



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

**VICERRECTORADO DE INVESTIGACIÓN
INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA**

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CENTRO DE POSTGRADOS

**MAESTRÍA EN SISTEMA DE GESTIÓN DE LA
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN
DEL TÍTULO DE MAGÍSTER EN SISTEMA DE GESTIÓN DE
LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS**

**ANÁLISIS DE TENDENCIAS Y HALLAZGOS DE PATRONES
DE COMPORTAMIENTO DE LAS REMUNERACIONES PARA
EL PERSONAL QUE LABORA EN EL EJÉRCITO
ECUATORIANO UTILIZANDO ALGORITMOS Y TÉCNICAS
DE MINERÍA DE DATOS APLICADO A LA INDUSTRIA DE
DEFENSA NACIONAL**

AUTORES: JÁCOME PANELUISA HERNÁN

MENESES BECERRA EDWIN RICARDO

DIRECTOR: MSC. NINAHUALPA QUIÑA GEOVANNI

SANGOLQUI

2017



DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA INFORMACIÓN E

INTELIGENCIA DE NEGOCIOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, “ANÁLISIS DE TENDENCIAS Y HALLAZGOS DE PATRONES DE COMPORTAMIENTO DE LAS REMUNERACIONES PARA EL PERSONAL QUE LABORA EN EL EJÉRCITO ECUATORIANO UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS APLICADO A LA INDUSTRIA DE DEFENSA NACIONAL” realizado por los señores Ing. HERNÁN JÁCOME PANELUISA y el Ing. EDWIN RICARDO MENESES BECERRA, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar a los señores HERNÁN JÁCOME PANELUISA Y EDWIN RICARDO MENESES BECERRA para que lo sustente públicamente.

Sangolquí, 31 de Marzo del 2017

Master Geovanni Ninahualpa Quiña

DIRECTOR



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

AUTORÍA DE RESPONSABILIDAD

Nosotros, HERNÁN JÁCOME PANELUISA, con cédula de identidad N° 1707493159 y EDWIN RICARDO MENESES BECERRA, con cédula de identidad N° 1709995128 declaramos que este trabajo de titulación “ANÁLISIS DE TENDENCIAS Y HALLAZGOS DE PATRONES DE COMPORTAMIENTO DE LAS REMUNERACIONES PARA EL PERSONAL QUE LABORA EN EL EJÉRCITO ECUATORIANO UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS APLICADO A LA INDUSTRIA DE DEFENSA NACIONAL” ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaramos que este trabajo es de nuestra autoría, en virtud de ello nos declaramos responsable del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 31 de Marzo del 2017

Hernán Jácome Paneluisa

C.C. 1707493159

Edwin Ricardo Meneses Becerra

C.C. 1709995128



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

AUTORIZACIÓN

Nosotros, HERNÁN JÁCOME PANELUISA y EDWIN RICARDO MENESES BECERRA, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca virtual de la institución el presente trabajo de titulación “ANÁLISIS DE TENDENCIAS Y HALLAZGOS DE PATRONES DE COMPORTAMIENTO DE LAS REMUNERACIONES PARA EL PERSONAL QUE LABORA EN EL EJÉRCITO ECUATORIANO UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS APLICADO A LA INDUSTRIA DE DEFENSA NACIONAL” cuyo contenido, ideas y criterios son de nuestra autoría y responsabilidad.

Sangolquí, 31 de Marzo del 2017

Hernán Jácome Paneluisa

C.C. 1707493158

Edwin Ricardo Meneses Becerra

C.C. 1709995128

DEDICATORIA

A Dios por cumplir con mi sueño y haberme dado la familia que siempre está pendiente brindándome el amor sincero y apoyo constante y a todos mis amigos que me impulsaron a seguir hasta cumplir con lo que tanto anhelaba.

A mis padres quienes me inculcaron el respeto, valores y la constancia de siempre luchar para llegar a conseguir los objetivos que uno tiene en la vida, a través de sus sueños han sido los pilares fundamentales para llegar a mis metas.

Hernán

A mi amada esposa Anita Torres por ser el puntal de mi vida y por el apoyo incondicional que me da todos los días, a mis amados hijos Dayanna y Matias porque por ellos trato de superarme cada vez más y darles un futuro mejor y a mi querida madre Cristina que en el cielo está muy orgullosa de su hijo.

Ricardo

AGRADECIMIENTO

Agradezco a Dios por haberme permitido cumplir con mi objetivo propuesto, a mi esposa Miriam por la paciencia y la comprensión brindada, a mis queridos hijos Danny, Erika y Odalis quienes estuvieron siempre junto a mí.

A la Universidad de las Fuerzas Armadas ESPE en cuyas aulas recibimos los conocimientos que permitieron cumplir mis sueños.

Al Magister Ninahualpa Quiña Geovanni quien acepto dirigir nuestro proyecto de investigación y por su sincera amistad y apoyo en el tema de investigación.

Al PHD. Waldo Hasperué quien nos brindó tutorías mediante video conferencia desde Argentina.

A la Comandancia General de la Fuerza Terrestre la sección remuneración por las facilidades prestadas.

Me resulta muy complicado enumerar a todos aquellos que han contribuido a que se haya terminado con éxito el proyecto propuesto en la investigación, sin embargo expreso mis sinceros agradecimientos.

Hernán

Quiero expresar mis agradecimientos al Señor Master en Auditoria de Sistemas Geovanni Ninahualpa Quiña por la guía prestada en el desarrollo de esta investigación, al PHD. Waldo Hasperué por la ayuda en la investigación científica para culminar con éxito este proyecto, al Ejército Ecuatoriano por brindar todas las facilidades para que esta tesis sea una de las mejores y a todos aquellos que de una u otra forma ayudaron en esta investigación.

Ricardo

ÍNDICE GENERAL

DEDICATORIA.....	v
AGRADECIMIENTO.....	vi
ÍNDICE GENERAL.....	vii
ÍNDICE DE FIGURAS	x
ÍNDICE DE TABLAS	xii
RESUMEN.....	xiii
ABSTRACT	xiv
CAPÍTULO 1	1
GENERALIDADES.....	1
1.1 Introducción.....	1
1.2 Antecedentes.....	1
1.3 Justificación e importancia	2
1.4 Planteamiento del problema	3
1.5 Formulación del problema.....	3
1.6 Hipótesis	4
1.7 Objetivo General.....	4
1.8 Objetivos Específicos	4
CAPÍTULO 2	5
FUNDAMENTACIÓN TEÓRICA.....	5
2.1 Conceptos de minería.....	5
2.2 Web Mining	6
2.3 Data Warehouse.....	7
2.4 Beneficios de un data warehouse.....	8
2.5 Estructura de los datos	8
2.6 Data Mart	9
2.7 Clasificación del data mart	9
2.7.1 Data mart dependiente	10
2.7.2 Data mart independiente	10
2.7.3 Data mart híbrido	10
2.8 Tipos de data marts.....	10
2.8.1 Data mart OLAP (Procesamiento analítico en línea).....	10
2.9 Definición de técnicas de minería de datos	10
2.9.1 Reglas de Asociación	11

2.9.2 Técnicas de «clustering».....	11
2.9.3 Redes bayesianas	11
2.9.4 Árboles de decisión.....	12
2.9.5 Redes neuronales	12
2.9.5.1 Redes neuronales supervisadas y por corrección.....	12
2.9.5.2 Redes neuronales artificiales.....	13
2.9.6 Regresión Lineal Simple.....	13
2.9.7 Regresión Lineal Múltiple	14
2.9.8 Regresión no lineal	14
2.10 Patrones de información.....	14
2.11 Inteligencia artificial y técnicas de minería de datos	16
2.11.1 Redes neuronales artificiales en minería de datos	17
2.12 Limpieza de datos	18
2.13 Diferencia entre modelo y metodología	19
2.13.1 Metodología KDD (proceso de extracción de conocimiento)	19
CAPÍTULO 3	23
PROPUESTA MODELO DE REMUNERACIONES Y MINERÍA DE DATOS	23
3.1 Fuentes de información para el análisis de las remuneraciones	23
3.1.1 Organigrama Estructural	23
3.1.2 Proceso de remuneraciones del personal del (E.E).....	24
3.1.3 Procedimientos de remuneraciones del personal del (E.E).....	25
3.2 Diagramas de flujo de procedimientos de remuneraciones	28
3.3 Arquitectura jerárquica del proceso de remuneraciones.....	31
3.4 Conocimiento disponible modelo entidad relación remuneraciones	32
3.5 Desarrollo del proceso ETL.....	35
3.5.1 Extracción de Datos	35
3.5.2 Análisis de la tabla historial del rol de pagos (shrl_hisro).....	35
3.6 Selección y descripción de los atributos para aplicar minería de datos.....	40
3.7 Consultas de tablas de la base de datos remuneraciones a analizar.....	41
3.8 Pre-procesamiento de los datos en la base remuneraciones	45
3.8.1 Load (carga).....	46
3.9 Correlación por peso de los atributos seleccionados	46
3.9.1 Correlación para proyección del salario del personal militar	46
3.9.2 Correlación para proyección de salarios a través de gráficos	47

3.10 Aplicación de algoritmos de minería de datos y análisis de resultados.....	49
3.10.1 Modelo proyección salario coroneles con redes neuronales.....	50
3.10.2 Modelo proyección salario teniente coronel con redes neuronales.....	52
3.10.3 Modelo proyección sargento segundo con redes neuronales	55
3.10.4 Modelo proyección salario coroneles con regresión lineal.....	58
3.10.5 Modelo proyección salario teniente coronel con regresión lineal.....	61
3.10.6 Modelo proyección salario sargento segundo con regresión lineal	63
Totales por grado con neural net y regresión lineal	67
3.10.8 Modelo de Predicción de sueldos militar con redes neuronales	69
3.10.8.1 Determinación del modelo adecuado para solucionar el problema .	72
3.10.9 Total proyección con redes neuronales y regresión lineal	76
3.10.10 Propuesta de modelo de remuneraciones	76
CAPÍTULO 4	80
RESULTADOS	80
4.1 Informe de resultados.....	80
4.1.1 Antecedentes	80
4.1.2 Investigación	80
4.1.3.1 Regresión Lineal	80
4.1.3.2 Redes Neuronales	82
4.1.3.3 Redes Neuronales con todos los grados militares.....	83
4.1.3.4 Proyección para el año 2016.....	84
4.1.3.5 Proyección para el año 2017	86
4.1.3.6 Proyección para el año 2018.....	89
CAPÍTULO 5	92
CONCLUSIONES Y RECOMENDACIONES	92
5.1 Conclusiones.....	92
5.2 Recomendaciones	93
REFERENCIAS BIBLIOGRÁFICAS	94

ÍNDICE DE FIGURAS

Figura 1: Minería web	7
Figura 2: Estructura de un (D.W).....	8
Figura 3: Data warehouse y data marts	9
Figura 4: Esquema del proceso KDD resumido.....	20
Figura 5: Organigrama estructural de la dirección del personal del E.E	23
Figura 6: Diagrama de flujo estudio salarial	28
Figura 7: Diagrama de flujo cálculo de remuneraciones	29
Figura 8: Diagrama de flujo cálculo de remuneraciones al exterior	30
Figura 9: Diagrama de flujo liquidación salarial.....	31
Figura 10: Jerarquía de remuneraciones	31
Figura 11: Jerarquía de procesos del salario neto	31
Figura 12: Rol de pagos, procesos y novedades	32
Figura 13: Descuentos, rubros, préstamos y ahorros	33
Figura 14: Poder voluntario y retenciones	34
Figura 15: Ventana principal de talend open studio for data quality	35
Figura 16: Ventana inicio análisis de columnas.....	36
Figura 17: Selección de nuevo análisis	36
Figura 18: Tipo de análisis	37
Figura 19: Nombre del análisis	37
Figura 20: Tabla de análisis	38
Figura 21: Columnas de análisis	38
Figura 22: Análisis de resultados	39
Figura 23: Registros analizados	40
Figura 24: Proceso leer base de datos y llevar a Excel	46
Figura 25: Modelo generación peso de atributo.....	46
Figura 26: Pesos de atributos	47
Figura 27: Barras con pesos de atributos	47
Figura 28: Modelo pesos atributos para gráficos	48
Figura 29: Pesos atributos para gráficos	48
Figura 30: Barras con pesos atributos para gráficos	49
Figura 31: Modelo coronel con red neuronal.....	50
Figura 32: Predicción salario total coronel con red neuronal	51
Figura 33: Predicción salario coronel individual con red neuronal	52
Figura 34: Modelo teniente coronel con red neuronal	53
Figura 35: Predicción salario total tcrnl con red neuronal	54
Figura 36: Predicción salario teniente coronel individual con red neuronal.....	55
Figura 37: Modelo sargento segundo con red neuronal	56
Figura 38: Predicción salario total sargento segundo con red neuronal.....	57
Figura 39: Predicción salario sargento segundo con red neuronal.....	58
Figura 40: Modelo coroneles con regresión lineal.....	58
Figura 41: Predicción salario total coronel con regresión lineal.....	59
Figura 42: Predicción salario coronel con regresión lineal.....	60
Figura 43: Modelo teniente coronel con regresión lineal.....	61

Figura 44: Predicción salario total teniente coronel con regresión lineal	62
Figura 45: Predicción salario teniente coronel con regresión lineal	63
Figura 46: Modelo sargento segundo con regresión lineal	64
Figura 47: Predicción salario total sargento segundo con regresión lineal	65
Figura 48: Predicción salario sargento segundo con regresión lineal	66
Figura 49: Subprocesos de todo el personal militar con redes neuronales	69
Figura 50: Suma de predicción por grados con redes neuronales	70
Figura 51: Predicción individual por grados con redes neuronales	70
Figura 52: Predicción de sueldo al 2016 por grados	71
Figura 53: Numérico del personal militar por grados	72
Figura 54: Suma de predicción por grados con regresión lineal	72
Figura 55: Proyección de salarios por provincias con n. n	73
Figura 56: Proyección de salarios por estado civil con n. n	74
Figura 57: Predicción del salario por fecha de estado.....	75
Figura 58: Modelo de predicción completo con redes neuronales.....	77
Figura 59: Predicción total por grados con redes neuronales	77
Figura 60: Suma de predicción total con redes neuronales.....	78
Figura 61: Suma de predicción total con regresión lineal.....	78
Figura 62: modelo de regresión lineal coroneles proyección 2016.....	81
Figura 63: Proyección sueldo coroneles individual 2016 con r. l.....	81

ÍNDICE DE TABLAS

Tabla 1: Totales por grado con neural net y regresión lineal.....	67
Tabla 2: Diferencia proyección neural net y regresión lineal	76
Tabla 3: Diferencia entre neural net, regresión lineal con datos reales	79

RESUMEN

Este documento de investigación presenta proyecciones de sueldos del personal militar según el grado en forma individual y total; valiéndose de la información de remuneraciones histórica del año 2007 hasta el 2015; lo que ha permitido a través de la utilización de técnicas de minería de datos como la regresión lineal y las redes neuronales comparar la que mejor se aproximaba a la realidad; permite generar proyecciones de salarios del año 2016 para comprobar los mismos con información real y a su vez realizar las proyecciones de salario del año 2017 y 2018; ayudando de esta forma en la generación de proyección de salarios para futuros años; para lograr esto en el primer capítulo se detalla el problema que existía para proyectar salario y se establece los objetivos a conseguir; en el segundo capítulo se describe el fundamento teórico de minería de datos con las técnicas y se detalla la metodología KDD que se aplicó en la investigación; en el tercer capítulo se analiza la base de datos de remuneraciones del personal militar utilizando el modelo físico y lógico de acuerdo a las tablas que fueron necesarios en nuestra investigación para generar modelos utilizando regresión lineal y redes neuronales a fin de determinar el modelo que realmente permita realizar la proyección del salario; en el capítulo 4 se presenta un informe a detalle de los resultados presentados por los modelos de regresión lineal y redes neuronales; En el último capítulo se presentó las conclusiones y recomendaciones a las que se ha llegado.

PALABRAS CLAVE:

MINERÍA DE DATOS

BASE DE DATOS

TÉCNICAS DE MINERÍA DE DATOS

KDD (PROCESO DE EXTRACCIÓN DEL CONOCIMIENTO)

PATRONES

MODELOS

RED NEURONAL

ABSTRACT

This research paper presents salary projections of military officers according to its range individually and in total form; using the historical remuneration information from 2007 to 2015. This fact has allowed the use of data mining techniques such as linear regression and neural networks in order to compare the best approach to reality. This approach allows to generate projections of salaries of year 2016 to compare with real information and to perform salary projections of years 2017 and 2018; contributing in this way in the generation of salary projection for future years. For achieving this goal, the first chapter describes the problem regarding salary projections and sets the objectives to be achieved; The second chapter describes the theoretical basis of data mining, the techniques, details, and the KDD methodology applied in the research; In the third chapter we analyze the remuneration database of military personnel using the physical and logical model according to the tables used in the research for generating models using linear regression and neural networks in order to determine the successful model that allows to realize the salary projection. Chapter 4 presents a detailed report of the results presented by linear regression models and neural networks. In the last chapter we present the conclusions and recommendations arrived.

KEY WORDS:

DATA MINING

DATABASE

DATA MINING TECHNIQUES

KDD (KNOWLEDGE DISCOVERY IN DATABASES)

PATTERNS

MODELS

NEURAL NETWORK

CAPÍTULO 1

GENERALIDADES

1.1 Introducción

Una vez comprendido los conceptos fundamentales en que se basa la minería de datos, en el presente trabajo investigativo se aprovechara de herramientas de minería de datos para lograr determinar patrones que ayuden a descubrir la tendencia que tiene las remuneraciones del personal militar del Ejército Ecuatoriano (E.E), tomando como referencia la base de datos histórico de años anteriores; esto ayudara para que el personal que procesa las remuneraciones y los altos mandos tomen decisiones acertadas y en beneficio del personal del (E.E.).

Aprovechando la información que se encuentra en las bases de datos se podrá aplicar métodos que permitan investigar cual fue el comportamiento de la información valiéndose de técnicas de minería de datos, así como el análisis estadístico en los diversos niveles jerárquicos con que cuenta el personal militar que trabaja para el (E.E.) esto permitirá entregar información al mando superior que en un principio no se podía utilizar en la institución, esto se puede presentar en un formato de información para la toma de decisiones en la (C.G.E.E).

El propósito de realizar este proyecto de investigación es poder determinar la forma de realizar proyecciones de sueldos, esto apoyará en poder saber cuánto será el presupuesto requerido para salarios y solicitar únicamente lo que sea necesario logrando que los altos mandos militares puedan realizar otros pedidos para otras actividades que son necesarias dentro del (E.E).

1.2 Antecedentes

En nuestro país no existen experiencias conocidas donde se hayan utilizado minería de datos dentro de las remuneraciones en el sistema de la defensa, es por esta razón la motivación para desarrollar esta investigación y aportar con este estudio; buscando tendencias que ayuden a generar modelos para aprovechar los mismos en el campo de proyecciones de sueldos de cualquier entidad pública o privada de nuestro país.

Muchas organizaciones y empresas tienen grandes volúmenes de información almacenadas en bases de datos que no han sido revisadas peor aún investigadas; lo que

realmente está siendo desaprovechada a fin de dar soluciones en diversos campos o niveles de las empresas.

Es por esta razón que la toma de decisiones a nivel del alto mando no se realiza de una manera técnica, sino basados en la experiencia del día a día de la persona que ha trabajado con estos datos como es el caso del personal del (E.E) que trabaja en la sección remuneraciones.

En la actualidad los investigadores están aprovechándose de técnicas de minería de datos para poder extraer conocimiento que se encuentra oculto en grandes volúmenes de información y una vez realizado el análisis de los mismos; están apoyando en la toma de decisiones encontrando modelos que realmente son considerados como herramienta fundamental para los altos mandos gerenciales dentro de una organización; señalando que, éste, es el motivo fundamental del presente trabajo investigativo.

1.3 Justificación e importancia

Se puede justificar este estudio utilizando minería de datos y generando patrones para realizar el control de los sueldos dentro del (E.E.), consiste en que permitirá a las personas encargadas de proyectar el presupuesto para salarios del personal militar; solicitar en forma oportuna la cantidad que sea necesaria para cada año, logrando optimizar los pedidos de presupuestos para el pago de salarios que perciben el personal militar del (E.E).

Adicionalmente la generación de modelos permitirán determinar según el grado de cada personal militar cuanto corresponde a cada uno de ellos, mediante la creación de patrones de minería de datos, se podría mantener segmentado al personal militar por grado de acuerdo a la función que cumple dentro del (E.E.); esto puede ayudar a un mejor manejo del personal militar y que sus remuneraciones estén acorde al cargo que desempeñan dentro de la institución. Permitirá contar con el personal militar capacitado de acuerdo al grado y puesto correspondiente y contar con ellos cuando sea necesario y proceder a realizar nuevos reclutamientos cuando sea necesario.

Estos modelos ayudarán al personal del (E.E) y en especial a las personas que realizan las proyecciones de sueldos a ser más eficientes en los pedidos que son necesarios, justificar a tiempo las solicitudes que realizan de presupuesto a las unidades financieras; reduciendo errores en las proyecciones de salarios, logrando que

los altos mandos realicen las peticiones justificando y demostrando técnicamente lo solicitado.

Indicando, además, que el presente trabajo de investigación no solo beneficiaría a nivel del departamento de remuneraciones del (E.E) sino a toda empresa pública que desea utilizar para realizar proyecciones de sueldos.

1.4 Planteamiento del problema

Las constantes variaciones de la variable remuneración que es de tipo cuantitativo pueden provocar muchos problemas al momento de crear proyecciones de sueldos, básicamente el pedido del presupuesto para cada uno de los años subsiguientes se realiza únicamente aumentando un dos por ciento al presupuesto utilizado en el año actual, es por esta razón que surge la necesidad de investigar y crear modelos de minería de datos que permitirán de una manera científica realizar una proyección de sueldos más cercana a la realidad, esto se da mediante la determinación de relaciones entre patrones de comportamiento de la información que se guarda en la base de datos de remuneraciones.

1.5 Formulación del problema

Básicamente una de las mayores dificultades dentro de las remuneraciones es el presupuesto que se necesita para remuneraciones en los próximos años del personal militar dentro del (E.E). Las principales dificultades son las siguientes:

- En este momento no existe ningún modelo de minería de datos para ayudar en la proyección de remuneraciones del personal militar del (E.E).
- La base de datos del sistema SIPER y el módulo de remuneraciones del personal militar del (E.E) no se está explotando adecuadamente para poder tomar decisiones oportunas.
- La remuneración que obtienen el personal militar del (E.E) no se justifica de una manera técnica en cuanto a la proyección de las remuneraciones de acuerdo a la función que desempeña dentro de la institución.

Para poder solucionar este problema, el proyecto de investigación realizará un análisis de las remuneraciones del personal militar que forma parte del (E.E) para lo cual se utilizará la base del sistema SIPER en el módulo de remuneraciones, lo que permitirá la toma de decisiones por parte de los altos mandos de la institución. Por medio de la utilización de técnicas de minería de datos se podrá entregar una solución

innovadora y revolucionaria para poder proyectar las remuneraciones del personal militar ya sea por grados o de una forma global.

El modelo de minería de datos generado se podrá utilizar dentro del (E.E.) a fin de optimizar la proyección en el campo de las remuneraciones.

