 años
   c) 4-8 años
   d) Más de 8 años

4. ¿Cuál es el área de la carrera de la cual se graduó? *
   a) Ciencias Económicas

b) Ciencias Médicas
c) Sistemas Informáticos
d) Ingeniería
e) Ciencias Jurídicas
f) Química y Farmacia
g) Humanidades y Arte
Otro: _____

5. ¿De qué carrera se graduó? *

_____

6. ¿De qué universidad se graduó? *

a) UNAH
b) Unitec
c) Católica
d) José Cecilio del Valle
e) UTH
f) Metropolitana
g) Politécnica
Otro: _____

7. ¿Se encuentra trabajando actualmente? *

a) Si
b) No

Si cuenta con un trabajo actualmente, pase a la siguiente sección (egresados universitarios con trabajo) de preguntas, de lo contrario pase a la sección de desempleados universitarios.

Egresados Universitarios con Trabajo

1. ¿Trabajó antes de graduarse de la universidad? *
   a) Si
   b) No

Si trabajó antes de graduarse de la universidad continúe a la pregunta 2, de lo contrario pase a la pregunta 5

2. ¿Tenía un trabajo relacionado con su carrera?
   a) Si
   b) No

Si su trabajo no era relacionado con el área continúe con la pregunta 3, de lo contrario pase a la pregunta 5

3. ¿Continuó con ese trabajo después de graduarse?
   a) Si
   b) No

Si continuó con el trabajo pase a la pregunta 4, de lo contrario pase a la pregunta 5

4. ¿Por qué razón continuó con el trabajo?
   a) Problemas económicos
   b) Por ganar experiencia laboral
   c) Falta de competencias técnicas para otro trabajo
   d) Porque el salario es bueno
   e) Porque no encontró un trabajo relacionado con su carrera
   f) Le gusta el área donde trabaja
   Otro: _____

Si contestó la pregunta 4 pase a la pregunta 6

5. ¿Cuánto tiempo después de graduarse de la universidad consiguió un trabajo?
   a) Menos de 6 meses

b) 6-12 meses
c) Más de 12 meses

6. ¿Hace cuánto tiempo se encuentra laborando en su trabajo actual? *

a) Menos de 6 meses
b) 6-12 meses
c) 1-2 años
d) Más de 2 años

7. ¿Cuál es en promedio su salario actual? *

a) Menos de L.8,000
b) L.8,000 - 15,000
c) L.15,000 - 25,000
d) L.25,000 - 35,000
e) Más de L.35,000

8. En su trabajo actual ¿Cuántas horas a la semana trabaja? *

a) Menos de 36 horas
b) Más de 36 horas

Si trabaja menos de 36 horas o si su salario es menor a L.8,000 continúe con la pregunta 9, delo contrario pase a la pregunta 11

9. ¿Le gustaría tener un segundo empleo?

a) Si
b) No

Si le gustaría tener un segundo empleo pase a la pregunta 10, de lo contrario pase a la pregunta 11

10. ¿Por qué motivos le gustaría tener un segundo empleo?

- Por tener mayores ingresos
- Por experiencia laboral
- Porque no le agrada su trabajo actual

Otro: _____

11. ¿Su trabajo actual está relacionado con su carrera?
    a) Si
    b) No

Si su trabajo no está relacionado con su carrera pase a la pregunta 12, de lo contrario pase a la pregunta 13

12. ¿Por qué tiene ese trabajo?
    a) Problemas económicos
    b) Por ganar experiencia laboral
    c) Falta de competencias técnicas para otro trabajo
    d) Porque el salario es bueno
    e) Porque no encontró un trabajo relacionado con su carrera
    f) Le gusta el área donde trabaja
    Otro: _____

Si contestó la pregunta 12 pase a la pregunta 15

13. ¿Aceptaría cambiar a un trabajo sin relación con su carrera?
    a) Si
    b) No

Si aceptaría cambiar de trabajo pase a la pregunta 14, de lo contrario pase a la pregunta 15

14. ¿Por qué aceptaría un trabajo sin relación con su carrera?
    a) Por ganar experiencia laboral
    b) Por un buen sueldo
    c) Por inconformidad en su trabajo actual
    d) Problemas económicos
    Otro: _____