El patrón de minería de datos generado dentro de esta investigación permitirá optimizar el pedido presupuestario a la dirección de finanzas del (E.E.) evitando pedir los recursos de una forma no adecuada o errónea.

1.6 Hipótesis

La propuesta de un modelo para control de las remuneraciones basado en técnicas de minería de datos permitirá la detección de posibles anomalías en el sistema actual de proyecciones de remuneración del personal militar del (E.E); debido a que se realizará un estudio de la base de datos tomando en cuenta el rol de pagos como variable dependiente en base a la información almacenada; se podrá comparar las variables independientes con la finalidad de generar datos de proyecciones de las remuneraciones y permitir que los altos mandos tomen decisiones oportunas para el beneficio de la institución.

1.7 Objetivo General

Descubrir y analizar las variaciones en las remuneraciones del personal militar del (E.E) a fin de determinar modelos de comportamiento a partir de los datos almacenados en la base de datos utilizando técnicas de minería de datos.

1.8 Objetivos Específicos

- Realizar una limpieza de los registros de la base de datos remuneraciones, mediante un procesamiento que permita la organización adecuada de los datos
- Aplicar minería de datos, empleando inteligencia de negocios para ejecutar algoritmos de búsqueda inteligente para encontrar las razones de cambio en el proceso de remuneraciones.
- Realizar el estudio de la base de datos a fin de implementar técnicas de minería de datos que permita extraer información valiosa y oportuna para lograr determinar soluciones y la toma oportuna de decisiones.
- Utilizar herramientas como RapidMiner y MS EXCEL 2013 para transformar la base de datos transaccional en información útil para el proceso de toma de decisiones.

- Se realizará gráficos para análisis a fin de obtener patrones de comportamiento en las remuneraciones.

CAPÍTULO 2

FUNDAMENTACIÓN TEÓRICA

2.1 Conceptos de minería

Es el proceso de hallar conocimiento beneficioso para una organización, que no ha sido explorado y se encuentra aún escondido; almacenado en grandes bases de datos que apoyará a la toma de decisiones gerencial de una empresa. La tarea fundamental de la minería de datos es encontrar modelos que ayuden a minimizar los errores aplicados en una organización. Para que este proceso sea seguro deberá ser investigado y automatizado; el uso de los patrones descubiertos deberá apoyar en la generación de proyectos que reporten beneficios en la toma de decisiones de acuerdo a la realidad del problema y de esta forma permitirá realizar proyecciones seguras. (AMORES & DIEGO, 2013)

Otro concepto de minería de datos se podría decir que es el proceso de detectar la información útil que se halla en grandes volúmenes de datos a través de análisis estadísticos a fin de encontrar patrones y tendencias de datos; los mismos son muy difíciles de entender por lo que se pueden recopilar y estudiar utilizando técnicas de minería de datos. (MICROSOFT, 2014)

Se podría definir a la minería de datos como un conjunto de técnicas y métodos que permiten investigar grandes bases de datos utilizando herramientas para análisis de datos como RapidMiner, donde el objetivo radica en encontrar patrones de datos iguales. (SINNEXUS, 2013)

La minería de datos surge para ayudar a indagar el contenido de una base de datos; para lo cual hace uso de experiencias estadísticas y de algoritmos de investigación parecidos a la inteligencia artificial¹ y a las redes neuronales; cuyo objetivo consisten en la interconexión de neuronas y la colaboración entre ellas dando como resultado un estímulo de salida lo que permitirá realizar un estudio de los resultados obtenido a fin de determinar el patrón o tendencia de los datos. (SINNEXUS, 2013).

La minería de datos consiste en la extracción de información que reside de manera implícita en las bases de datos de una organización. Dicha información al inicio

¹ **La Inteligencia Artificial** es la rama de la ciencia de la computación dedicada al desarrollo de agentes racionales no vivos; se podría decir que tiende a la creación de máquinas que puedan pensar.

mientras no se investigue será desconocida y una vez encontrada el patrón de tendencia; podrá resultar útil para tomar decisiones gerenciales. Se podría decir que la minería de datos prepara, sondea y explora los datos para sacar la información oculta en ellos.

2.2 Web Mining

El web mining es el proceso dirigido a extraer conocimiento, previamente escondido y potencialmente útil, que se encuentra en Internet. Del mismo modo que la minería de datos constituye un elemento de análisis básico dentro de los procesos de negocio que existe en una empresa, en nuestro caso es fundamental en la investigación de remuneraciones del personal militar que labora en el (E.E), ya que permitirá obtener modelos de conocimiento útiles a partir de la información de la base de datos existentes en la web; esta herramienta permite explotar el conocimiento procedente de la interacción personal con la organización; detallando modelos de análisis que determinen patrones de las formas de uso y consumo, descubriendo características de los distintos grupos o segmentos de clientes que hagan uso de este medio; en nuestra investigación permitirá buscar alguna relación con el tema propuesto. (Molina, 2012)

Se podría decir que las fases de la minería en la web son las siguientes:

Recopilación de los datos, limpieza de los datos, transformación de los datos, razonamiento y conclusiones. Los log de los servidores web se encuentran formados por la dirección ip, la fecha de la visita, la página que fue visitada url, el tamaño, el browser y el sistema operativo.

Existen tres clases de minería en la web: minería de contenido en la web, minería de estructura en la web y minería de uso en la web

- La minería de uso en la web permite descubrir lo que los usuarios buscan en Internet.
- La minería de contenido en la web busca información de texto, imagen, sonido o datos de video en la web que sean útiles.
- La minería de estructura en la web permite investigar cómo se encuentran los hipervínculos dentro de una página web, de esta forma permite dar un informe estructural de la página y el sitio web. (Molina, 2012).

Lo señalado anteriormente se observa (ver figura 1).

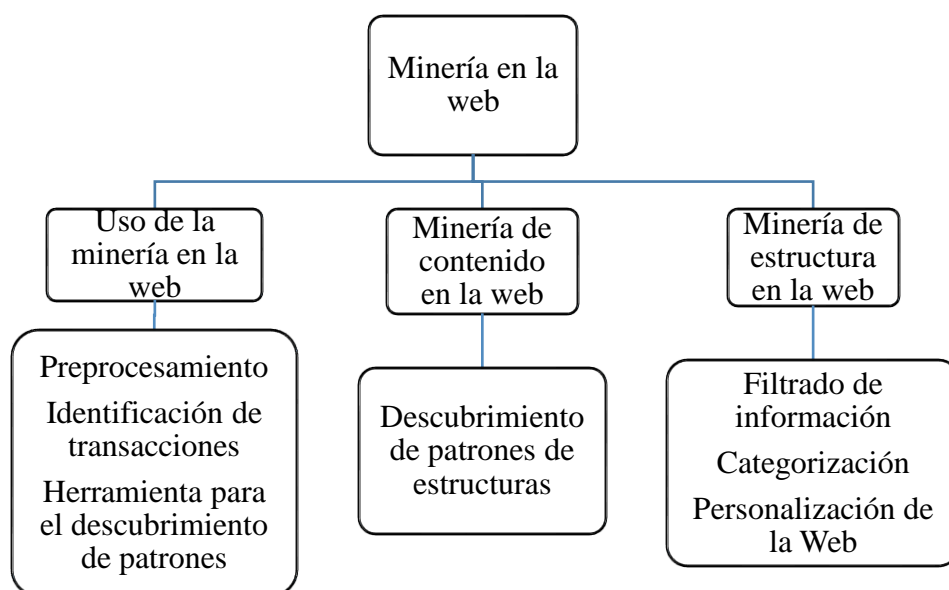


Figura 1: Minería web

Fuente: (Velasco López , 2013)

2.3 Data Warehouse

Un data warehouse (DW) es un almacén de información organizados en base al interés de la organización, que concentra un gran volumen de información, la cual se distribuye por medio de diversas herramientas de consulta y de creación de informes orientadas a la toma de decisiones.

Se podría decir que es la plataforma que concentra la información de la empresa; guardando la misma y presentando indicadores claves de la empresa para la toma de decisiones.

Se podría decir que en una universidad la data warehouse se organiza alrededor de sujetos tales como estudiantes, clases y profesores; es decir en conjunto forman un almacén de información a la cual se accede mediante herramientas de consulta y se generan informes de acuerdo a la necesidad que ayudan en la toma de decisiones.

A continuación se muestra (ver figura 2) la estructura de un (D.W).

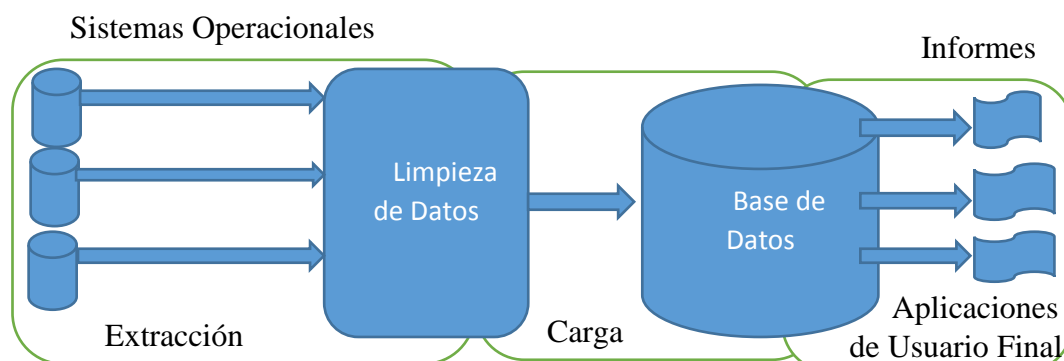


Figura 2: Estructura de un (D.W)

En la figura se puede visualizar la estructura del (D.W), la misma que se inicia con la extracción de información desde diferentes sistemas operacionales de la empresa, el siguiente paso es realizar una limpieza y organización de esos datos, los mismos que son almacenados en el (D.W), para finalmente en función de estos elaborar informes predefinidos.

Cabe señalar que para poder analizar los datos con fiabilidad es necesario que en estos exista una cierta estructuración y coherencia.

2.4 Beneficios de un data warehouse

- Proporciona una herramienta para la toma de decisiones basándose en información integrada y global de la organización.
- Permite aplicar diferentes técnicas de análisis y estadísticas para poder relacionar datos específicos de la organización.
- Proporciona la capacidad de estudiar datos del pasado y de predecir situaciones futuras de una organización.
- Permite aprovechar la tecnología que permita en base a la investigación generar informes para la toma de decisiones de manera oportuna. (Luzón, 2014).

2.5 Estructura de los datos

Para que los datos, a ser analizados sean considerados como íntegros, es necesario relacionar los distintos archivos y bases de datos de manera que se pueda extraer conclusiones validas; sin embargo se podría tener los siguientes inconvenientes:

- Poseer diferentes tipos de datos representando un mismo concepto. (Ejemplo el problema de la representación de los años con 2 o 4 dígitos).

- Tener diferentes claves para representar un mismo elemento. (Diferente codificación).
- Tener diferentes niveles de precisión para representar un dato.
- Los archivos pueden estar en diferentes sistemas informáticos y soportes. (Molina, 2012).

2.6 Data Mart

Un data mart (D.M) es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica; en nuestra investigación remuneraciones del (E.E); se caracteriza por seleccionar los datos de la organización en forma óptima para analizar la información en forma minuciosa de los procesos que afecten a dicho departamento. (Castillo & Palomino, 2013).

Los (D.M), tienen las mismas características de integración y orientación temática; sin pérdida de información igual que el D.W. Se puede deducir que mientras el D.W central comienza a crecer de una forma muy rápida, es necesario solamente una pequeña cantidad de información contenidos en el data warehouse. Para crear una data mart no se debe realizar solamente una copia o réplica de la información, sino se debe realizar una segmentación de los datos así como también se los debe consolidar de una manera adecuada. (Luzón, 2014).

A continuación se muestra (ver figura 3) el data warehouse y data marts.

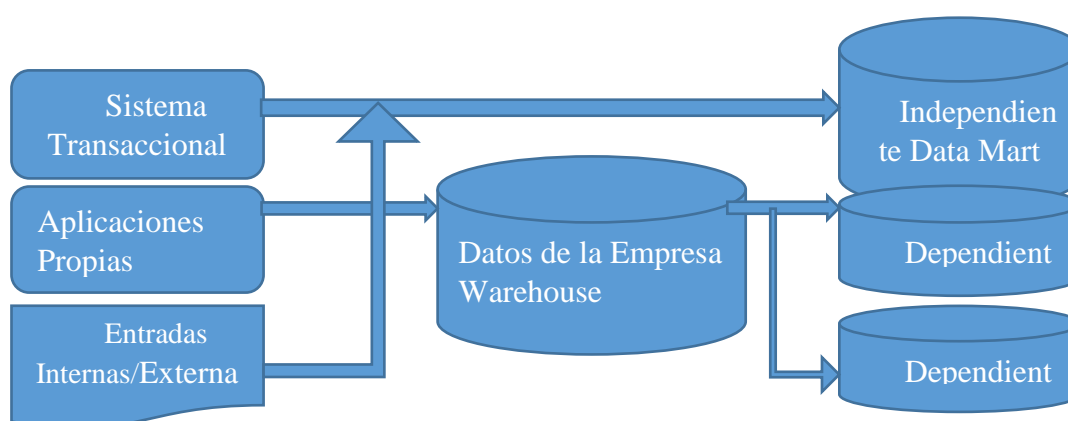


Figura 3: Data warehouse y data marts

2.7 Clasificación del data mart

A continuación se detalla la clasificación de (D.M)

2.7.1 Data mart dependiente

Los (D.M) dependientes son aquellos que reciben los datos desde una (DW); la fuente de los datos es única.

2.7.2 Data mart independiente

Son aquellos que toman sus datos directamente desde los sistemas transaccionales y no dependen de otros (D.W); se alimenta generalmente de las empresas.

2.7.3 Data mart híbrido

Los (D.M) híbridos permiten combinar las fuentes de datos de un (D.W) con otras fuentes de datos tales como los sistemas transaccionales.

2.8 Tipos de data marts

2.8.1 Data mart OLAP (Procesamiento analítico en línea)

Se base en la lectura a las base de datos, lo más tradicional que se realiza son consultas con escasa actualizaciones, inserciones o eliminaciones.

2.8.2 Data mart OLTP (Procesamiento de transacciones en línea)

Pueden basarse en un simple extracto del (D.W), no obstante, lo común es introducir mejoras en su rendimiento a través de agregaciones y filtrados; aprovechando las características de acceso a los datos ya que permite la lectura y escritura en las tareas. Se puede deducir que un (D.M) de este tipo permite el manejo de la información con facilidad. (Castillo & Palomino, 2013).

2.9 Definición de técnicas de minería de datos

Las diferentes técnicas a utilizarse para el desarrollo de este trabajo son las siguientes.

- Revisión de artículos científicos².
- Observación.
- Pruebas.

Dentro de esto se aplicara los modelos descriptivos y predictivos:

Los modelos descriptivos permiten realizar búsquedas en los datos que se encuentran en grandes volúmenes de información a fin de encontrar modelos que ayuden en la investigación y poder tomar decisiones acertadas.

Los modelos predictivos ayudan a genera proyecciones futuras a través de ajustes realizados a los modelos en construcción; como el cambio de los valores de los

² Es un trabajo de investigación realizada por investigadores en un tema específico.

coeficientes en los ciclos de entrenamiento, tasa de aprendizaje e impulsos que se podría ajustar en una red neuronal por lo que se conocen como aprendizaje supervisado ya permite analizar y comparar que tan ajustado está a la realidad. La investigación realizada proveerá de una proyección de las remuneraciones a un determinado año según sea el caso del personal militar que labora en el (E.E.) y poder solucionar los problemas que se puedan presentar tomando decisiones acertadas en relación a la información entregada por nuestro producto.

A continuación se explica las técnicas de minería de datos:

2.9.1 Reglas de Asociación

Son utilizadas cuando el objetivo es realizar análisis exploratorios, buscando relaciones dentro del conjunto de datos; por ejemplo ventas de mercancías ya que son especialmente interesantes a la hora de comprender las rutinas de compra de los compradores y constituyen algo fundamental en la relación de las ofertas y ventas cruzada.

2.9.2 Técnicas de «clustering»

Se puede decir que esta técnica permite concentrar rangos de datos de acuerdo a sus características siempre y cuando exista semejanza en los mismos; al utilizarlos en un proceso de análisis permite encontrar patrones que existen en la base de datos que se encuentra sujeta a estudio.

Esta técnica también se le llama de agrupamiento, ya que permite la identificación de características o grupos donde los elementos guardan gran semejanza entre sí y muchas diferencias con los de otros grupos. Se aplica sobre datos numéricos; para lo cual se debe identificar los grupos de análisis. (Luzón, 2014)

2.9.3 Redes bayesianas

Se puede determinar que esta red se apoya en el teorema de Bayes la misma que requiere de estadística; y a su vez está se encarga de pronosticar la posibilidad de que ocurra un acontecimiento para lo que requiere de información mínima que debe corresponder a una determinado grupo; es decir representa todos los posibles acontecimientos que pueden ocurrir y genera predicciones de los mismos. (Calderón, 2014).

2.9.4 Árboles de decisión

Se determina que estos árboles permiten representar de forma gráfica una serie de reglas sobre la decisión que se debe tomar en la asignación de un valor de salida a un determinado registro.

Razón por lo que se establece que estos árboles permiten una clasificación controlada ya que dependen de una variable sobre la cual se va a realizar la investigación. Una de las ventajas se podría decir que permite realizar análisis de una forma fácil; ya que se establecen reglas de forma sencilla que aportarían en la investigación y ayudarían en la toma de decisiones en el ámbito organizacional. (Gutierrez J. , 2012)

2.9.5 Redes neuronales

Esta técnica ayuda a generar proyecciones; ya que permiten encontrar comportamientos similares en las bases de datos y la determinación de características comunes.

Una de las características de las redes neuronales es la capacidad de aprender y generar nuevo conocimiento a partir de datos parciales; ya que realizan ajustes con los datos que existen en una base de datos y pueden determinar proyecciones que ayuden en la toma de decisiones de una empresa. La parte fundamental para aplicar una red neuronal consiste en las entradas de la misma; donde las variables independientes deben ser relevantes dentro de la investigación con relación a la variable dependiente fruto de la investigación.

Existen tres tipos principales de redes neuronales de acuerdo al tipo de aprendizaje utilizado (Supervisado, No Supervisado y Por Corrección), cada una de ellas es aplicable en diferentes casos dependiendo de las necesidades y los objetivos.

2.9.5.1 Redes neuronales supervisadas y por corrección

Requieren un grupo de datos de control contra los cuales se debe verificar los resultados obtenidos con el fin de generar conocimiento, debido a las características del proyecto en el cual no se conocen los resultados esperados no se consideran a las técnicas supervisadas las más adecuadas.

Se utilizará la red neuronal por corrección debido a que se tiene variables independientes, que servirán como datos de entrada, y disponiendo de los datos de salida como objetivo a cumplir; los cuales permitirán ir comparando los resultados de

salida obtenidos y a la vez se cambiarán los pesos en función de la diferencia que se obtenga en los resultados proyectados.

Las técnicas no supervisadas no requieren datos de control para verificar los resultados obtenidos debido a que considera todos los datos de entrada como variables aleatorias a partir de las cuales puede generar conocimiento.

Entre las técnicas no supervisadas más usadas que cuentan con más soporte, se encuentra los mapas auto-organizativos o red neuronal de Kohonen. Esta es una arquitectura de sistemas de clasificación que permite encontrar patrones con características comunes en una población determinada. (Gutierrez E. , 2014).

2.9.5.2 Redes neuronales artificiales

Las redes neuronales artificiales son sistemas que procesan la información, están basadas su estructura y funcionamiento en las redes neuronales biológicas. Su estructura está dada por los nodos que no son más que un conjunto de elementos simples, estas son las neuronas que se conectan mutuamente por medio de un peso que es un valor numérico que puede ser modificado. Estas redes neuronales tienen una semejanza muy parecida al cerebro humano, por ejemplo los dos tipos de neuronas pueden aprender de casos anteriores por experiencia o por casos nuevos y mediante la observación de características relevantes a partir de una gran cantidad de entradas que representan dentro de la investigación una parte muy importante; para nuestro estudio se tomará como referencia información de salarios anteriores, y compararándolos con los actuales luego de la predicción. (Gutierrez E. , 2014)

2.9.6 Regresión Lineal Simple

Es una forma de regresión lineal simple que consiste de una variable X conocida como independiente y una variable dependiente Y; con lo que la ecuación que se determina es $Y=bx+a$; a través de esta se puede modelar datos de una organización cuyo objetivo sea el generar predicciones hacia el futuro como hacia atrás de un evento motivo de una investigación. En la ecuación se asume que a y b se considera que son los coeficientes de regresión que especifican la intersección con el eje de ordenadas, y la pendiente de la recta, respectivamente. Estos factores a y b se calculan utilizando el procedimiento de los mínimos cuadrados que disminuyen el error entre los datos reales y la valoración de la línea. (Gutierrez E. , 2014).

El soporte matemático es: $Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$

2.9.7 Regresión Lineal Múltiple

Es una ampliación de regresión lineal que implica más de una variable predictora, y permite que la variable a obtener como solución Y, sea programada como una función lineal de un vector multidimensional. (Gutierrez E. , 2014)

2.9.8 Regresión no lineal

En muchas ocasiones los datos no muestran una dependencia lineal. Esto es lo que sucede si, por ejemplo, la variable respuesta depende de las variables independientes según una función polinómica, dando lugar a una regresión polinómica que puede planearse agregando las condiciones polinómicas al modelo lineal básico. (Gutierrez E. , 2014)

2.10 Patrones de información

El objetivo de la minería de datos es obtener modelos de datos, para lo que ha de averiguar qué tipo de moldes pueden ser encontrados utilizando las técnicas de minería de datos; estas pueden ser descriptivas o predictivas.

- Tareas descriptivas caracterizan las propiedades generales de los datos en una base de datos.
- Tareas predictivas realizan supuestos en los datos generados con el fin de poder generar predicciones.

A continuación se describe los patrones y sus funcionalidades que pueden ser obtenidas mediante la utilización de minería de datos.

Descripción de clases: La idea de trabajar con este tipo de modelos es poder asociar datos en clases o conceptos. Hay tres formas de encarar este enfoque.

- Caracterización de datos se basa en realizar una sumariación de las características generales de una clase en particular de datos. Los resultados de este tipo de análisis suele presentarse en la forma de reglas de caracterización.
- Discriminación de datos es una comparación entre las características generales de los objetos de una clase respecto a las de otro conjunto contrastante. En este caso se habla de reglas de diferenciación.
- Combinación de ambas utilizar conjuntamente la caracterización con la diferenciación.

Análisis de asociación: Es el descubrimiento de reglas de asociación que muestran condiciones del tipo atributo-valor que ocurren con frecuencia dentro de un conjunto de datos. **Clasificación y predicción:** Por medio de este proceso se buscan modelos o funciones que distingan diferentes tipos de datos para describirlos con el objetivo de predecir un determinado número de objetos, de que clase son y sus características detalladas.

Análisis de clusters: Trata de agrupar variables a investigar de acuerdo a grupos semejantes, de esta forma intenta determinar el comportamiento que tienen cada grupo a fin de poder tomar decisiones que ayuden en una investigación. Por esta razón se dice que esta técnica se clasifica como aprendizaje no supervisado.

Análisis de infrecuentes: Una gran volumen de información puede contener datos que no cumplen con el comportamiento general que tiene la información o el patrón de los datos. Esos datos son los infrecuentes; la mayoría de los métodos de minería de datos descartan estos datos como si se trataran de ruido o anomalías.

Análisis progresivo: Detalla y ajusta la periodicidad o la tendencia que poseen determinados datos que tienen un comportamiento variable en el tiempo. Ha de entenderse que los métodos de minería de datos generan modelos más de lo que se espera; lo que da lugar a una variedad de reglas que serán motivos de investigación ya que no todo lo que se genera arroja resultados que sirven en el proceso de investigación por lo que se deberá analizar y determinar cuáles son los que realmente sirven a la organización a fin de poder tomar decisiones adecuadas.

Un modelo es atractivo si cumple con las siguientes realidades:

- Es sencillamente descifrable para las personas.
- Es legal, proporciona confianza y seguridad, para trabajar con otro grupo de datos sean estos nuevos o que se haya tomado como prueba en la investigación.
- Tiene una ventaja permitida.
- Formula un conocimiento impactante.

Un modelo también será útil si sirve para aprobar una suposición que el interesado pretende confirmar. Se puede sacar como conclusión que los puntos señalados con anterioridad son en gran medida individuales ya que dependen de los datos proporcionados y que dispone el beneficiario de los mismos.

Para determinar con mayor exactitud y precisión si un modelo es útil, es necesario crear medidas que permitan controlar qué variables se está utilizando en la investigación, a fin de determinar si un modelo es o no útil; ha de analizarse la estructura de los modelos encontrados y la estadística que permitió generar dicho modelo. Se puede establecer que las reglas de asociación utilizan el soporte y la confianza como medidas tomadas en forma imparcial. (Dandretta, 2013)

Los científicos deben pasar mucho tiempo estudiando los datos crudos³, para así conocerlos en su totalidad, ver como lucen, y ver que cubren estos datos y que no. Deben entender la estructura, calidad y que campo envuelven los datos.

Es por esto que se aplican métodos estadísticos y de minería de datos como reglas de asociación, descubrimiento de modelos secuenciales, clustering y reglas de clasificación, para detectar modelos interesantes de información. Se trata de un grupo de técnicas orientadas a evaluar las semejanzas y las diferencias que se pueden dar en la información.

2.11 Inteligencia artificial y técnicas de minería de datos

Se define a la inteligencia artificial como un conjunto combinado de varias ciencias de la computación, filosofía y fisiología, estos agrupan varios campos como son los sistemas inteligentes, los sistemas expertos e incluido la cibernética, el común de los denominadores de estos sistemas son la creación de máquinas que tienen la capacidad de pensar si fuera esta la palabra correcta casi como los humanos.

El objetivo principal es crear una máquina que pueda realizar trabajos parecidos o iguales a los realizados por el hombre, son un gran número de aplicaciones que utilizan estas máquinas como lo son en el ámbito de la medicina, ingeniería, idiomas, juegos, realidad virtual, esto da lugar a que esta ciencia sea una de las más confiables en cada una de las ramas antes expuestas.

Por esta razón los sistemas que administran las bases de datos son cada vez más sofisticados y tienen mejores funciones y características, la estructuración de la información y el desarrollo de nuevos algoritmos para insertar datos, borrarlos y localizarlos son más fiables, el crear máquinas que realizan el trabajo de los seres

³ Datos crudos: son aquellos datos que fueron almacenados en una base de datos y que no han tenido ningún tipo de procesamiento antes de ser analizados.

humanos es una clara muestra de que cada vez es más próxima a la misma inteligencia humana. (Mamani, 2014)

2.11.1 Redes neuronales artificiales en minería de datos

Las redes neuronales artificiales han sido entrenadas para la realización de funciones complejas en variados campos de aplicación. Hoy en día pueden ser entrenadas para la solución de problemas que son complejos para sistemas computacionales comunes o para el ser humano.

La idea de las redes neuronales artificiales fue imaginada originalmente como un intento de modelar el cerebro humano, esto es, entender y explicar cómo funciona y opera el cerebro. La meta es crear un modelo capaz en emular el proceso humano de reflexión y comprensión del suceso que está siendo investigado. La mayor parte de los trabajos iniciales en redes neuronales fue realizada por fisiólogos y no por ingenieros.

Una característica importante de las redes neuronales artificiales es que son altamente flexibles al ruido y fuertes frente a fallos estructurales: la eliminación o mal funcionamiento de un porcentaje importante de unidades de proceso no provoca un “colapso”, sino una disminución progresiva en el rendimiento de la red, características que posee el cerebro humano. Sin embargo, la característica más valorada de las redes neuronales artificiales es su capacidad de aprendizaje. La claridad de las redes neuronales artificiales y su poder de predicción tienen una relación estrecha, que se comporta de manera inversa, entre más sencilla sea la forma del modelo, más fácil será su comprensión, pero tendrá menor capacidad para encontrar diferencias sutiles o demasiado variadas. (Aluja, 2013)

De la investigación se puede decir claramente como las técnicas de minería de datos ayudan a resolver problemas empresariales, en nuestro estudio a descubrir patrones en las remuneraciones del personal que labora en el (E.E). Una de las grandes ventajas al manejar una gran cantidad de información con minería de datos y redes neuronales es que se puede ayudar en varios ámbitos uno de ellos es el ámbito científico donde se puede resolver varios problemas, utilizando los algoritmos estadísticos ya implementados en la minería de datos que puede ayudar a realizar análisis de datos para la toma de decisiones a niveles de altos mandos, en este trabajo de investigación se procedió a generar nuestro propio modelo de redes neuronales para la predicción de sueldos utilizando minería de datos.

2.12 Limpieza de datos

Dentro de una determinada base de datos no siempre estos están correctos o completos, para resolver estos problemas se debe llenar los datos faltantes, arreglar los datos con ruido o dañados, estos datos faltantes o mal ingresados pueden dañar el proceso de minería de datos, dando como resultado información no confiable. A continuación se muestra una cantidad de métodos para la limpieza de los datos:

Para valores faltantes se tiene lo siguiente:

- Ignorar la tupla: Esto se da cuando el valor que falta es el que caracteriza a la tupla.
- Completar el valor manualmente: Esto se puede dar solamente en algunos casos puntuales porque para solucionar este inconveniente se requiere demasiado tiempo.
- Usar una constante para completar el valor: Este método permite llenar los valores faltantes con un valor constante pero que tan eficiente es esto, por ejemplo si faltase un valor de tiempo en el grado de varios militares y se carga con una constante no es conveniente porque no todos tienen el mismo valor en ese campo y esto provocaría que la minería de datos no sea exitosa por lo tanto la información emitida por esta no sea confiable.
- Usar el promedio del atributo para completar el valor.
- Usar el promedio del atributo de los elementos de la misma clase para completar el valor: Es similar al anterior pero utilizando algún criterio para definir diferentes clases de tuplas.
- Usar el valor más probable para completar el valor: Se puede utilizar varios métodos estadísticos para llenar estos valores faltantes como son los formulismos bayesianos, arboles, regresión, estos pueden dar un valor muy aproximado a la realidad.

Ruido: El ruido es un error aleatorio que se produce en una variable medida.

- Vecinos: Se debe realizar un promedio entre tuplas vecinas en cuanto a los valores locales, esto permite llenar con valores aproximados la información faltante, es como el método que se utiliza para el arreglo de fotografías.
- Clustering: Esta técnica agrupa y separa los datos en conjuntos, si algún dato no cae en alguno de estos grupos esta técnica simplemente la toma como ruido dentro de la información.

- Inspección automática y manual: Se buscan los infrecuentes con algoritmos combinados con revisión humana (valores no esperados).
- Regresión: Consiste en alisar los datos haciendo que se ajusten con una función.
- Inconsistencias: Para corregir estos errores se puede utilizar herramientas específicas como son las de ingeniería del conocimiento en la que se detectan restricciones sabidas de los datos utilizados, es posible también detectarlas manualmente utilizando una referencia externa.

2.13 Diferencia entre modelo y metodología

Se puede especificar que algunos modelos se los conoce como metodologías pero en realidad son un conjunto de actividades que permiten realizar un determinado trabajo siguiendo un conjunto de tareas específicas, se puede decir que la diferencia entre método y metodología es que la metodología indica cómo se deben llevar a cabo las tareas específicas en un proceso determinado y el método especifica que hacer en un determinado proceso.

La definición específica de metodología vendría a ser el conjunto de actividades específicas que debe tener un proceso para llevar a cabo un determinado trabajo, para cada actividad se debe tomar en cuenta la forma en la que se llevara adelante así como también sus entradas y salidas que son muy importantes para cumplir con el objetivo.

2.13.1 Metodología KDD (proceso de extracción de conocimiento)

El hallazgo de conocimiento en bases de datos; constituye encontrar información válida siguiendo un proceso secuencial compuesto por distintas etapas y fases que van desde la preparación de los datos hasta la interpretación y propagación de los resultados.

Los patrones obtenidos de la investigación deberán ser válidos para generar proyecciones con nuevos datos, aportaran nuevo conocimiento a la aplicación en estudio y serán útiles para el usuario final en la toma de decisiones.

KDD es un proceso repetitivo y participativo. Repetitivo ya que la salida de alguna de las etapas puede retroceder a pasos anteriores; esto se debe porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad. Es participativo porque generalmente un experto en el dominio del problema, debe ayudar a la preparación de los datos y validación del conocimiento extraído.

El modelo de proceso KDD se resume en las siguientes fases:

- Selección de los datos sobre los que se trabajará.
- Pre-procesamiento de los datos, donde se realiza la validación de los datos incorrectos y ausentes a fin de obtener información que ayude a realizar la investigación con datos de calidad que realmente ayuden en la exploración de la información.
- Transformación de los datos y reducción de la dimensión de la base de datos a fin de quedarse con lo que realmente es necesario para el proceso de investigación.
- Minería de datos permite obtener patrones de datos a fin tomar decisiones
- Interpretación y evaluación del nuevo conocimiento de acuerdo a la información presentada en el proyecto de investigación.

A continuación se resume el proceso KDD (ver figura 4).

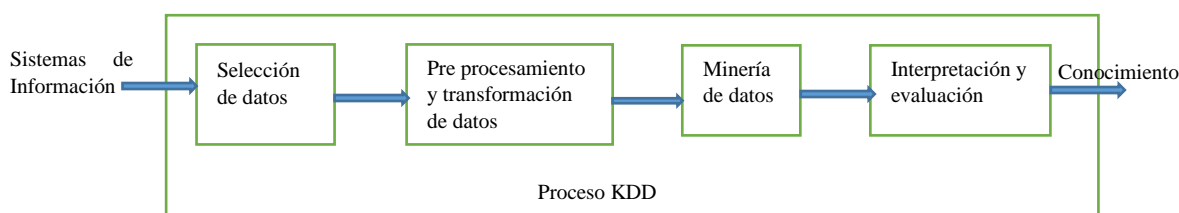


Figura 4: Esquema del proceso KDD resumido

Fuente: (Hernández Orallo, 2014)

Se puede establecer que en la etapa de minería de datos es donde se realiza la extracción de patrones a partir de los datos validados previamente según la investigación que se esté desarrollando. El término KDD y minería de datos se utilizan indistintamente para hacer referencia al proceso completo de descubrimiento de conocimiento.

Fayyad establece nueve etapas para llevarlo a cabo el proceso anterior. (Moine, 2013)

- Conocimiento de la aplicación a investigar

En esta etapa se debería recolectar toda la información disponible especialmente en bases de datos históricas y relevantes sobre la aplicación motivo de la investigación e identificar los objetivos del proceso que persigue la indagación desde el punto de vista del usuario. Este se considera un punto importante donde se debe explorar cada uno

de los datos de la empresa motivo del estudio; ya que de este dependerá el futuro de la investigación. (Moine, 2013)

- Creación del conjunto de datos

Esta etapa consiste en la elección de las fuentes de datos que se utilizarán en la investigación, la integración de las mismas y la selección de los atributos que conformarán la base de datos que ayudaran a explotar la información oculta. En este paso podría requerirse la construcción de un almacén de datos⁴ donde se almacene los datos necesarios para la investigación. (Moine, 2013)

- Limpieza y pre-procesamiento de los datos

En esta fase se deberían llevar a cabo tareas como limpieza de ruido o datos absurdos y tratamiento de datos faltantes a fin de obtener una base de datos útil para la investigación. (Moine, 2013)

- Reducción y proyección de los datos

En este paso se descubren características útiles de representación de los datos dependiendo del objetivo de la tarea de minería; se puede determinar las variables independientes que aportaran en el proceso de investigación. Se incluye la utilización de técnicas de reducción de la dimensionalidad de acuerdo al peso que tiene tomando como referencia la totalidad de la base de datos motivo de la investigación; lo que permite reducir la cantidad de variables en discusión o para encontrar representaciones que aporte al proyecto de investigación. En esta etapa es habitual la transformación de los datos, calculando nuevos atributos o bien determinando los existentes con otro formato. (Moine, 2013)

- Determinar la tarea de minería de datos

En esta etapa se determina cual será la tarea de la minería de datos con la que se realizará el estudio tales como agrupamiento, regresión, clasificación, o asociación; teniendo en cuenta los objetivos determinados en la etapa, una vez analizado cada uno de las variables dependientes motivos de la investigación y de acuerdo a los objetivos planteados. (Moine, 2013)

- Determinar el algoritmo de minería

⁴ Almacén de datos, o D.W, es un repositorio de fuentes heterogéneas de datos, integrados y organizados bajo un esquema unificado para facilitar su análisis y dar soporte a la toma de decisiones.

Una vez realizado la investigación con varios algoritmos y analizados los resultados que proyectan se llega a determinar qué modelo o modelos y parámetros se aplicarán para la búsqueda de patrones sobre los datos que se está investigando de una empresa u organización. Realmente es fundamental analizar los resultados que arrojan los modelos ya que estos serán los que permitan tomar decisiones. (Moine, 2013)

- Minería de datos

En esta etapa se aplican los algoritmos y técnicas seleccionadas al conjunto de datos en búsqueda de los patrones de interés de la investigación para lo que previamente se debió realizar un estudio detallado de los resultados generados. (Moine, 2013)

- Interpretación

Comprende la interpretación de los patrones encontrados, generando gráficos y traduciendo los mismos en términos comprensibles por el usuario; del análisis realizado dependerá la toma de decisiones. (Moine, 2013)

- Utilización del nuevo conocimiento

En esta fase se implementa el conocimiento descubierto, apoyando con el mismo la toma de decisiones o bien informando a las partes interesadas. Incluye la verificación y resolución de potenciales conflictos con conocimiento descubierto previamente; si todo el proceso KDD se llevó en forma minuciosa el nuevo conocimiento será realmente útil para la empresa u organización; debido a que realmente será una fuente de información para la toma de decisiones. (Moine, 2013)

Si bien KDD define las fases generales del proceso de minería de datos, no especifica qué actividades puntuales hay que realizar en cada una, quedando la definición de las mismas a criterio del equipo de trabajo; esto permite al investigador realmente adaptarse de acuerdo a como vaya desarrollando el proceso de extracción del conocimiento sobre el tema que sea motivo de su investigación.

CAPÍTULO 3

PROPUESTA MODELO DE REMUNERACIONES Y MINERÍA DE DATOS

3.1 Fuentes de información para el análisis de las remuneraciones

3.1.1 Organigrama Estructural

La estructura orgánica de la dirección de personal militar del E.E, está conformada por los siguientes niveles: Directivo, asesoramiento y apoyo, y un nivel operativo, como se detalla a continuación (ver figura 5).

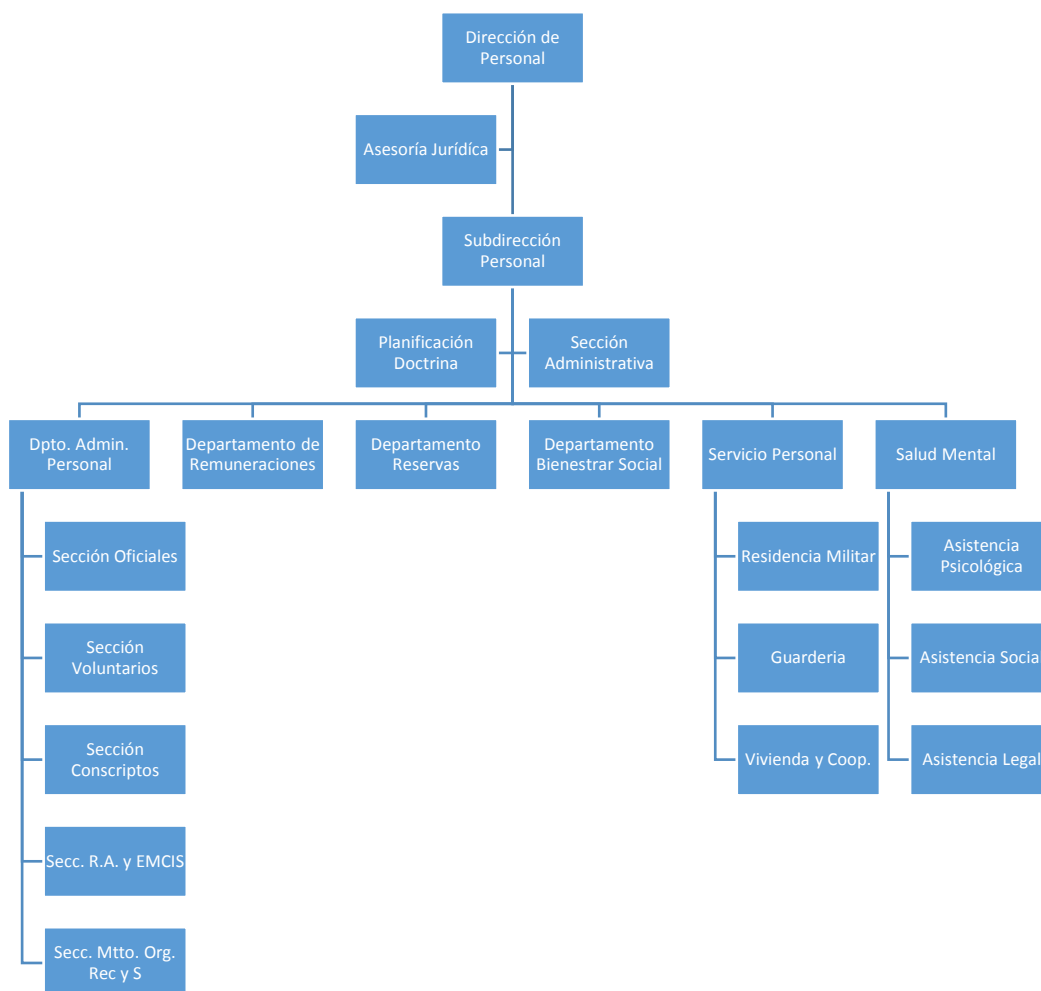


Figura 5: Organigrama estructural de la dirección del personal del E.E

Fuente: (C.G.E.E, 2015)

3.1.2 Proceso de remuneraciones del personal del (E.E)

Rol de pagos. El proceso de rol de pagos determina todos los rubros de ingreso y egreso que tiene el personal del (E.E) consultando los valores de los parámetros establecidos para calcular el sueldo, rancho o refrigerio, compensaciones, los descuentos de ley, impuesto a la renta y los descuentos varios; registrando el valor de cada rubro de acuerdo a su codificación. Una vez establecido el rol de pagos definitivo se procede a cerrar el mes, este subproceso actualiza los estados de todos los descuentos e indica al administrador de procesos que se ha cerrado el rol, para que no se pueda ejecutar. Entonces se imprime: la liquidación individual del personal en la forma pre impresa, descuentos varios, retenciones judiciales, aportes del instituto ecuatoriano de seguridad social (I.E.S.S) y el instituto de seguridad social de las fuerzas armadas (ISSFA), aportes de cesantía, comprobante de pago de sueldos, auditoría de sueldos, gastos de representación, balance, cooperativas, anticipo de sueldo. (C.G.E.E, 2015)

Descuentos. De acuerdo a los reglamentos establecidos se ingresan al sistema de forma manual o mediante medios magnéticos datos acerca de: descuentos por prestaciones del IESS e ISSFA, retenciones judiciales, descuentos ordenados por el (E.E), descuentos de las unidades militares, poderes voluntarios, cooperativas y casas comerciales con autorización (E.E).

Una vez que se ha ejecutado el proceso de rol de pagos y se ha efectuado el cierre de mes, se procede a exportar los datos de descuentos ingresados con sus respectivas novedades, para que las entidades financieras, comerciales, judiciales y de bienestar social, confirmen en sus respectivos sistemas estos descuentos. (C.G.E.E, 2015)

Rubros anuales. El proceso de remuneraciones anuales determina los rubros de ingreso y egreso que tiene el personal del (E.E) por concepto de decimotercero, decimocuarto, compensaciones al personal militar, en los meses correspondientes de acuerdo a los reglamentos de remuneraciones para el personal del (E.E). (C.G.E.E, 2015)

Liquidaciones. El proceso de liquidaciones determina los valores proporcionales por concepto de decimotercero, decimocuarto, compensaciones, que se paga al

personal que se encuentra con la baja del (E.E), este personal puede ser militar o civil y puede ser pagado una sola vez⁵. (C.G.E.E, 2015)

Retroactivos. El proceso de retroactivos determina los rubros proporcionales de ingreso y egreso del personal para una fecha anterior determinada por cambios en la base de cálculo. (C.G.E.E, 2015)

Reincorporaciones. El proceso de reincorporaciones determina los rubros de ingreso y egreso que se deben calcular entre las fechas de salida y reingreso del personal militar del (E.E). Para realizar las cancelaciones de los valores correspondientes entre estas fechas. (C.G.E.E, 2015).

Remuneraciones en el exterior. El proceso de remuneraciones en el exterior determina todas las asignaciones que percibirá el personal que sale al exterior por razones de servicio. Para calcular dichas asignaciones se requiere de los parámetros establecidos de acuerdo a la función y la jerarquía del personal, y los parámetros de coeficiente de costo de vida. Una vez realizado el cálculo se presenta un listado para realizar el pago a los países en donde se encuentra el personal. (C.G.E.E, 2015).

Desvinculación Laboral. El proceso de desvinculación laboral se calcula al personal de generales, coroneles, sub oficiales mayores y suboficiales primeros que se retiran del (E.E) una vez que han cumplido todo el tiempo en el grado especificado, los valores para el cálculo de este rubro está dado por la tabla especificada por el ministerio de relaciones laborales. (C.G.E.E, 2015).

3.1.3 Procedimientos de remuneraciones del personal del (E.E)

- Estudio salarial
- Cálculo de remuneraciones
- Cálculo de remuneraciones al exterior
- Liquidación salarial por bajas, cesaciones o reincorporaciones por fallo judicial.

Estudio salarial se encarga de realizar estudios salariales en forma permanente para llegar a determinar si las remuneraciones actuales se ajustan a las necesidades del personal militar, reserva activa y empleados civiles de acuerdo al costo de vida.