15. En promedio ¿Cuántas personas que usted conoce de su misma carrera están trabajando actualmente en algo no relacionado con su carrera? *

a) 1-5
b) 6-10
c) Más de 10
d) Ninguno

16. En promedio ¿Cuántas personas que usted conoce de su misma carrera están desempleados actualmente? *

a) 1-5
b) 6-10
c) Con la tecnología de
d) Más de 10
e) Ninguno

17. Selecciones las principales razones por las que cree que un profesional universitario está desempleado (donde 1 es "baja" y 5 es "alta") *

	1	2	3	4	5
Falta de experiencia laboral					
Falta de competencias técnicas					
Sueldos muy bajos					
Poca oferta de trabajo en las empresas					
Crisis económica del país					
Sobrepoblación en la carrera					

Egresados Universitarios Desempleados

1. ¿Cuánto tiempo tiene de estar desempleado? *

a) Menos de 6 meses
b) 6-12 meses
c) Más de 12 meses

2. Selecciones las principales razones por las que cree que está desempleado (donde 1 es "baja" y 5 es "alta") *

	1	2	3	4	5
Falta de experiencia laboral					
Falta de competencias técnicas					
Sueldos muy bajos					
Poca oferta de trabajo en las empresas					
Crisis económica del país					
Sobrepoblación en la carrera					

3. En el último mes ¿En cuántas empresas ha finalizado un proceso de selección de trabajo, pero no ha sido contratado? *

a) 1-5
b) 6-10
c) Más de 10
d) Ninguna

4. ¿De cuántas plazas vacantes se ha informado en el último año y que no se apegan a su carrera?*

a) 1-5
b) 6-10
c) Más de 10
d) Ninguna

5. ¿Desde que se graduó ha tenido algún trabajo? *

a) Si
b) No

Si su respuesta es SI pase a la pregunta 6, de lo contrario pase a la pregunta 11

6. ¿Su trabajo anterior, estaba relacionado con su carrera?

a) Si
b) No

Si su respuesta es No pase a la pregunta 7, de lo contrario pase a la pregunta 8

7. ¿Por qué tenía ese trabajo?

a) Problemas económicos
b) Por ganar experiencia laboral
c) Falta de competencias técnicas para otro trabajo
d) Porque el salario es bueno
e) Porque no encontró un trabajo relacionado con su carrera
f) Le gusta el área donde trabaja

Otro: _____

8. ¿Por qué motivo perdió su trabajo?

a) Salarios muy bajos
b) Falta de competencias técnicas
c) Por buscar un mejor trabajo
d) Lo despidieron

Otro: _____

9. ¿Cuenta con experiencia laboral en el área en la que se graduó?

a) Si
b) No

10. ¿Aceptaría un trabajo sin relación con su carrera? *
    a) Si
    b) No

Si aceptará un trabajo pase a la pregunta 11, de lo contrario pase a la pregunta 13

11. ¿Por qué aceptaría un trabajo sin relación con su carrera?
    a) Por ganar experiencia laboral
    b) Por un buen sueldo
    c) Por decepción
    d) Problemas económicos
    Otro: _____

12. ¿Con qué promedio de salario aceptaría trabajar?
    a) Menos de L.8,000
    b) Más de L.8,000

Si contestó la pregunta 12 pase a la pregunta 15

13. ¿Por qué no aceptaría un trabajo sin relación con su carrera?
    a) Por orgullo
    b) Porque los salarios son muy bajos
    c) Porque quiere ejercer su profesión
    d) Porque no tiene necesidad de trabajar
    Otro: _____

14. Si logrará obtener un trabajo relacionado con su carrera, ¿Con qué salario aceptaría trabajar?
    a) Menos de L.8,000
    b) L.8,000 - 15,000
    c) L.15,000 - 25,000
    d) L.25,000 - 35,000
    e) Más de L.35,000

15. En promedio ¿Cuántas personas que usted conoce de su misma carrera están trabajando actualmente en algo no relacionado con su carrera? *

a) 1-5
b) 6-10
c) Más de 10
d) Ninguno

16. En promedio ¿Cuántas personas que usted conoce de su misma carrera están desempleados actualmente? *

a) 1-5
b) 6-10
c) Más de 10
d) Ninguno

Muchas Gracias por su colaboración