Es responsable de ejecutar este procedimiento la sección de remuneraciones del departamento de estímulos y remuneraciones de la dirección de personal en

⁵ Comandancia General del (E.E). (10 de 5 de 2015). Manual de remuneraciones

coordinación con la sección de planificación e ingreso, estado mayor planificador, la dirección de finanzas y la dirección de bienestar social del (E.E). (C.G.E.E, 2015)

Cálculo de remuneraciones se procede a ingresar la orden general en donde consta la nómina del personal con altas y bajas, reincorporaciones, licencias sin sueldo, disponibilidades; modificaciones en índices o parámetros de cálculo; además la información relacionada con las retenciones judiciales, ISSFA, IESS, FONIFA, unidades militares, asociaciones, cooperativas, promociones y otras entidades. Las principales salidas de este proceso son la actualización de los índices y parámetros de cálculo, además el ingreso de datos de los descuentos voluntarios al sistema de remuneraciones, con esta información se genera el cálculo de bonos y décimos, roles de pagos y los siguientes reportes; balance general, listado de auditoría, valor líquido a recibir por unidades, liquidaciones individuales, listado de retenciones judiciales y poderes voluntarios, aportaciones y préstamos al ISSFA, fondo de reservas para militares y EMCIS, impuesto a la renta, aportaciones y préstamos al IESS, bonificaciones por ascensos, etc. (C.G.E.E, 2015)

Cálculo de remuneraciones al exterior en este procedimiento la entrada es la orden general en la cual consta la nómina del personal que debe recibir las asignaciones económicas, para lo cual la sección de remuneraciones calcula los gastos de compensación, residencia, representación, otros ingresos y emite reportes de las asignaciones al exterior a la sección administrativa de la dirección de personal la cual elabora la orden de gasto. (C.G.E.E, 2015)

Liquidación salarial por bajas, cesación o reincorporaciones por fallo judicial el desarrollo de este procedimiento de liquidación conlleva a revisar de antemano las ordenes generales y determinar la nómina de personal dado de baja o cesante, con esta información la sección de remuneraciones calcula los componentes de sueldo y adicionales y determina el valor a pagar por liquidación, el mismo que es enviado a la sección administrativa para la elaboración de la orden de gasto respectiva.

El desarrollo del procedimiento de reincorporaciones tiene como principal entrada la verificación de personal reincorporado por fallos judiciales en las ordenes generales, en donde la sección de remuneraciones calcula el valor a pagar tomando en cuenta los derechos de reconocimiento económica y beneficios de ley, para que posteriormente

la sección administrativa elabore la orden de gasto y la dirección de finanzas reincorpore la nómina de personal en los roles mensuales. (C.G.E.E, 2015)

3.2 Diagramas de flujo de procedimientos de remuneraciones

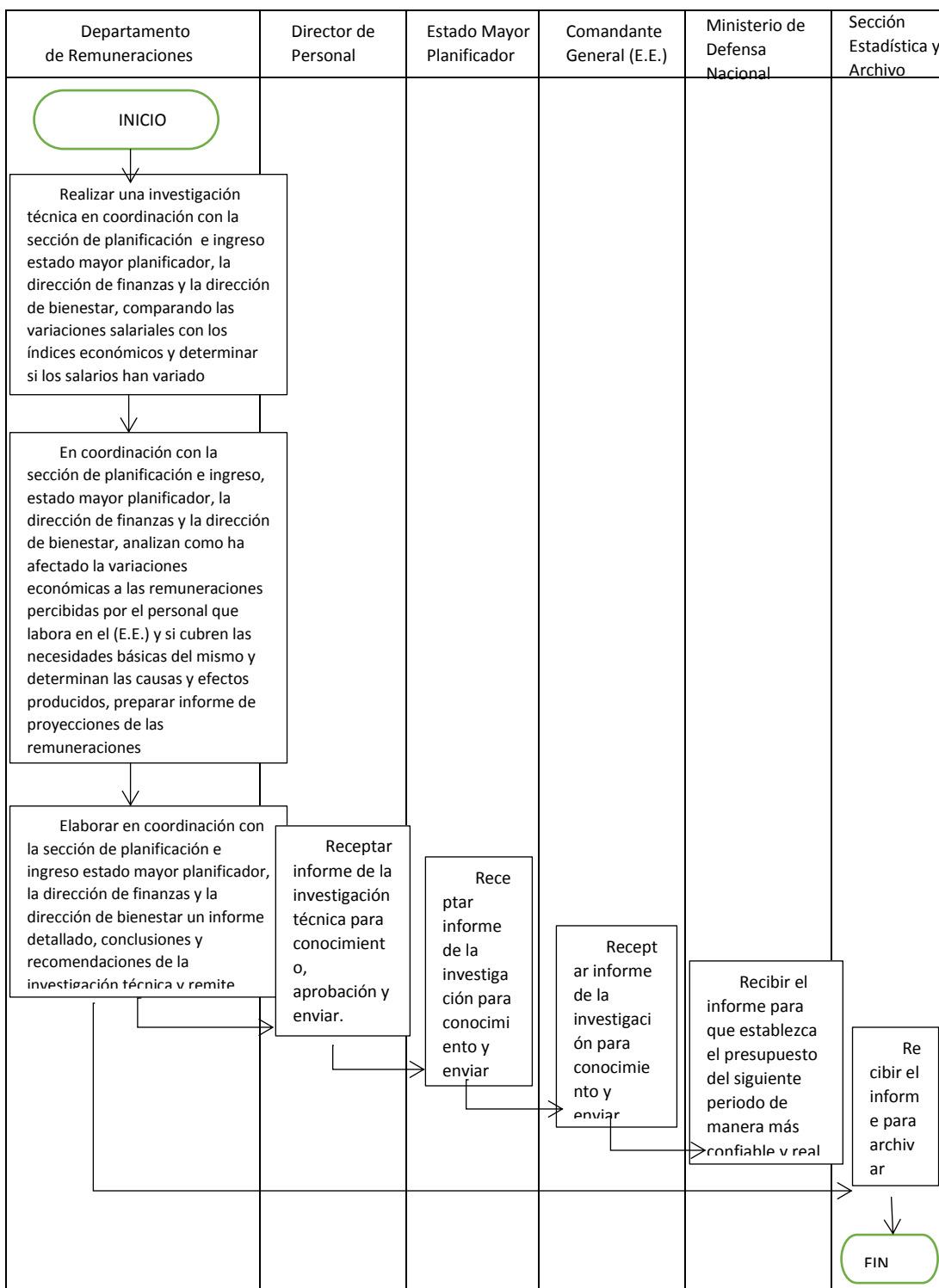


Figura 6: Diagrama de flujo estudio salarial

Fuente: (C.G.E.E, 2015)

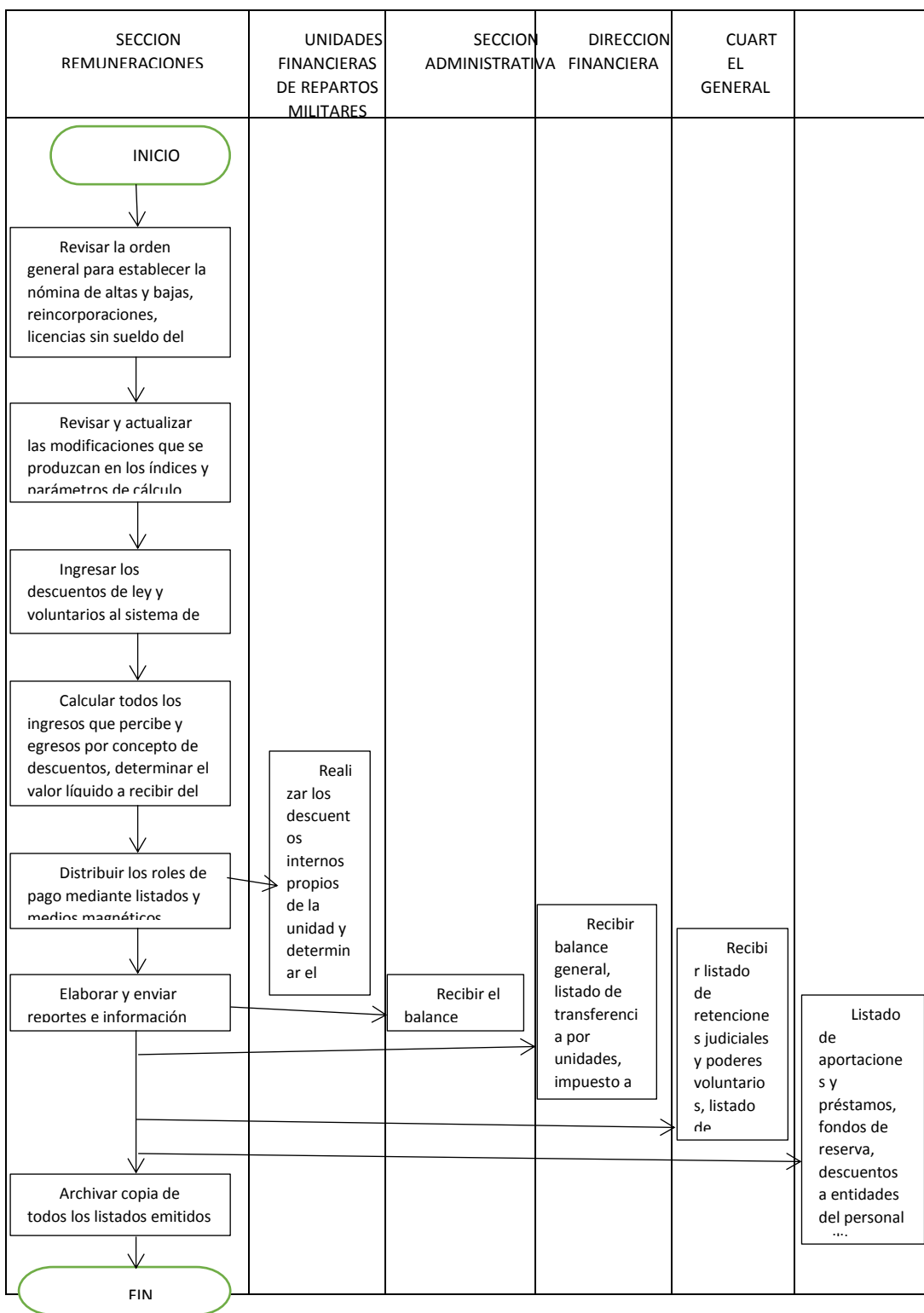


Figura 7: Diagrama de flujo cálculo de remuneraciones

Fuente: (C.G.E.E, 2015)

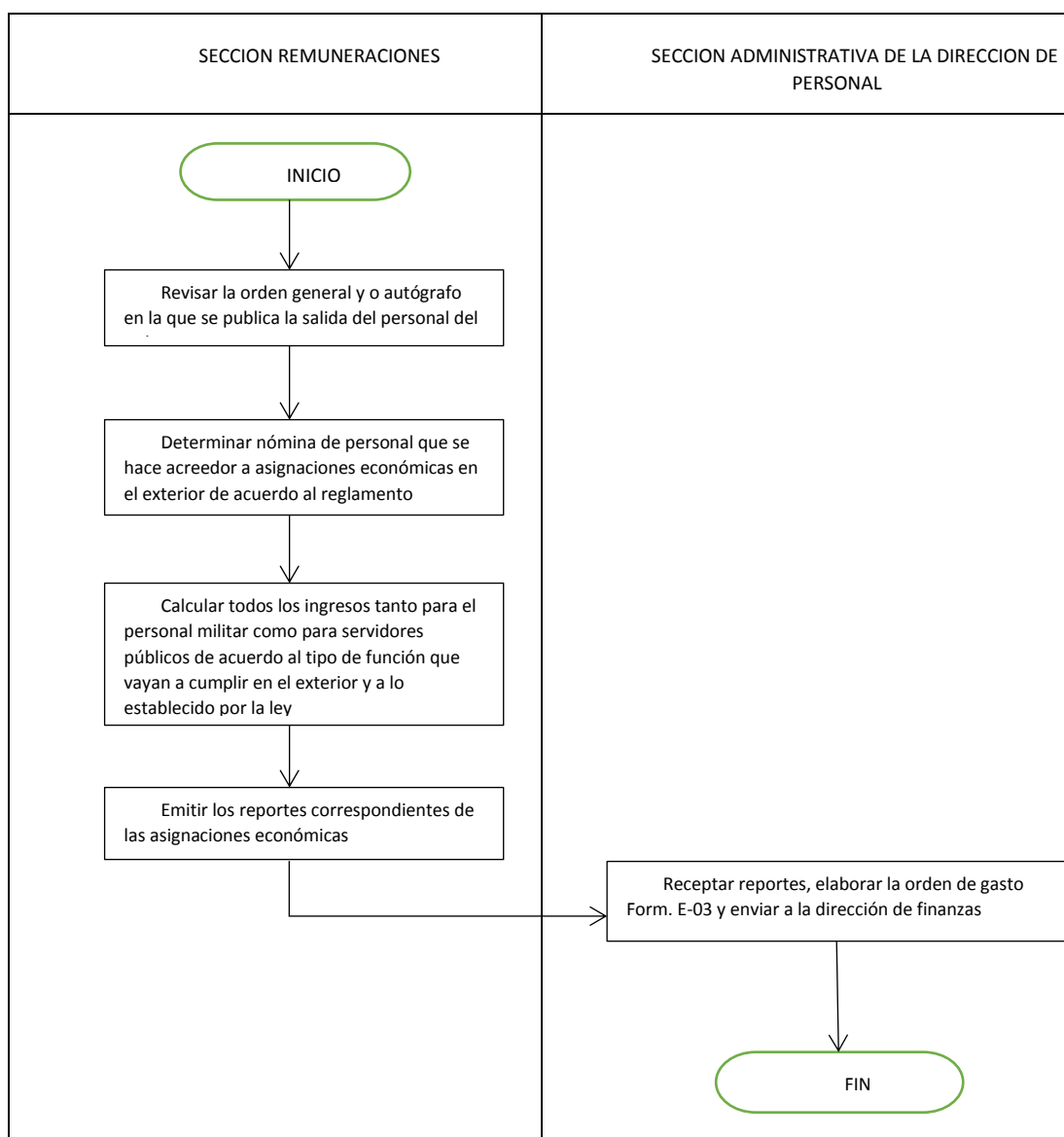


Figura 8: Diagrama de flujo cálculo de remuneraciones al exterior

Fuente: (C.G.E.E, 2015)

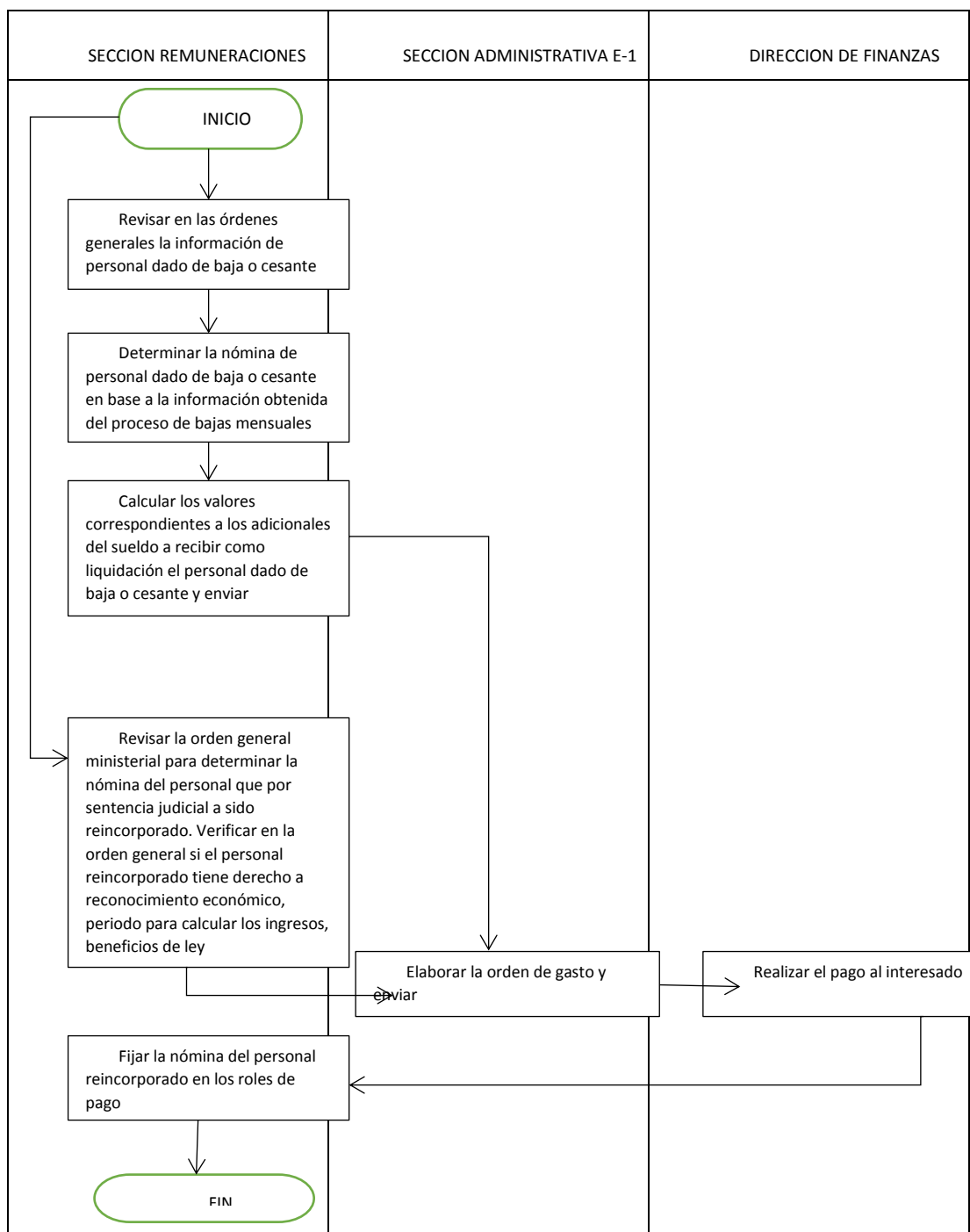


Figura 9: Diagrama de flujo liquidación salarial

Fuente: (C.G.E.E, 2015)

3.3 Arquitectura jerárquica del proceso de remuneraciones

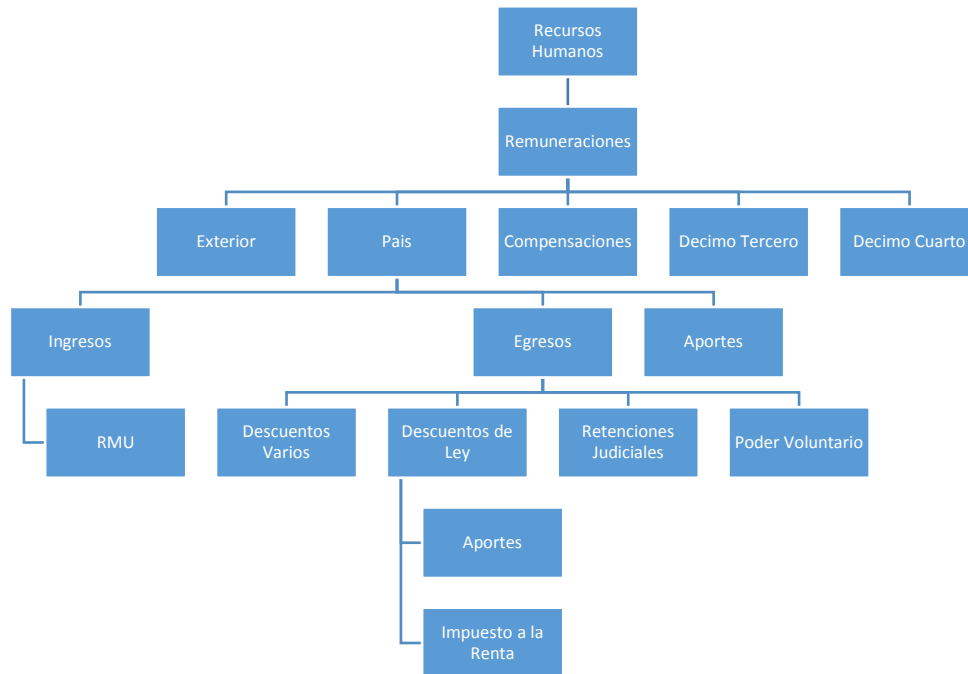


Figura 10: Jerarquía de remuneraciones

Fuente: (C.G.E.E, 2015)

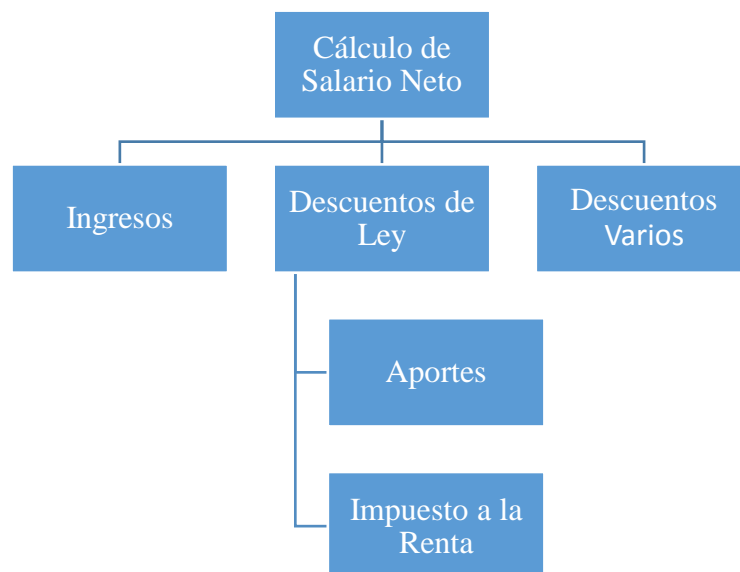


Figura 11: Jerarquía de procesos del salario neto

Fuente: (C.G.E.E, 2015)

3.4 Conocimiento disponible modelo entidad relación remuneraciones

ROL DE PAGOS, CONTROL DE PROCESO Y NOVEDADES

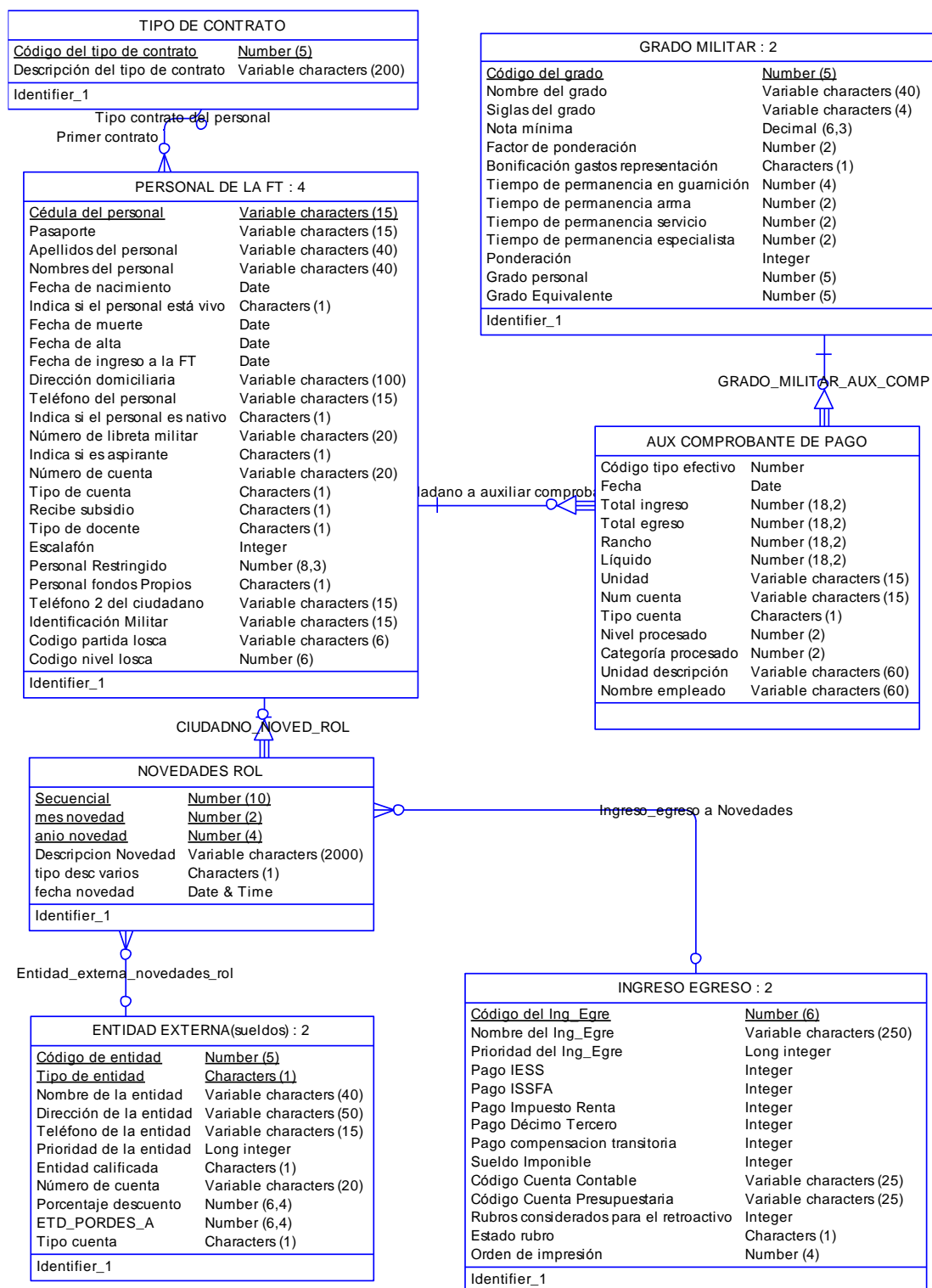


Figura 12: Rol de pagos, procesos y novedades

REMUNERACIONES

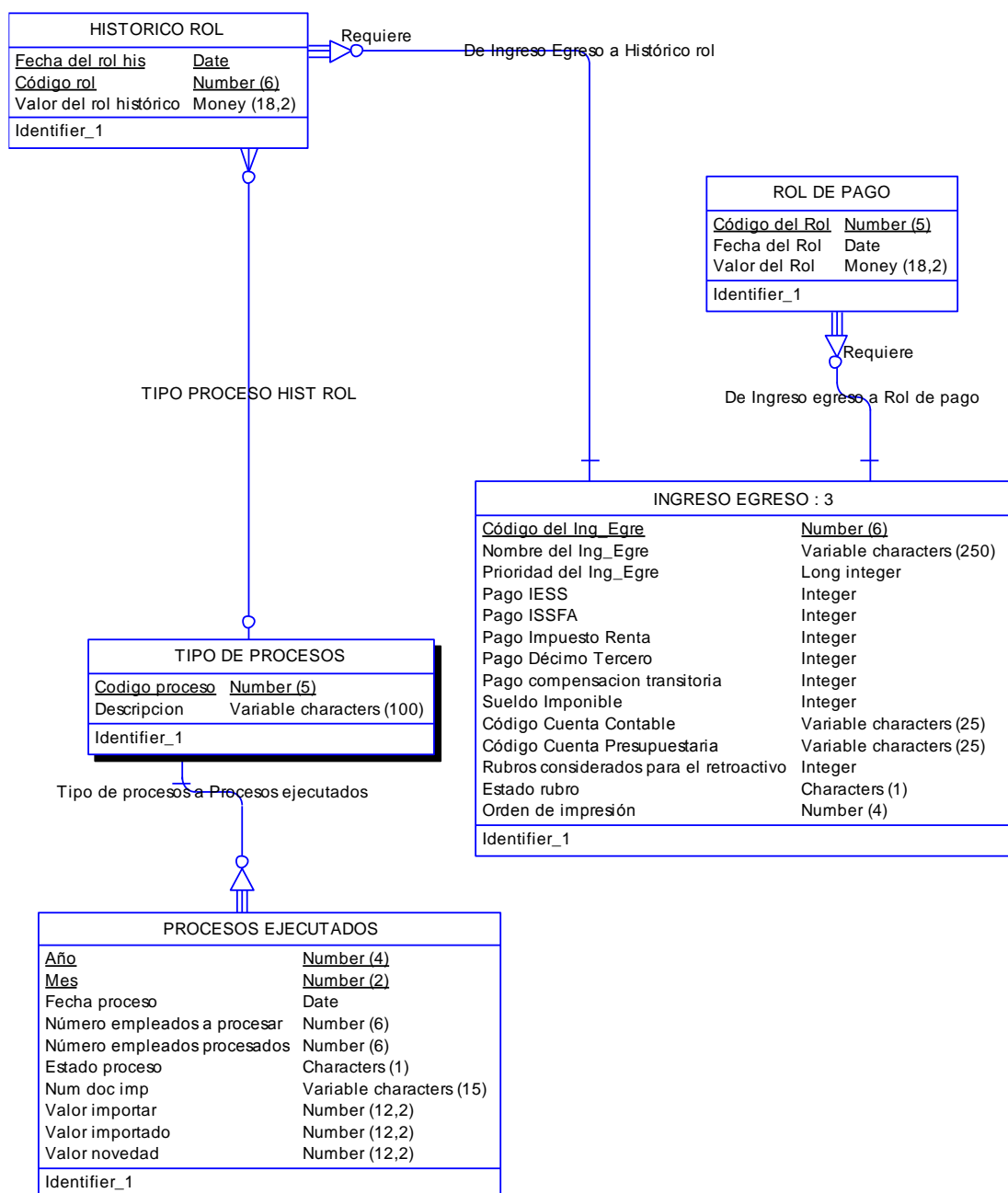


Figura 13: Descuentos, rubros, préstamos y ahorros

DESCUENTOS: RUBROS FIJOS Y VARIABLES, PRESTAMOS Y AHORROS

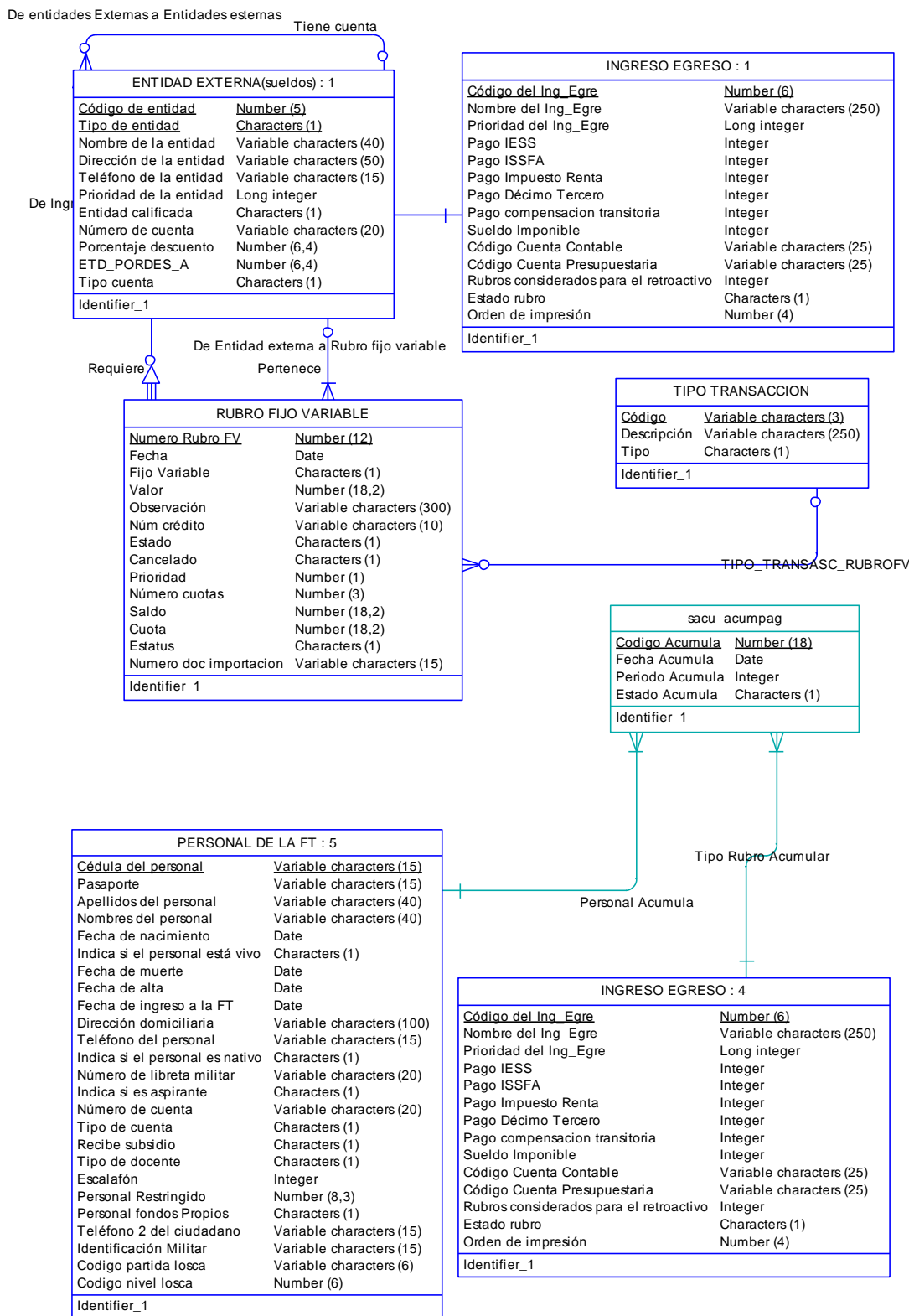


Figura 14: Poder voluntario y retenciones

3.5 Desarrollo del proceso ETL

3.5.1 Extracción de Datos

El E.E cuenta con una base de datos relacional dentro de la cual se encuentra la tabla histórica de remuneraciones, que es de donde se ha obtenido la información de los años 2007 al 2015 para poder realizar las proyecciones salariales de los años 2016 al 2018.

3.5.2 Análisis de la tabla historial del rol de pagos (shrl_hisro)

Se ha utilizado la herramienta Talend Open Studio for Data Quality 6.2.1, este aplicativo permite realizar diferentes tipo de análisis dentro de una base de datos, en esta ocasión se realizara el análisis de la tabla shrl_hisro que es la que tiene la información histórica que se va a analizar (ver figura 15).

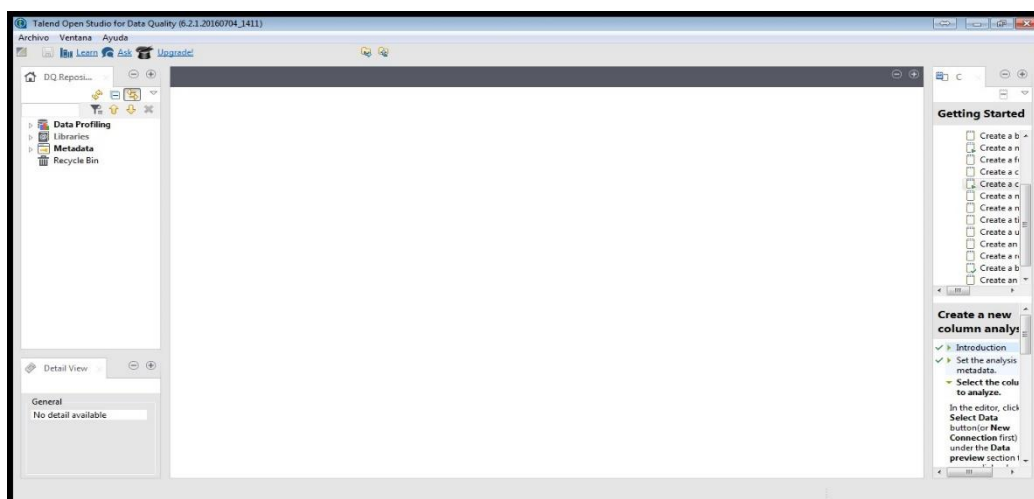


Figura 15: Ventana principal de talend open studio for data quality

El análisis de columnas es lo que se utilizará para saber en qué estado se encuentra cada uno de los campos de la tabla shrl_hisro, ayudándose de las estadísticas arrojadas se podrá utilizar la herramienta RapidMiner para, si es necesario, suprimir o completar un determinado dato en la tabla especificada (ver figura 16).



Figura 16: Ventana inicio análisis de columnas

En primer lugar se debe crear un nuevo análisis (ver figura 17).

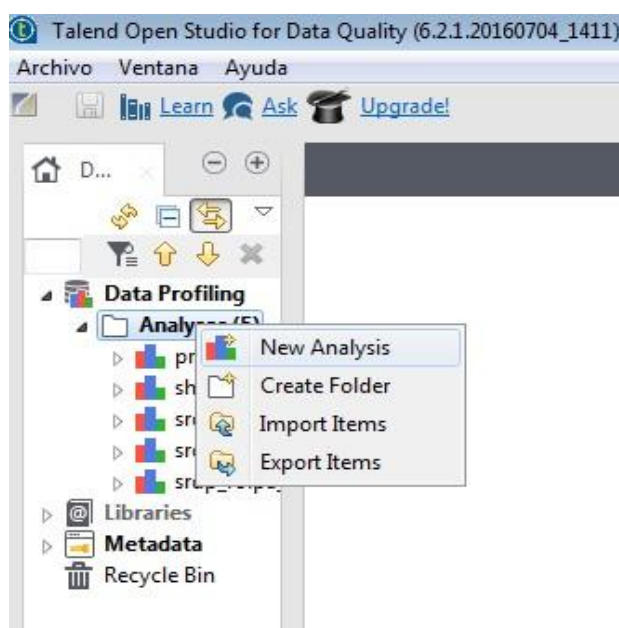


Figura 17: Selección de nuevo análisis

Es necesario especificar el tipo de análisis que se quiere realizar (ver figura 18).

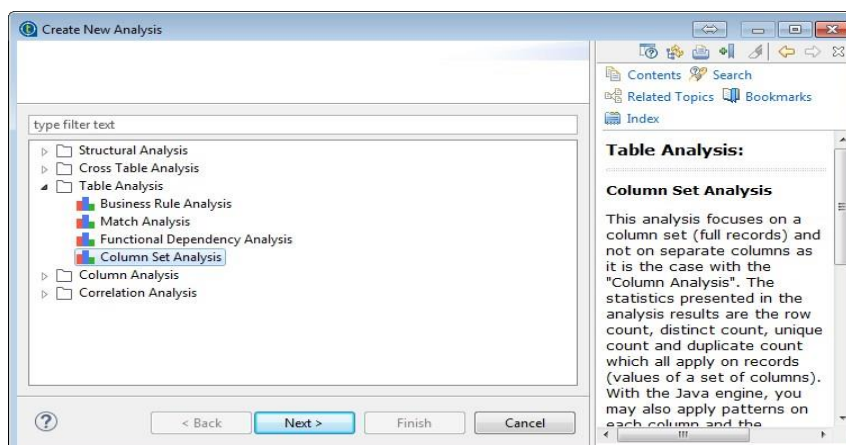


Figura 18: Tipo de análisis

A continuación se procede a registrar el nombre que va a tener nuestro análisis así como también una breve descripción de lo que va a realizar esta herramienta (ver figura 19).

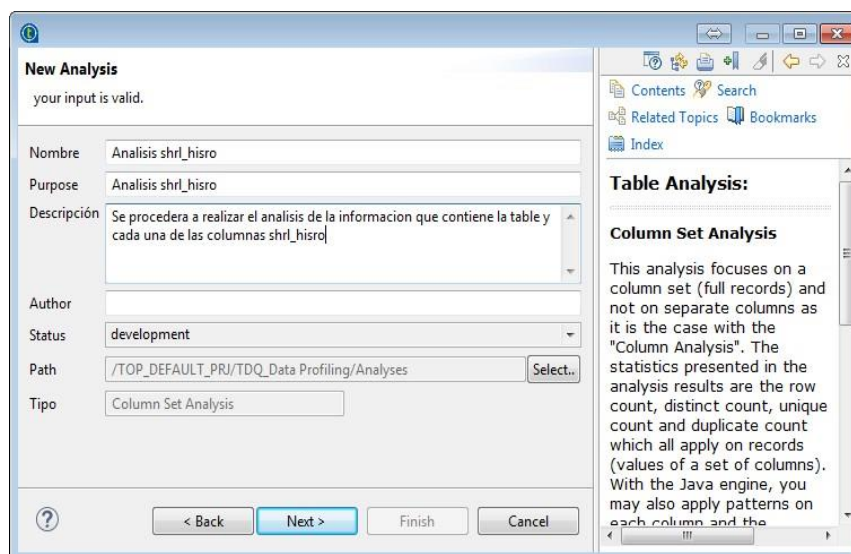


Figura 19: Nombre del análisis

En la siguiente etapa es necesario especificar la base de datos y la tabla a ser analizada (ver figura 20).

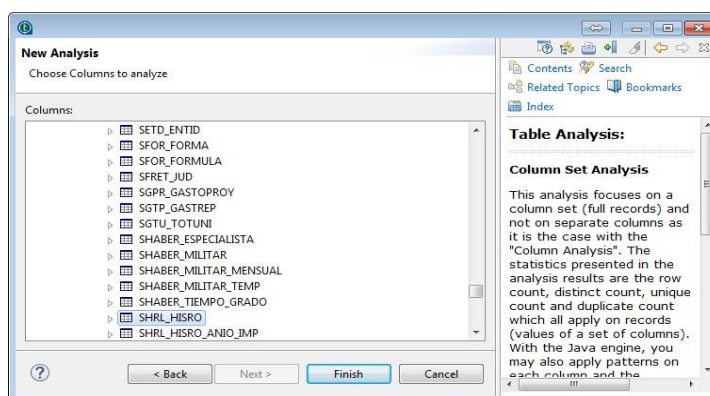
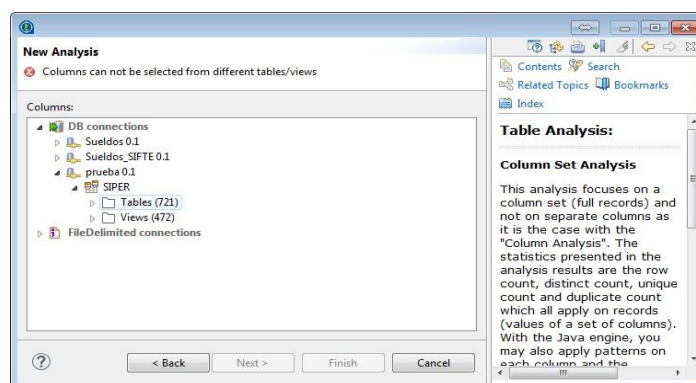


Figura 20: Tabla de análisis

A continuación la herramienta muestra las columnas a ser analizadas, cabe destacar que se debe parametrizar algunos datos para que se pueda realizar el análisis (ver figura 21).

TIE_CODIGO	IES_CODIGO	MEM_CEDULA	HRL_FECHA	HRL_CODIGO	TIP_CODIGO	HRL_VALRO	CODIGO	MES
1	2	3880	0501563753	2002-01-30 00:00:...	3341	1	6	1
2	3	51015	0501563753	2002-01-30 00:00:...	3341	1	2	1
3	2	3141	0501928527	2002-01-30 00:00:...	3804	1	69.6	1
4	2	3880	0501928527	2002-01-30 00:00:...	3804	1	6	1
5	1	3111	0501928790	2002-01-30 00:00:...	3805	1	107.9	1
6	2	3112	0501928790	2002-01-30 00:00:...	3805	1	43.1	1
7	2	3131	0501928790	2002-01-30 00:00:...	3805	1	64.7	1
8	1	3111	0501923064	2002-01-30 00:00:...	3795	1	135.8	1
9	2	3112	0501923064	2002-01-30 00:00:...	3795	1	54.3	1

Figura 21: Columnas de análisis

decir tenga una sola clave primaria. Además se puede apreciar que no existen registros en blanco.

A continuación se puede visualizar la misma información (ver figura 23).

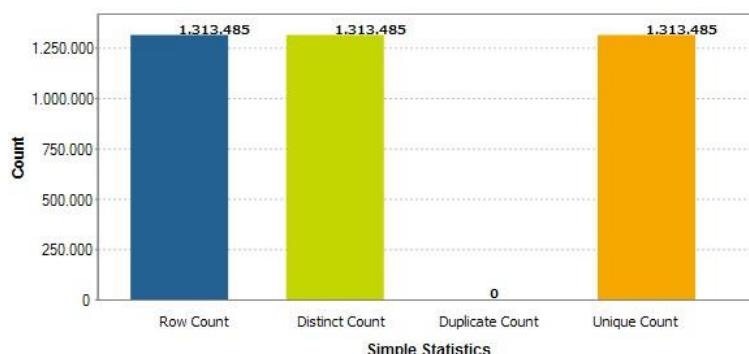


Figura 23: Registros analizados

3.6 Selección y descripción de los atributos para aplicar minería de datos

Se eligieron los siguientes campos para realizar la investigación utilizando técnicas de minería de datos; tomando en cuenta que cada uno de ellos puede formar una variable independiente con relación al salario anual que viene a ser la variable dependiente, estos campos seleccionados permitirán llegar a los objetivos propuestos, MEM_CEDULA, SEX_CODIGO, CODGRADOHIST, NOMGRADOHIST, TIEMPOGRADOHIST, TIEMPOSERVICIO, EDAD, GRADO, ECI_CODIGO, ESTADO, FECHAESTADO, PROVINCIA, ANIO, SALARIO.

Descripción de los atributos seleccionados

- MEM_CEDULA.- Este campo registra el número de cédula de la persona que se está realizando el cálculo de remuneraciones del año especificado.
- SEX_CODIGO.- Este campo muestra el sexo de la persona en ese registro.
- CODGRADOHIST.- Este campo muestra el código del grado de la persona que se encuentra en el registro especificado.
- NOMGRADOHIST.- Este campo muestra el nombre del grado que tiene la persona que se encuentra en el registro especificado.
- TIEMPOGRADOHIST.- Este campo muestra el tiempo del grado que tiene en el año especificado de la persona en ese registro.
- TIEMPOSERVICIO.- Este campo muestra el tiempo de servicio en el año especificado de la persona en ese registro.

- EDAD.- Este campo muestra la edad de la persona que se encuentra en el registro especificado.
- GRADO.- Este campo muestra el grado de la persona que se encuentra en el registro especificado.
- ECI_CODIGO.- Este campo muestra el estado civil de la persona que se encuentra en el registro especificado.
- ESTADO.- Este campo muestra el estado de la persona que se encuentra en el registro especificado, se debe tomar en cuenta que en este campo solo están las personas activas y con disponibilidad debido a que se pretende proyectar los valores a ser pagados en un futuro.
- FECHAESTADO.- Este campo permite visualizar la fecha en la que se registra el estado de la persona que se encuentra en el registro especificado.
- PROVINCIA.- Este campo muestra la provincia de nacimiento de la persona que se encuentra en el registro especificado.
- ANIO.- Este campo muestra el año en que la persona percibe el salario.
- SALARIO.- Este campo muestra la suma de las RMU del año especificado perteneciente a la persona en ese registro.

3.7 Consultas de tablas de la base de datos remuneraciones a analizar

Consulta historial del personal militar

```
SELECT pm.mem_cedula,
       pm.sex_codigo,
       f_grado_hist_r (pm.mem_cedula, '31/08/' || TO_CHAR (sh.anio), 3)
       CodgradoHist,
       f_grado_hist_r (pm.mem_cedula, '31/08/' || TO_CHAR (sh.anio), 4)
       NomgradoHist,
       F_TIEMPO_GRADO_FECHA_R (
       pm.mem_cedula,
       TO_DATE ('31/08/' || TO_CHAR (sh.anio), 'dd/mm/yyyy')
       )
       tiempogradoHist,
       (f_edad1 (pm.mem_f_alta, TO_DATE ('31/12/' || TO_CHAR
       (sh.anio), 'dd/mm/yyyy'))))
```

```

    tiemposervicio,
    (f_edad1 (pm.mem_fecnac, TO_DATE ('31/12/'|| TO_CHAR (sh.anio),
    'dd/mm/yyyy'))))
    edad,
    (SELECT gm.gmi_nombre
    FROM pgmi_g_mil gm
    WHERE gm.gmi_codigo = pm.gmi_codigo)
    grado,
    pm.eci_codigo,
    (CASE f_estado (pm.mem_cedula)
    WHEN 1 THEN 'ACTIVO'
    WHEN 2 THEN 'BAJA'
    WHEN 3 THEN 'DISPOSICION'
    WHEN 4 THEN 'DISPONIBILIDAD'
    END)
    estado,
    f_fecha_estado (pm.mem_cedula,
    f_estado (pm.mem_cedula))
    FechaEstado,
    (SELECT mp.prv_descri
    FROM mprv_provi mp
    WHERE mp.prv_codigo = pm.prv_codigo)
    provincia,
    sh.anio,
    SUM (sh.hrl_valro)
FROM shrl_hisro sh, pmem_milit pm
WHERE pm.mem_cedula = sh.mem_cedula
AND sh.anio IN
    (2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015)
AND sh.ies_codigo IN (3111, 9456)
AND pm.tef_codigo IN (1, 4)
AND f_estado (pm.mem_cedula) IN (1, 3, 4)

```

```

AND SH.HRL_FECHA >= pm.mem_f_alta and sh.anio >
to_number(to_char(pm.mem_f_alta,'yyyy'))
GROUP BY pm.mem_cedula, sh.anio, pm.sex_codigo,
pm.mem_f_alta,pm.mem_fecnac,pm.eci_codigo,pm.gmi_codigo,pm.prv_codigo
ORDER BY pm.mem_cedula, sh.anio;

```

Esta consulta permite obtener la información de los militares para poder alimentar cada uno de los modelos utilizados para la predicción de sueldos, en este caso se obtiene solo la información del personal militar que se encuentra activo, a disposición o disponibilidad debido a que solo a esos estados se realiza el pago mensual de las remuneraciones, al personal que ya salió con la baja no se toma en cuenta en este estudio porque los pagos los realiza el ISSFA, la información tomada del personal militar es desde el año 2007 al año 2015 debido a que se puede hacer un mejor análisis y predicción con esa cantidad de datos.

Consulta proyección salario personal militar

```

SELECT pm.mem_cedula,
       pm.sex_codigo,
       f_grado_hist_r (pm.mem_cedula, '31/08/' || TO_CHAR (sh.anio + 1), 3)
CodgradoHist,
       f_grado_hist_r (pm.mem_cedula, '31/08/' || TO_CHAR (sh.anio + 1),
4)NomgradoHist,
       F_TIEMPO_GRADO_FECHA_R (pm.mem_cedula,TO_
DATE ('31/08/' || TO_CHAR (sh.anio + 1), 'dd/mm/yyyy'))
tiempogradoHist,
       (f_edad1 (pm.mem_f_alta,
                TO_DATE ('31/12/' || TO_CHAR (sh.anio + 1),
'dd/mm/yyyy'))))
       tiemposervicio,
       (f_edad1 (pm.mem_fecnac,
                TO_DATE ('31/12/' || TO_CHAR (sh.anio + 1),
'dd/mm/yyyy'))))

```

```

    edad,
    (SELECT gm.gmi_nombre
     FROM pgmi_g_mil gm
     WHERE gm.gmi_codigo = pm.gmi_codigo)
    grado,
    pm.eci_codigo,
    (CASE f_estado (pm.mem_cedula)
     WHEN 1 THEN 'ACTIVO'
     WHEN 2 THEN 'BAJA'
     WHEN 3 THEN 'DISPOSICION'
     WHEN 4 THEN 'DISPONIBILIDAD'
     END)
    estado,
    f_fecha_estado (pm.mem_cedula, f_estado (pm.mem_cedula))
FechaEstado,
    (SELECT mp.prv_descri
     FROM mprv_provi mp
     WHERE mp.prv_codigo = pm.prv_codigo)
    provincia,
    (case sh.anio when 2015 then 2016 else sh.anio
end)anio
    FROM shrl_hisro sh, pmem_milit pm
    WHERE pm.mem_cedula = sh.mem_cedula
    AND sh.anio IN
                                -2015
    AND sh.ies_codigo IN (3111, 9456)
    AND pm.tef_codigo IN (1,4)
    AND f_estado (pm.mem_cedula) IN (1, 3, 4)
    AND SH.HRL_FECHA >= pm.mem_f_alta
GROUP BY pm.mem_cedula,
    sh.anio,
    pm.sex_codigo,

```

```
pm.mem_f_alta,  
pm.mem_fecnac,  
pm.eci_codigo,  
pm.gmi_codigo,  
pm.prv_codigo  
ORDER BY pm.mem_cedula, sh.anio;
```

Esta consulta permite obtener la información que se va a utilizar para la predicción de los sueldos; se debe tomar en cuenta ya que para la predicción se necesita varios registros por cada persona y no es necesario tener en varias columnas el valor de los sueldos anteriores, todos los campos seleccionados en esta consulta deben ser registrados en el modelo para posteriormente poder esta información utilizarla para realizar predicciones de años futuros.

En esta consulta solo se obtiene personal activo, a disposición y disponibilidad puesto que ellos cobran sus sueldos normales y se encuentran dentro del presupuesto del Ejército Ecuatoriano.

3.8 Pre-procesamiento de los datos en la base remuneraciones

En esta etapa se realizó el filtrado de los registros, donde los salarios sean mayores que cero, debido a que los mismos no aportan con nada en la investigación que se está realizando. Además, se ha creado archivos de extensión xlsx que permitirá realizar la minería de datos con mayor velocidad de procesamiento; a fin de encontrar patrones de comportamiento del personal militar. El proceso que se indica a continuación ha permitido realizar la consulta a la base de datos de acuerdo a los atributos a utilizar; filtrar los registros con datos reales y exportar a Excel con el propósito de mejorar la velocidad de procesamiento de los registros involucrados en la búsqueda de patrones que permitan determinar la proyección de salarios (ver figura 24).

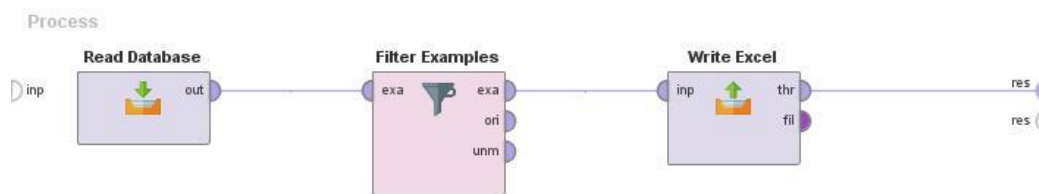


Figura 24: Proceso leer base de datos y llevar a Excel

3.8.1 Load (carga)

Aplicando la información obtenida con las sentencias sql (select), se procede a obtener la información de las remuneraciones del personal militar de los años 2007 al 2015, generando un archivo Excel que ha sido utilizado para realizar las proyecciones de los salarios correspondientes a los años 2016 al 2018 utilizando la herramienta RapidMiner.

3.9 Correlación por peso de los atributos seleccionados

3.9.1 Correlación para proyección del salario del personal militar

Se utilizó el siguiente modelo para poder determinar la relevancia de cada uno de los atributos seleccionados en nuestra investigación con relación al atributo dependiente salario; lo cual permitió obtener los siguientes resultados (ver figura 25).

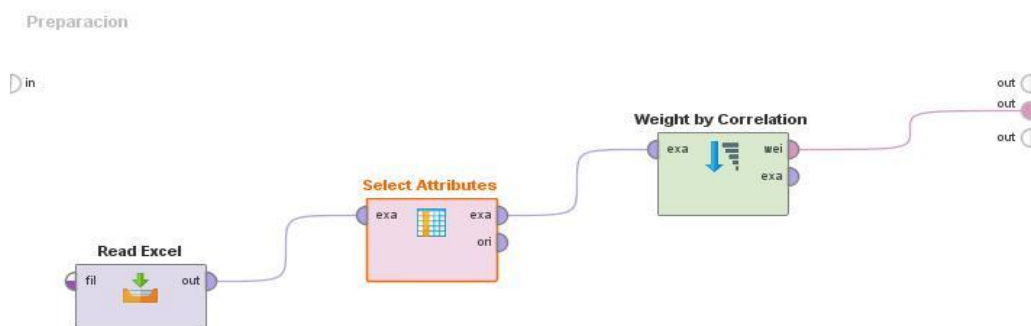


Figura 25: Modelo generación peso de atributo

En el modelo presentado se seleccionaron de la base de datos a investigar; los atributos anio, codgradohist, memcedula, salario y tiempogradohist para realizar la correlación y poder determinar la importancia de cada uno de ellos en la búsqueda de la predicción del salario del personal militar (ver figura 26).

atribute	weight
tiempogradohist	0.173
año	0.419
codgradohist	0.812

Figura 26: Pesos de atributos

Seguidamente, se puede apreciar en un diagrama de barras (ver figura 27).

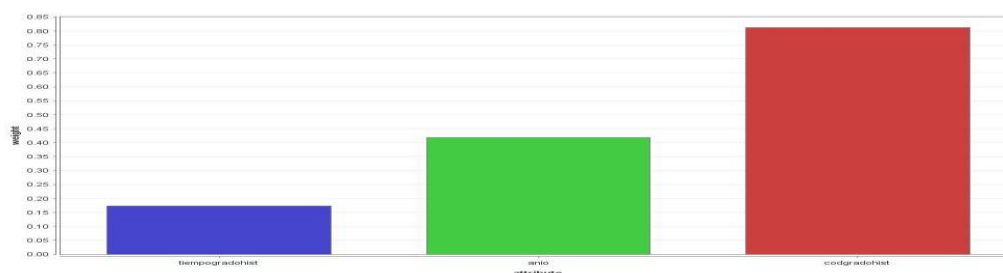


Figura 27: Barras con pesos de atributos

Se puede observar que el atributo que más relevancia tiene para nuestra investigación es el código del grado histórico que tiene el personal militar con un 81.2% de correlación, seguido del año con un 41.9%, y por último el tiempo en el grado histórico con el 17.3%, con respecto al salario que se busca. Por lo que se ha de tomar una especial atención a cada uno de estos atributos; debido a que conoce el peso que representa para nuestras proyecciones, los atributos que no son relevantes al modelo se ha eliminado con excepción de la cédula que es considerado como atributo clave.

3.9.2 Correlación para proyección de salarios a través de gráficos

Se utilizó el siguiente modelo para poder determinar la relevancia de cada uno de los atributos seleccionados para diseñar gráficos de predicción con relación al atributo dependiente salario (ver figura 28).

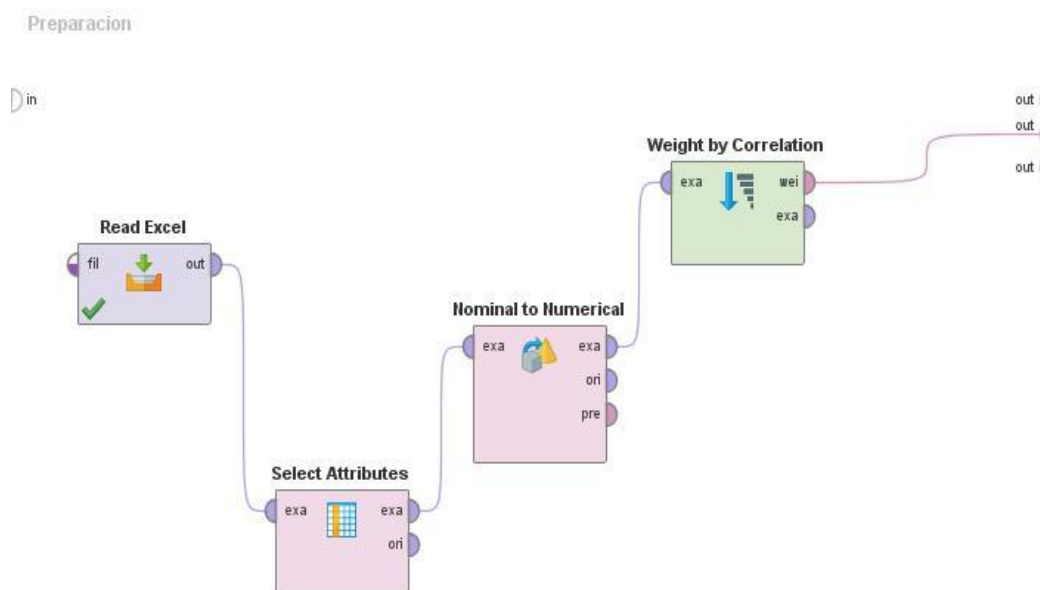


Figura 28: Modelo pesos atributos para gráficos

En el modelo presentado se seleccionaron de la base de datos a investigar los atributos eci_codigo, edad, fechaestado, grado, provincia y sex_codigo para realizar la correlación y poder determinar la importancia de cada uno de ellos en la búsqueda de la predicción del salario a través de gráficos para el personal militar; lo que permitirá realizar diversos tipos de análisis según el atributo.

A continuación se puede apreciar los pesos (ver figura 29).

attribute	weight
provincia	0.029
eci_codigo	0.076
sex_codigo	0.082
grado	0.345
fechaestado	0.391
edad	0.608

Figura 29: Pesos atributos para gráficos

Seguidamente se puede observar los pesos de los atributos en un diagrama de barras (ver figura 30).

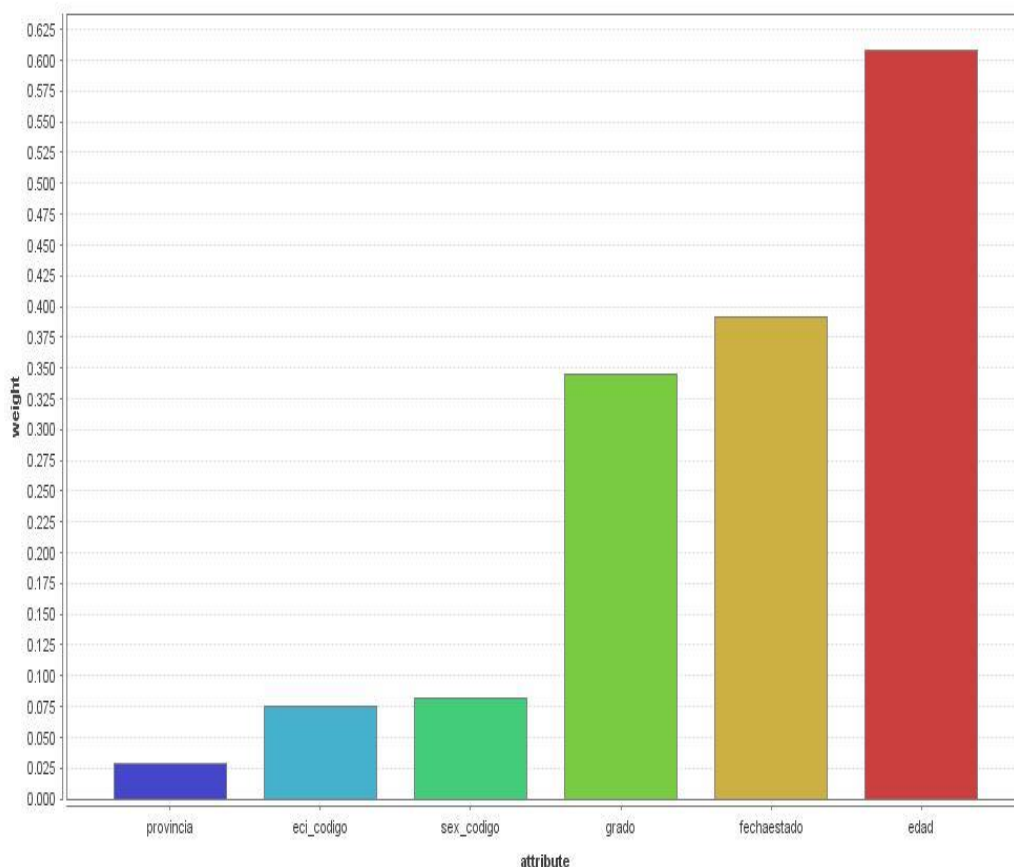


Figura 30: Barras con pesos atributos para gráficos

Se puede observar el peso de los atributos para la generación de gráficos con relación al salario; estos son los siguientes: la edad 60,8%, fechaestado 39,1%, grado 34,5%, sex_codigo 8,2%, eci_codigo 7,6% y provincia 2,9%. Por lo que se ha de tomar especial atención a los porcentajes más relevantes para nuestra investigación; se puede mirar que la variable de mayor importancia es la edad y la menos importante es la provincia.

3.10 Aplicación de algoritmos de minería de datos y análisis de resultados

En la actualidad existe un gran volumen de información almacenada en bases de datos; aprovechando la minería de datos se procede a extraer conocimiento utilizando

técnicas de minería de datos; se diseñará modelos que permita realizar proyecciones de remuneraciones del personal militar del Ejército Ecuatoriano.

3.10.1 Modelo proyección salario coroneles con redes neuronales

A continuación se presenta en la (figura 31) los procesos que permiten generar la proyección de salarios de los coroneles aplicando redes neuronales.

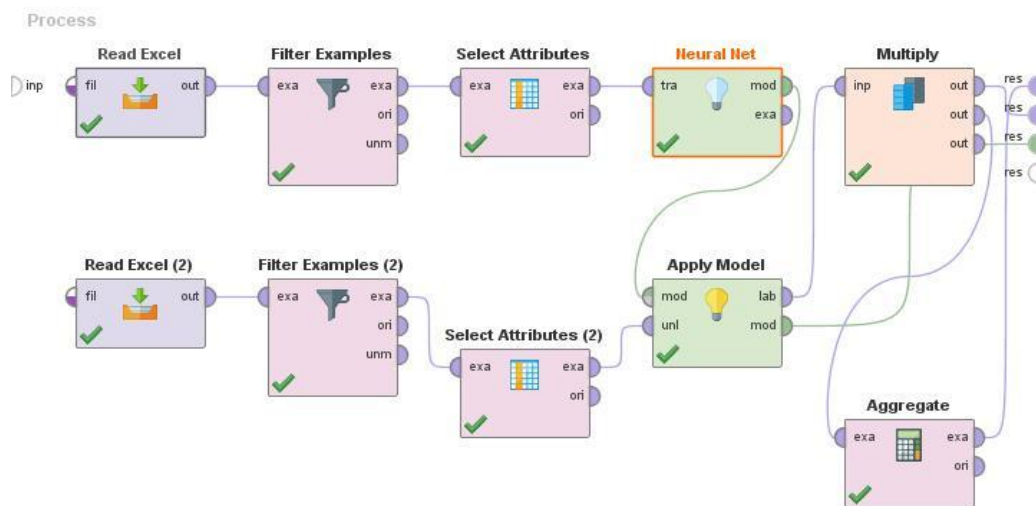


Figura 31: Modelo coronel con red neuronal

Para realizar este modelo se utilizó la base de datos con los datos del personal militar tomando en cuenta información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015; filtrando por el código de grado histórico igual a 4 que corresponde al grado de coronel; seguidamente se seleccionó los siguientes atributos año, codgradohist, men_cedula, salario, tiempo gradohist considerados los atributos relevantes para la predicción del salario. Además se seleccionó la base de datos de predicción del personal militar para el año 2016; filtrando por el código del grado histórico igual a 4; luego se eligió todos los atributos de la base de datos predicción militares activos 2016 y se aplicó el modelo Neural Net para la predicción del salario de cada personal militar y la suma del total de la predicción del salario necesario para el año 2016 a través del operador Aggregate.

A continuación se presenta en la (figura 32) la predicción del salario para el año 2016 de los coroneles en forma individual que genera la red neuronal con un total de 208 registros.

Row No.	mem_cedula	prediction(salario)	sex_codigo	codgradohist	nomgradohist	tiempograd...	tiemposervi...	edad	grado	eci_codigo
1	101061679	53822.302	M	4	CRNL	6	29	62	CORONEL	S
2	101719987	49500.997	M	4	CRNL	2	29	50	CORONEL	C
3	101789089	53822.302	M	4	CRNL	6	32	54	CORONEL	C
4	101943314	51210.340	M	4	CRNL	4	31	52	CORONEL	C
5	102286226	50448.715	M	4	CRNL	3	30	50	CORONEL	C
6	200938223	49500.997	M	4	CRNL	2	29	51	CORONEL	D
7	200955573	49500.997	M	4	CRNL	2	29	50	CORONEL	C
8	200987600	44413.233	M	4	CRNL	0	28	48	CORONEL	C
9	300743028	44413.233	M	4	CRNL	0	28	55	CORONEL	C
10	300832706	49500.997	M	4	CRNL	2	29	50	CORONEL	C
11	300879665	51210.340	M	4	CRNL	4	31	52	CORONEL	C
12	400637757	49500.997	M	4	CRNL	2	29	54	CORONEL	C
13	400646642	53822.302	M	4	CRNL	6	32	54	CORONEL	D
14	400675146	52283.732	M	4	CRNL	5	31	53	CORONEL	C
15	400684220	52283.732	M	4	CRNL	5	31	53	CORONEL	C
16	400686804	49500.997	M	4	CRNL	2	29	50	CORONEL	C

Row No.	sum(predict...
1	10329463.450

Figura 32: Predicción salario total coronel con red neuronal

Se puede apreciar la predicción del salario individual del año 2016 de todos los coroneles; se ha de tomar en cuenta que se tomó información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 por número de cédula. El total del salario necesario para el personal militar con el grado de coronel en el año 2016 será de 10'329.463,450 dólares. Se ha de tomar en cuenta que en esta predicción intervienen únicamente los coroneles; que la precisión es realmente casi perfecta con relación a la información real de cada uno de los salarios. Se toma únicamente a cada uno de los grados de forma individual porque la variación de salarios entre algunos grados

militares es bastante alta y esta variación da un resultado erróneo si se toma a todos los grados militares por igual.

A continuación se presenta en la (figura 33) la predicción del salario en un gráfico de dispersión para el año 2016 de los coroneles en forma individual que genera la red neuronal.

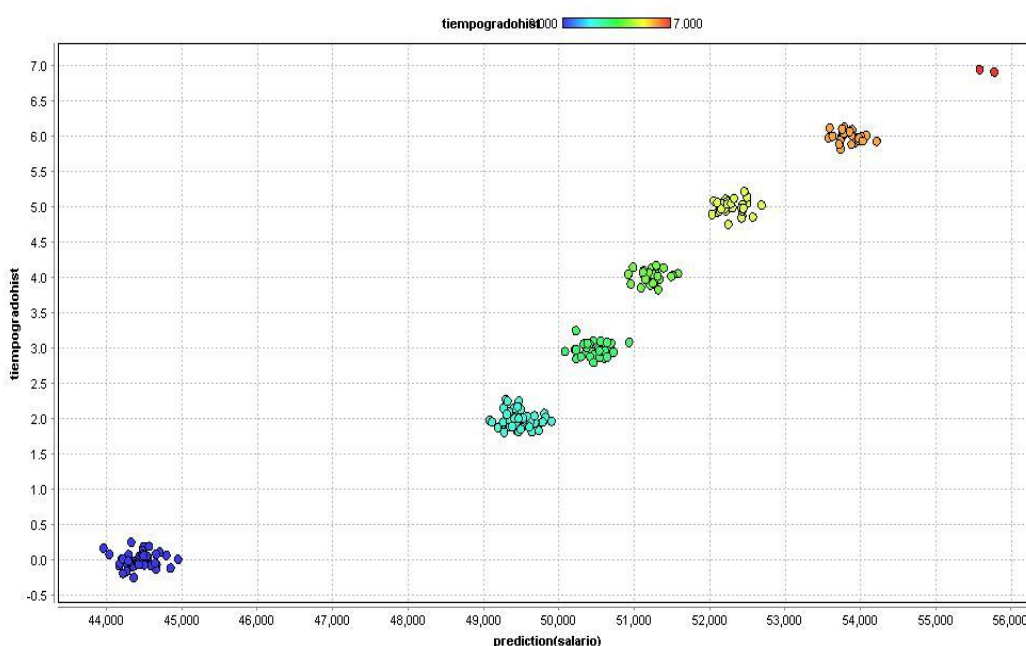


Figura 33: Predicción salario coronel individual con red neuronal

Del gráfico se puede predecir que personal militar con grado de coronel que tiene entre 1,5 y 2,5 años de tiempo en el grado ganarán entre 49.000 y 50.000 dólares en el año 2016; un número reducido de coroneles con 6,5 a 7 años de tiempo en el grado tendrán un salario anual en el 2016 entre 55.000 y 56.000 dólares y entre 44000 y 45000 dólares con tiempo en el grado menor a medio año existe un número considerable de personal militar.

3.10.2 Modelo proyección salario teniente coronel con redes neuronales

A continuación se presenta en la (figura 34) los procesos que permiten generar la proyección de salarios de los tenientes coroneles aplicando redes neuronales.

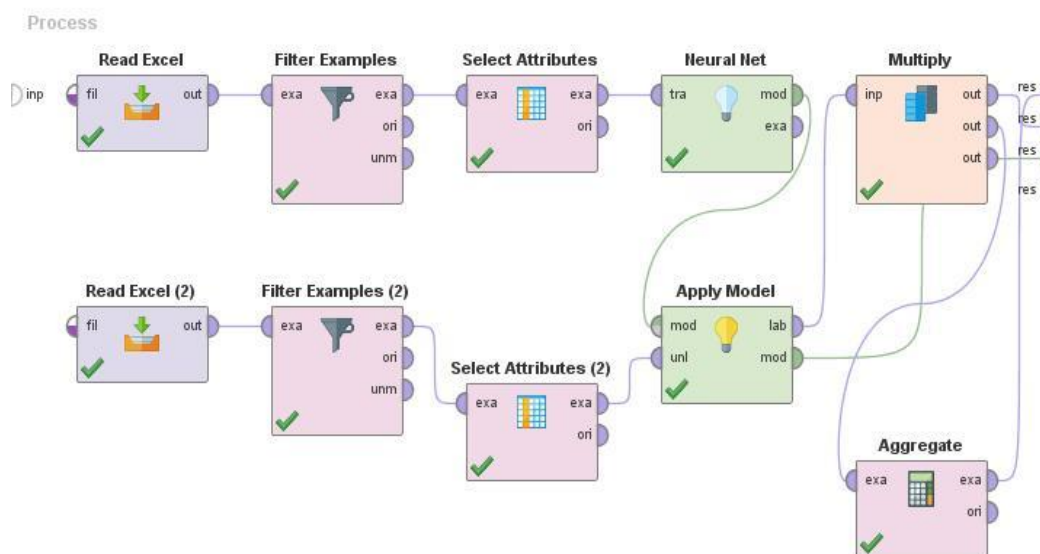


Figura 34: Modelo teniente coronel con red neuronal

Para realizar este modelo se utilizó la base de datos con los datos del personal militar tomando en cuenta información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015; filtrando por el código de grado histórico igual a 5 que corresponde al grado de teniente coronel; seguidamente se seleccionó los siguientes atributos año, codgradohist, men_cedula, salario, tiempoegradohist considerados los atributos relevantes para la predicción del salario.

Además se seleccionó la base de datos de predicción del personal militar para el año 2016; filtrando por el código del grado histórico igual a 5; luego se eligió todos los atributos de la base de datos predicción militares activos 2016 y se aplicó el modelo Neural Net para la predicción del salario de cada personal militar y la suma del total de la predicción del salario necesario para el año 2016 a través del operador Aggregate.

A continuación (ver figura 35) se presenta la predicción del salario para el año 2016 de los tenientes coroneles en forma individual que genera la red neuronal con un total de 328 registros.

Row No.	mem_cedula	prediction(s...	sex_codigo	codgradohist	nomgradohist	tiempograd...	tiemposervi...	edad	grado	eci_codigo	estado
1	101505493	41592.501	M	5	TCRN	5	26	59	TENIENTE C...	C	ACTIVO
2	101661023	39182.567	M	5	TCRN	2	24	47	TENIENTE C...	C	ACTIVO
3	102114683	39182.567	M	5	TCRN	2	24	46	TENIENTE C...	C	ACTIVO
4	102133576	38305.842	M	5	TCRN	0	23	45	TENIENTE C...	S	ACTIVO
5	102157328	39182.567	M	5	TCRN	2	24	45	TENIENTE C...	C	ACTIVO
6	102245479	38305.842	M	5	TCRN	0	23	45	TENIENTE C...	C	ACTIVO
7	102253218	38305.842	M	5	TCRN	0	21	51	TENIENTE C...	C	ACTIVO
8	102366986	39958.863	M	5	TCRN	3	26	49	TENIENTE C...	C	ACTIVO
9	102634615	39958.863	M	5	TCRN	3	26	50	TENIENTE C...	C	ACTIVO
10	102634854	39958.863	M	5	TCRN	3	26	48	TENIENTE C...	C	ACTIVO
11	102636016	39958.863	M	5	TCRN	3	26	48	TENIENTE C...	C	ACTIVO
12	102642691	41592.501	M	5	TCRN	5	27	48	TENIENTE C...	C	ACTIVO
13	102657780	42135.343	M	5	TCRN	6	28	49	TENIENTE C...	C	ACTIVO
14	102659190	38305.842	M	5	TCRN	0	23	46	TENIENTE C...	C	ACTIVO
15	102678372	40826.783	M	5	TCRN	4	27	47	TENIENTE C...	C	ACTIVO
16	102730124	39182.567	M	5	TCRN	2	24	49	TENIENTE C...	C	ACTIVO

Row No.	sum(prediction(salario))
1	13159848.749

Figura 35: Predicción salario total tcrln con red neuronal

Se puede apreciar la predicción del salario individual del año 2016 de todos los tenientes coroneles; se ha de tomar en cuenta que se tomó información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 por número de cédula. El total del salario necesario para el personal militar con el grado de teniente coronel en el año 2016 será de 13'159.848,749 dólares.

A continuación en la (figura 36) se muestra la predicción del salario en un gráfico de dispersión para el año 2016 de los tenientes coroneles en forma individual que genera la red neuronal.

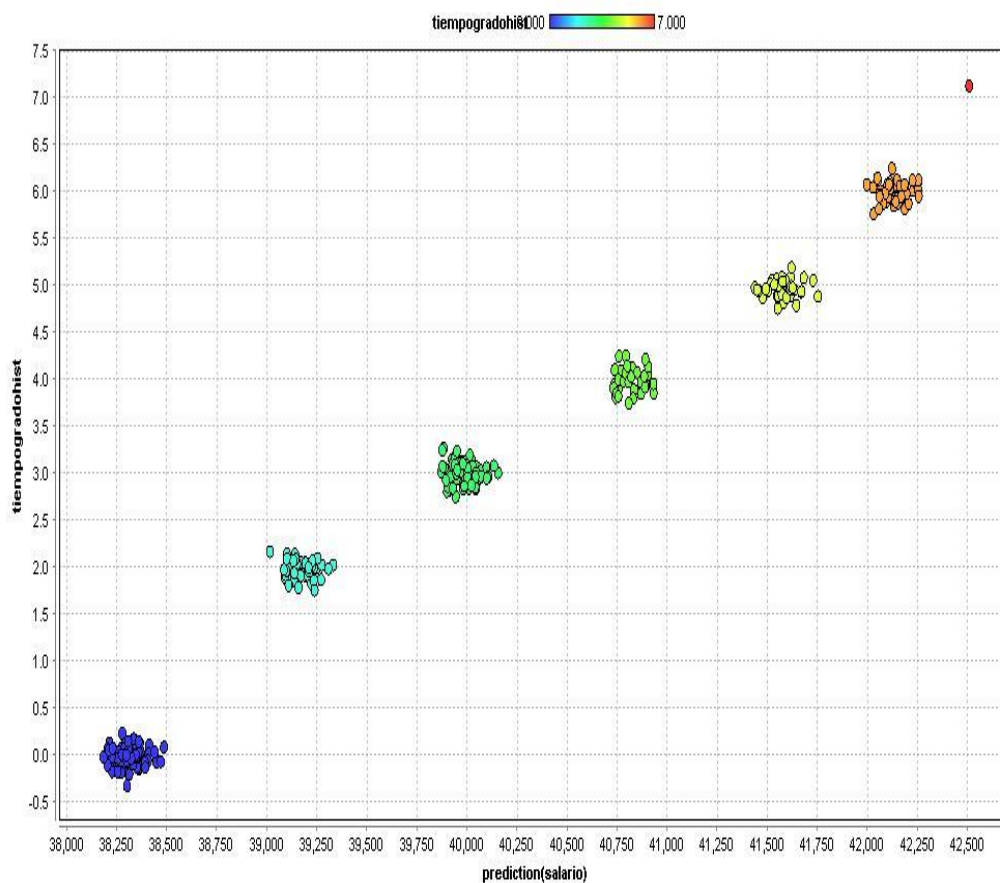


Figura 36: Predicción salario teniente coronel individual con red neuronal

Del gráfico se puede predecir que personal militar con grado de teniente coronel que tiene entre 2,5 y 3,5 años de tiempo en el grado ganarán entre 39.750 y 40.250 dólares en el año 2016; en cambio un número reducido de tenientes coroneles con 7 a 7,5 años de tiempo en el grado tendrán un salario anual en el 2016 entre 42.500 y más dólares.

Se puede observar que entre 38.100 y 38.500 dólares existen un número considerable de oficiales con menos de medio año de tiempo en el grado.

3.10.3 Modelo proyección sargento segundo con redes neuronales

A continuación se muestra en la (figura 37) los procesos que permiten generar la proyección de salarios de los sargentos segundos aplicando redes neuronales.

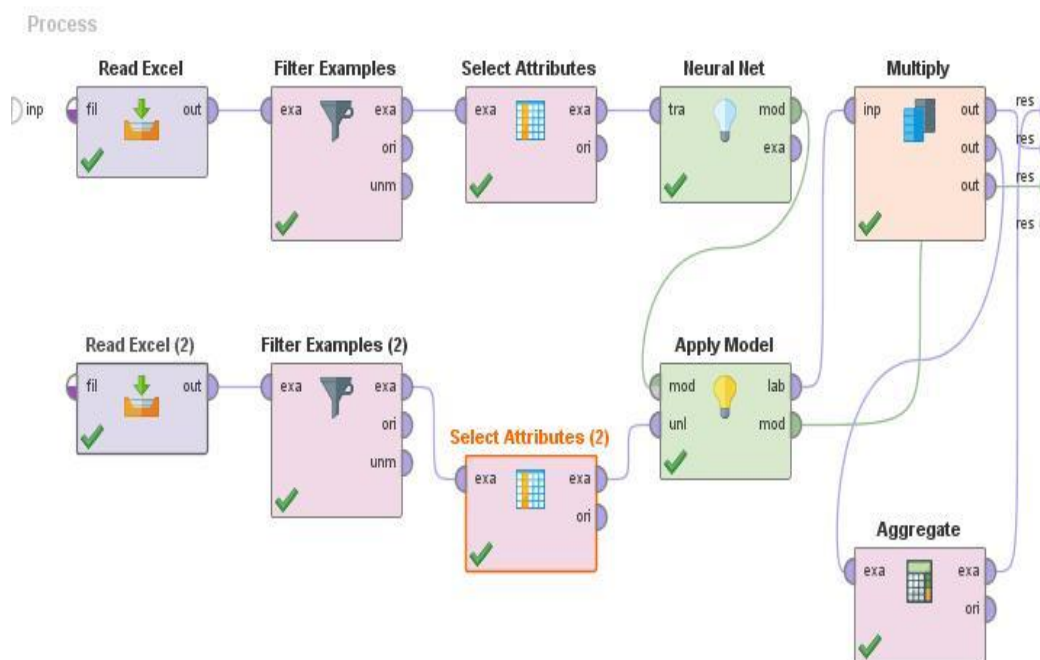


Figura 37: Modelo sargento segundo con red neuronal

Para realizar este modelo se utilizó la base de datos con los datos del personal militar tomando en cuenta información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015; filtrando por el código de grado histórico igual a 15 que corresponde al grado de sargento segundo; seguidamente se seleccionó los siguientes atributos año, codgradohist, men_cedula, salario, tiempoegradohist considerados los atributos relevantes para la predicción del salario.

Además se seleccionó la base de datos de predicción del personal militar para el año 2016; filtrando por el código del grado histórico igual a 15; luego se eligió todos los atributos de la base de datos predicción militares activos 2016 y se aplicó el modelo Neural Net para la predicción del salario de cada personal militar y la suma del total de la predicción del salario necesario para el año 2016 a través del operador Aggregate.

A continuación en la (figura 38) se presenta la predicción del salario para el año 2016 de los sargentos segundos en forma individual que genera la red neuronal con un total de 5788 registros.

Row No.	mem_cedula	prediction(salario)	sex_codigo	codgradohist	nomgradohist	tiempograd...	tiemposervi...	edad	grado	eci_codigo
1	102144987	16277.118	M	15	SGOS	5	20	41	SARGENTO...	C
2	102677143	16277.118	M	15	SGOS	5	20	39	SARGENTO...	C
3	103334967	15920.911	M	15	SGOS	4	19	40	SARGENTO...	C
4	103416376	15611.214	M	15	SGOS	3	18	39	SARGENTO...	S
5	103586137	16277.118	M	15	SGOS	5	19	38	SARGENTO...	D
6	103819033	15343.101	M	15	SGOS	2	16	40	SARGENTO...	C
7	103916680	15920.911	M	15	SGOS	4	19	39	SARGENTO...	C
8	103925517	15611.214	M	15	SGOS	3	17	36	SARGENTO...	C
9	104090725	14913.574	M	15	SGOS	0	15	38	SARGENTO...	C
10	201191293	16685.035	M	15	SGOS	6	21	46	SARGENTO...	C
11	201195112	16685.035	M	15	SGOS	6	21	47	SARGENTO...	U
12	201196177	15920.911	M	15	SGOS	4	19	45	SARGENTO...	C
13	201244266	16685.035	M	15	SGOS	6	21	45	SARGENTO...	U
14	201245990	16685.035	M	15	SGOS	6	21	43	SARGENTO...	C
15	201261583	16685.035	M	15	SGOS	6	21	44	SARGENTO...	C
16	201263603	16277.118	M	15	SGOS	5	20	44	SARGENTO...	C

Row No.	sum(prediction(salario))
1	90857663.365

Figura 38: Predicción salario total sargento segundo con red neuronal

Se puede apreciar la predicción del salario individual del año 2016 de todos los sargentos segundos; se ha de tener en cuenta que se tomó información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 por número de cédula. El total del salario necesario para el personal militar con el grado de sargento segundo en el año 2016 será de 90'857.663,365 dólares.

A continuación en la (figura 39) se muestra la predicción del salario en un gráfico de dispersión para el año 2016 de los sargentos segundos en forma individual que genera la red neuronal.

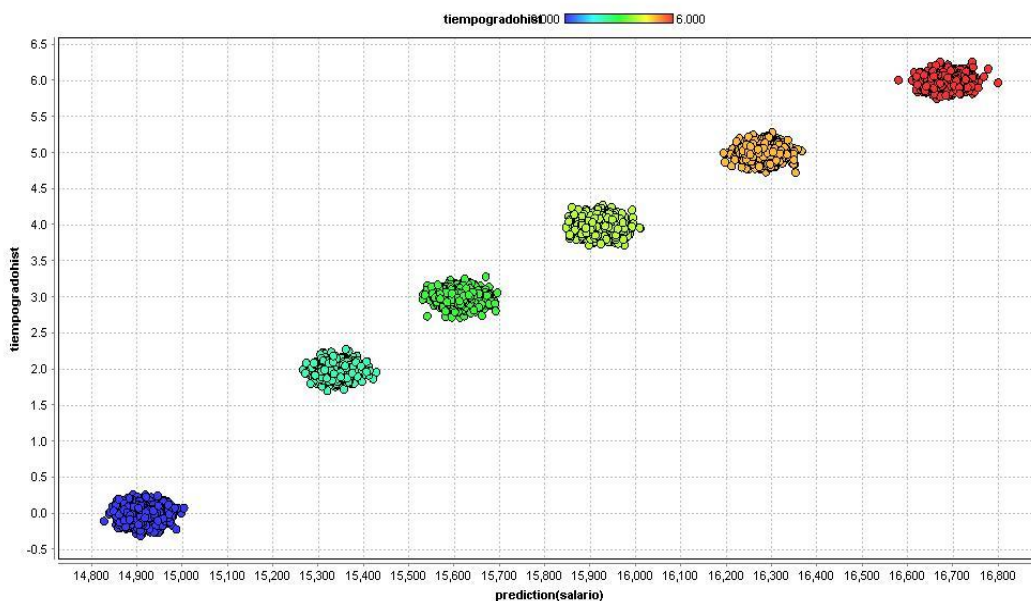


Figura 39: Predicción salario sargento segundo con red neuronal

Del gráfico se puede predecir que personal militar con grado de sargento segundo que tiene entre 1,5 y 2,5 años de tiempo en el grado ganarán entre 15.300 y 15.600 dólares, entre 2,6 y 3,4 años de tiempo en el grado ganarán entre 15.500 y 15.700 dólares, los grupos de tiempos en el grado y el salario de predicción según el gráfico se puede apreciar que son casi similares.

3.10.4 Modelo proyección salario coroneles con regresión lineal

A continuación se muestra en la (figura 40) los procesos que permiten generar la proyección de salarios de los coroneles aplicando regresión lineal.

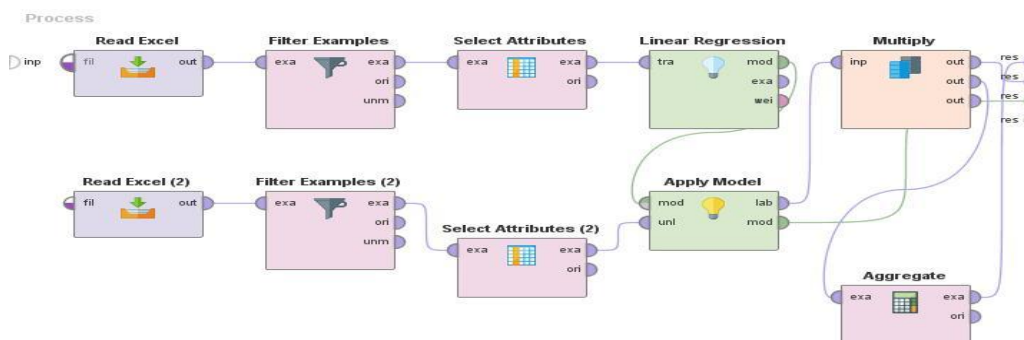


Figura 40: Modelo coroneles con regresión lineal

Para realizar este modelo se utilizó la base de datos con los registros del personal militar tomando en cuenta información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015; filtrando por el código de grado histórico igual a 4 que corresponde al grado de coronel; seguidamente se seleccionó los siguientes atributos año, codgradohist, men_cedula, salario, tiempogradohist considerados los atributos relevantes para la predicción del salario. Además se seleccionó la base de datos de predicción del personal militar para el año 2016; filtrando por el código del grado histórico igual a 4 y se aplicó el modelo de regresión lineal para la predicción del salario de cada personal militar y la suma del total de la predicción del salario necesario a través del operador Aggregate.

A continuación en la (figura 41) se muestra la predicción del salario en forma individual y total de los coroneles utilizando regresión lineal.

Row No.	mem_cedula	prediction(salario)	sex_codigo	codgradohist	nomgradohist	tiempograd...	tiemposervi...	edad	grado	eci_codigo	estado
1	101061679	56816.795	M	4	CRNL	6	29	62	CORONEL	S	ACTIVO
2	101719987	52068.853	M	4	CRNL	2	29	50	CORONEL	C	ACTIVO
3	101789089	56816.795	M	4	CRNL	6	32	54	CORONEL	C	ACTIVO
4	101943314	53942.824	M	4	CRNL	4	31	52	CORONEL	C	ACTIVO
5	102282226	53005.838	M	4	CRNL	3	30	50	CORONEL	C	ACTIVO
6	200938223	52068.853	M	4	CRNL	2	29	51	CORONEL	D	ACTIVO
7	200955573	52068.853	M	4	CRNL	2	29	50	CORONEL	C	ACTIVO
8	200987600	50194.882	M	4	CRNL	0	28	48	CORONEL	C	ACTIVO
9	300743028	50194.882	M	4	CRNL	0	28	55	CORONEL	C	ACTIVO
10	300832706	52068.853	M	4	CRNL	2	29	50	CORONEL	C	ACTIVO
11	300879665	53942.824	M	4	CRNL	4	31	52	CORONEL	C	ACTIVO
12	400637757	52068.853	M	4	CRNL	2	29	54	CORONEL	C	ACTIVO
13	400646642	56816.795	M	4	CRNL	6	32	54	CORONEL	D	ACTIVO
14	400675146	54879.809	M	4	CRNL	5	31	53	CORONEL	C	ACTIVO
15	400694220	54879.809	M	4	CRNL	5	31	53	CORONEL	C	ACTIVO
16	400696804	52068.853	M	4	CRNL	2	29	50	CORONEL	C	ACTIVO

Row No.	sum(prediction(salario))
1	11001789.685

Figura 41: Predicción salario total coronel con regresión lineal

Se puede apreciar la predicción del salario individual del año 2016 de todos los coroneles; se ha de tener en cuenta que se tomó información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 por número de cédula. El total del salario necesario para el personal militar con el grado de coronel en el año 2016 será de 11001789,685 dólares.

A continuación se presenta en la (figura 42) la predicción del salario en un gráfico de dispersión para el año 2016 de los coroneles en forma individual que genera la regresión lineal.

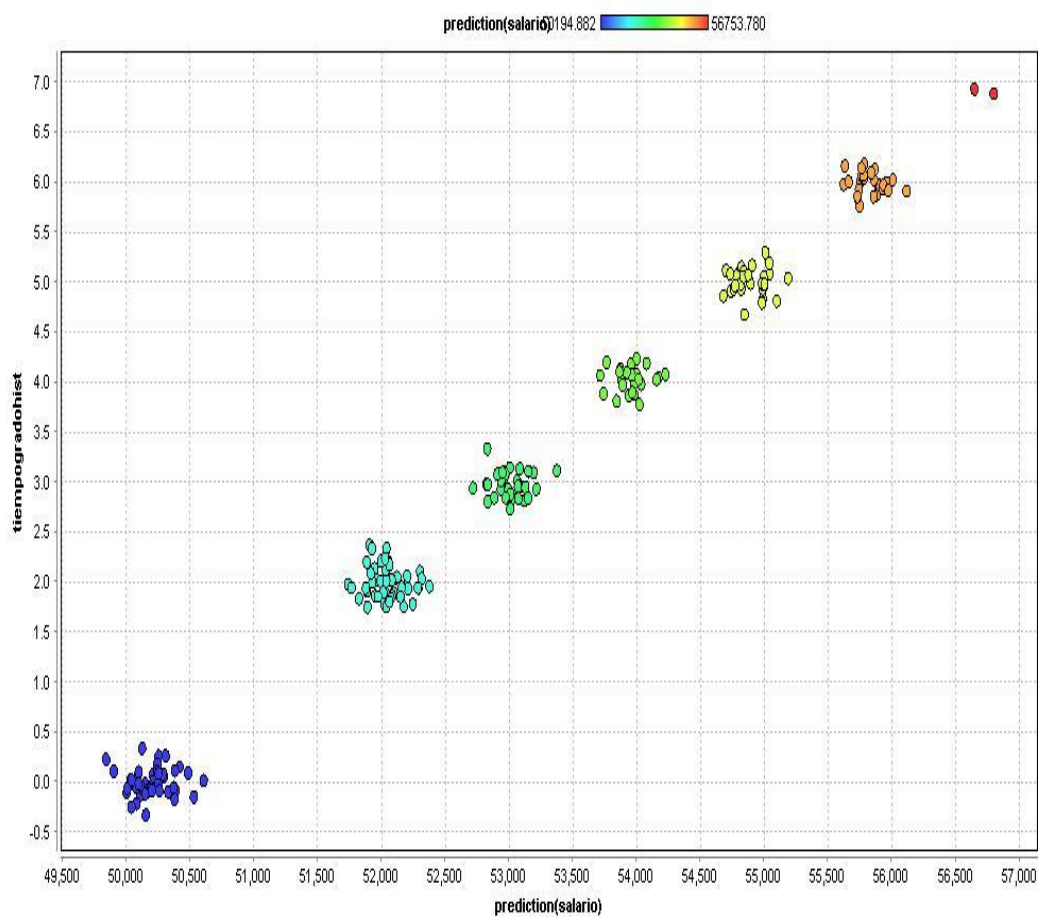


Figura 42: Predicción salario coronel con regresión lineal

Del gráfico se puede predecir que personal militar con grado de coronel que tiene entre 1,5 y 2,5 años de tiempo en el grado ganarán entre 51.500 y 52.500 dólares en el

año 2016 en cambio un número reducido de coroneles con 6,5 a 7 años de tiempo en el grado tendrán un salario anual en el 2016 entre 56.500 y 57.000 dólares.

3.10.5 Modelo proyección salario teniente coronel con regresión lineal

A continuación se muestra en la (figura 43) los procesos que permiten generar la proyección de salarios de los tenientes coroneles aplicando regresión lineal.

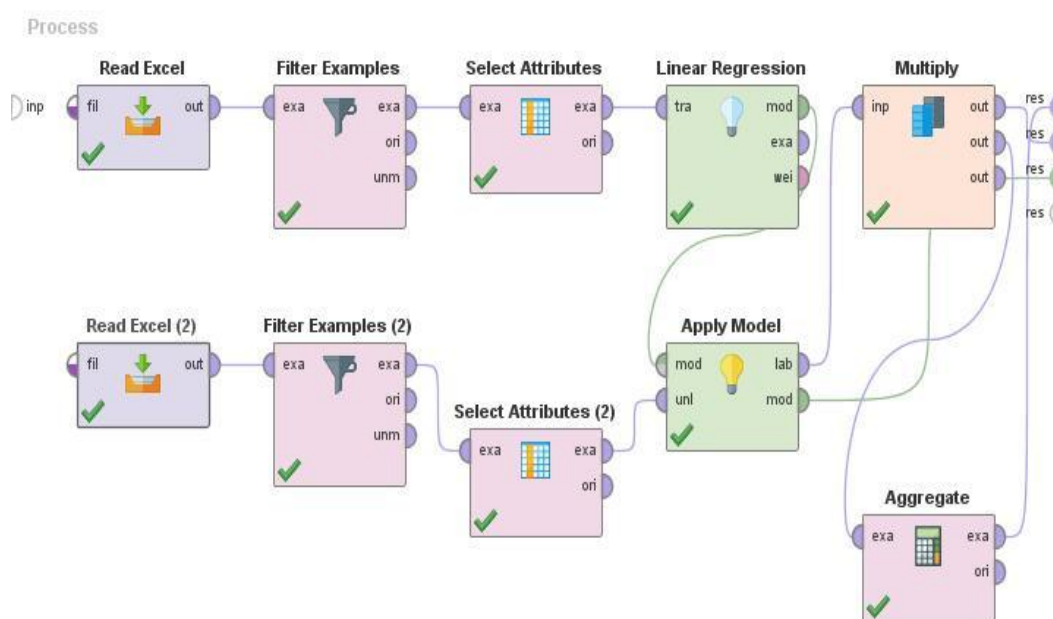


Figura 43: Modelo teniente coronel con regresión lineal

Para realizar este modelo se utilizó la base de datos con los registros del personal militar tomando en cuenta información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015; filtrando por el código de grado histórico igual a 5 que corresponde al grado de teniente coronel; seguidamente se seleccionó los siguientes atributos año, codgradohist, men_cedula, salario, tiempogradohist considerados los atributos relevantes para la predicción del salario. Además se seleccionó la base de datos de predicción del personal militar para el año 2016; filtrando por el código del grado histórico igual a 5 y se aplicó el modelo de regresión lineal para la predicción del salario de cada personal militar y la suma del total de la predicción del salario necesario a través del operador Aggregate.

A continuación en la (figura 44) se muestra la predicción del salario en forma individual y total de los tenientes coroneles utilizando regresión lineal.

Row No.	mem_cedula	prediction(salario)	sex_codigo	codgradohist	nomgradohist	tiempograd...	tiemposervi...	edad	grado	eci_codigo	estado
1	101505493	46371.226	M	5	TCRN	5	26	59	TENIENTE C...	C	ACTIVO
2	101661023	43439.304	M	5	TCRN	2	24	47	TENIENTE C...	C	ACTIVO
3	102114683	43439.304	M	5	TCRN	2	24	46	TENIENTE C...	C	ACTIVO
4	102133576	41484.689	M	5	TCRN	0	23	45	TENIENTE C...	S	ACTIVO
5	102157328	43439.304	M	5	TCRN	2	24	45	TENIENTE C...	C	ACTIVO
6	102245479	41484.689	M	5	TCRN	0	23	45	TENIENTE C...	C	ACTIVO
7	102253218	41484.689	M	5	TCRN	0	21	51	TENIENTE C...	C	ACTIVO
8	102366986	44416.611	M	5	TCRN	3	26	49	TENIENTE C...	C	ACTIVO
9	102634615	44416.611	M	5	TCRN	3	26	50	TENIENTE C...	C	ACTIVO
10	102634854	44416.611	M	5	TCRN	3	26	48	TENIENTE C...	C	ACTIVO
11	102636016	44416.611	M	5	TCRN	3	26	48	TENIENTE C...	C	ACTIVO
12	102642691	46371.226	M	5	TCRN	5	27	48	TENIENTE C...	C	ACTIVO
13	102657780	47348.533	M	5	TCRN	6	28	49	TENIENTE C...	C	ACTIVO
14	102659190	41484.689	M	5	TCRN	0	23	46	TENIENTE C...	C	ACTIVO
15	102678372	45393.919	M	5	TCRN	4	27	47	TENIENTE C...	C	ACTIVO
16	109730124	43439.304	M	5	TCRN	2	24	49	TENIENTE C...	C	ACTIVO

Row No.	sum(prediction(salario))
1	14579398.798

Figura 44: Predicción salario total teniente coronel con regresión lineal

Se puede apreciar la predicción del salario individual del año 2016 de todos los tenientes coroneles; se ha de tener en cuenta que se tomó información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 por número de cédula. El total del salario necesario para el personal militar con el grado de teniente coronel será de 14'579.398,798 dólares.

A continuación se presenta en la (figura 45) la predicción del salario en un gráfico de dispersión para el año 2016 de los tenientes coroneles en forma individual que genera la regresión lineal.

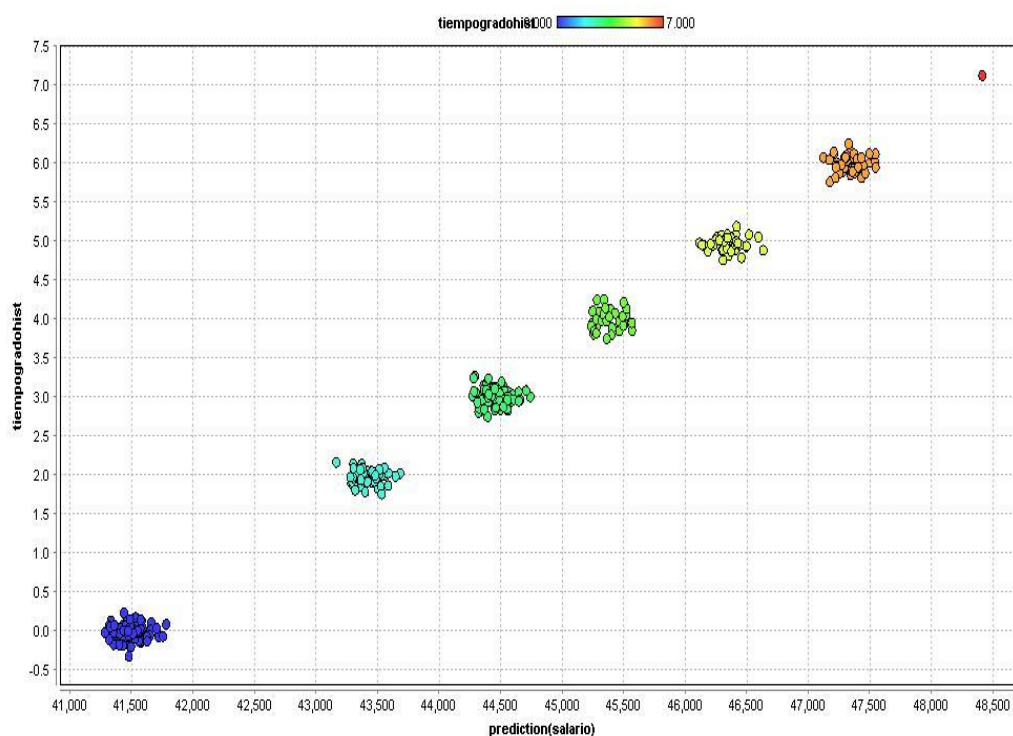


Figura 45: Predicción salario teniente coronel con regresión lineal

Del gráfico se puede predecir que personal militar con grado de teniente coronel que tiene entre 1,5 y 2,5 años de tiempo en el grado ganarán entre 43.000 y 44.000 dólares en el año 2016; se observa que encada grupo de tiempo en el grado ganaran como se visualiza; en cambio con años en el grado mayores a 7 tendrán un salario anual en el año 2016 entre 48.000 y 48.500 dólares.

3.10.6 Modelo proyección salario sargento segundo con regresión lineal

A continuación se muestra en la (figura 46) los procesos que permiten generar la proyección de salarios de los sargentos segundos aplicando regresión lineal.

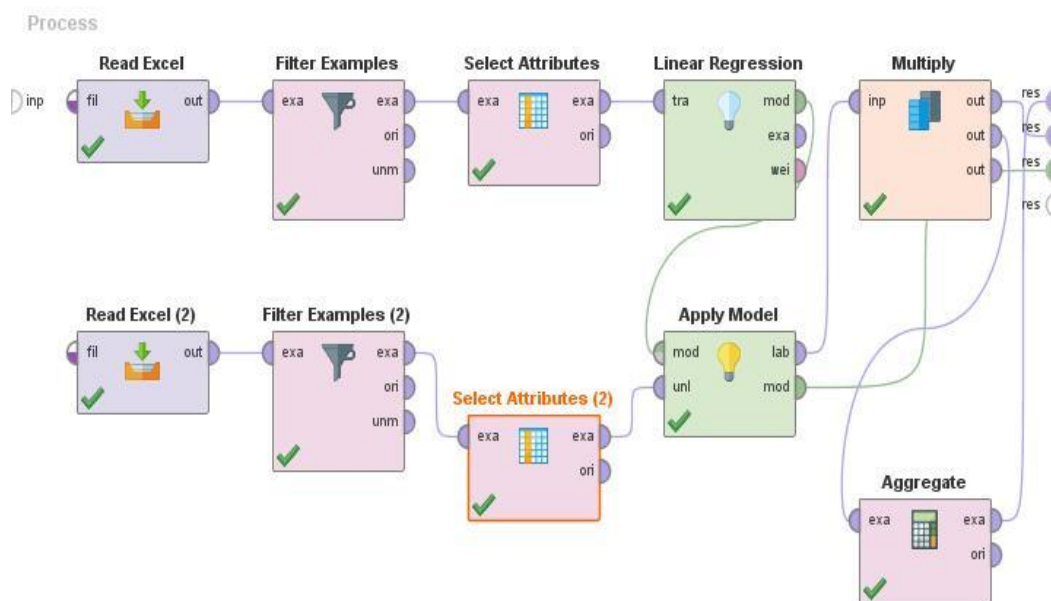


Figura 46: Modelo sargento segundo con regresión lineal

Para realizar este modelo se utilizó la base de datos con los registros del personal militar tomando en cuenta información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015; filtrando por el código de grado histórico igual a 15 que corresponde al grado de sargento segundo; seguidamente se seleccionó los siguientes atributos `anio`, `codgradohist`, `men_cedula`, `salario`, `tiempogradohist` considerados los atributos relevantes para la predicción del salario. Además se seleccionó la base de datos de predicción del personal militar para el año 2016; filtrando por el código del grado histórico igual a 15 y se aplicó el modelo de regresión lineal para la predicción del salario de cada personal militar y la suma del total de la predicción del salario necesario a través del operador `Aggregate`.

A continuación en la (figura 47) se muestra la predicción del salario en forma individual y total de los sargentos segundos utilizando regresión lineal.

Row No.	mem_cedula	prediction(salario)	sex_codigo	codgradohist	nomgradohist	tiempograd...	tiemposervi...	edad	grado	eci_codigo	estado
1	102144987	18274.931	M	15	SGOS	5	20	41	SARGENTO...	C	ACTIVO
2	102677143	18274.931	M	15	SGOS	5	20	39	SARGENTO...	C	ACTIVO
3	103334967	17923.280	M	15	SGOS	4	19	40	SARGENTO...	C	ACTIVO
4	103416376	17571.628	M	15	SGOS	3	18	39	SARGENTO...	S	ACTIVO
5	103586137	18274.931	M	15	SGOS	5	19	38	SARGENTO...	D	ACTIVO
6	103819033	17219.976	M	15	SGOS	2	16	40	SARGENTO...	C	ACTIVO
7	103916680	17923.280	M	15	SGOS	4	19	39	SARGENTO...	C	ACTIVO
8	103925517	17571.628	M	15	SGOS	3	17	36	SARGENTO...	C	ACTIVO
9	104090725	16516.673	M	15	SGOS	0	15	38	SARGENTO...	C	ACTIVO
10	201191293	18626.583	M	15	SGOS	6	21	46	SARGENTO...	C	ACTIVO
11	201195112	18626.583	M	15	SGOS	6	21	47	SARGENTO...	U	ACTIVO
12	201196177	17923.280	M	15	SGOS	4	19	45	SARGENTO...	C	ACTIVO
13	201244266	18626.583	M	15	SGOS	6	21	45	SARGENTO...	U	ACTIVO
14	201245990	18626.583	M	15	SGOS	6	21	43	SARGENTO...	C	ACTIVO
15	201261583	18626.583	M	15	SGOS	6	21	44	SARGENTO...	C	ACTIVO
16	201283603	18274.931	M	15	SGOS	5	20	44	SARGENTO...	C	DISPONIBLE

Row No.	sum(prediction(salario))
1	101731659.905

Figura 47: Predicción salario total sargento segundo con regresión lineal

Se puede apreciar la predicción del salario individual del año 2016 de todo el personal con el grado de sargento segundo; se ha de tener en cuenta que se tomó información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015 por número de cédula. El total del salario necesario para el personal militar con el grado de sargento segundo es de 101'731.659,905 dólares.

A continuación se presenta en la (figura 48) la predicción del salario en un gráfico de dispersión para el año 2016 de los sargentos segundos en forma individual que genera la regresión lineal.

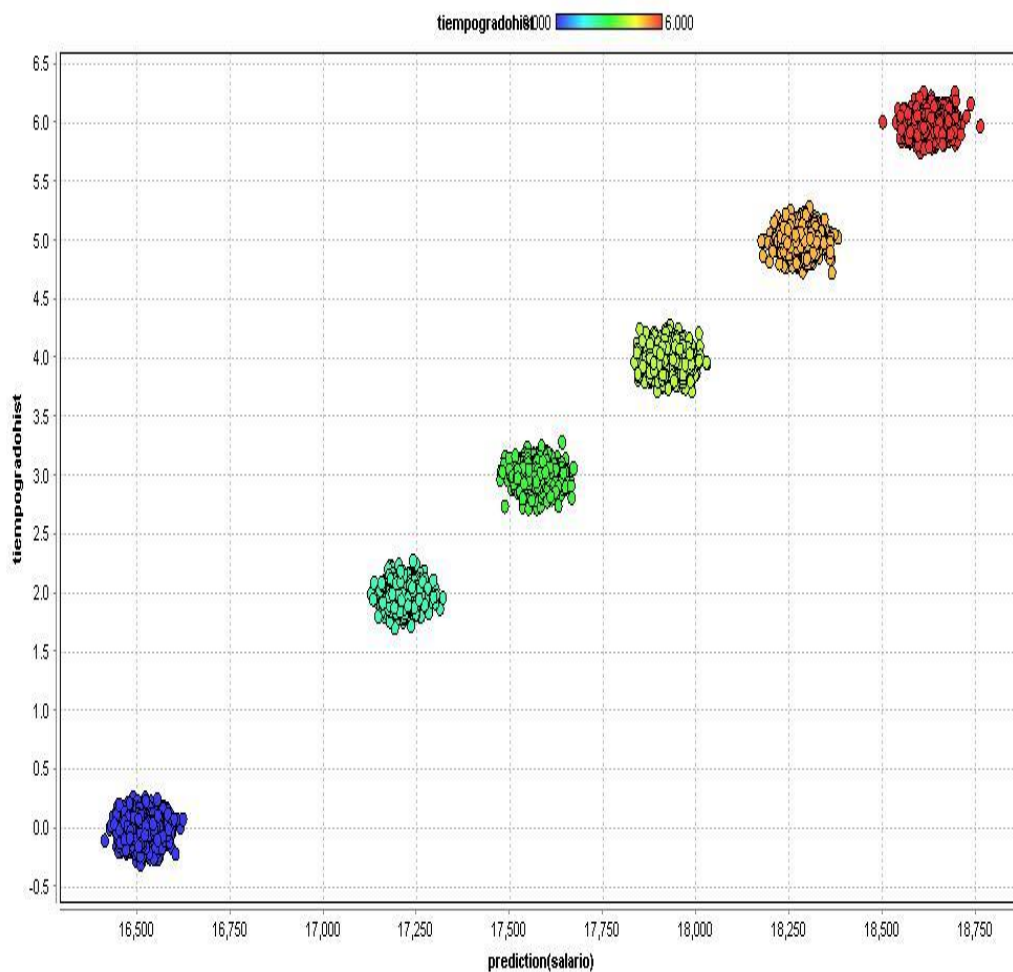


Figura 48: Predicción salario sargento segundo con regresión lineal

Del gráfico se puede predecir que personal militar con grado de sargento segundo que tiene entre 5,5 y 6,5 años de tiempo en el grado ganarán entre 18.500 y 18.750 dólares en el año 2016; se observa que encada grupo de tiempo en el grado ganarán como se visualiza; en cambio con años en el grado menores a 0,5 tendrán un salario anual en el año 2016 entre 16.000 y 16.600 dólares.

3.10.7 Total proyección con redes neuronales y regresión lineal por grados

A continuación se muestra en la (tabla 1) los salarios proyectados según el grado utilizando redes neuronales y regresión lineal del personal militar del (E.E).

Tabla 1:**Totales por grado con neural net y regresión lineal**

NOMBRE	GRADO	NEURAL NET	REGRESION LINEAL
GENERAL DE BRIGADA	3	870621,273	877043,084
CORONEL	4	10329463,450	11001789,685
TENIENTE CORONEL	5	13159848,749	14579398,798
MAYOR	6	24040307,698	28040689,087
CAPITAN	7	22255037,349	26960097,200
TENIENTE	8	13086618,525	15070589,807
SUBTENIENTE	9	7944769,483	9164867,853
SUBOFICIAL PRIMERO	12	6274893,752	6393286,636
SUBOFICIAL SEGUNDO	13	4956562,726	5100174,144
SARGENTO PRIMERO	14	35781067,236	39710475,532
SARGENTO SEGUNDO	15	90857663,365	101731659,905
CABO PRIMERO	16	79569226,699	88796169,143
CABO SEGUNDO	17	50913030,314	59620895,314
SOLDADO	18	41632955,887	46831924,884
		401'672.066,506	453'879.061,072

De la tabla se concluye que la mejor proyección salarial es la de redes neuronales, porque realizando la comparación con los datos reales del año 2016 son los que más se aproximan.

Una vez realizado el análisis de los valores con redes neuronales con tasas de aprendizajes de 0.3 e impulsos de 0.2 con un ciclo de entrenamiento de 500 y con regresión lineal comparado con los valores reales de predicción del año 2016, se

determina que por los cambios en los grados y los retiros del personal militar no es conveniente sumar las proyecciones en forma individual para general la proyección total.

Sin embargo es recomendable para realizar predicciones parciales por grados ubicar en las redes neuronales tasas de aprendizajes de 0.4 e impulsos de 0.3 con un ciclo de entrenamiento de 500 lo que genera una mejor aproximación, esto se puede afirmar porque la configuración de estos rangos se los realizó modificando cada uno de estos parámetros hasta llegar al que daba la mínima diferencia de error; y a la vez se recomienda trabajar con redes neuronales que son las que realmente se aproximan a los salarios individuales.

A continuación se puede visualizar en la (figura 49) la red neuronal por subprocesos que contiene cada uno de los grados del personal militar que permite determinar el total individual y general del salario por grado; clasificado por código de grado lo que permite llegar a determinar el total de la proyección del salario.

3.10.8 Modelo de Predicción de sueldos militar con redes neuronales

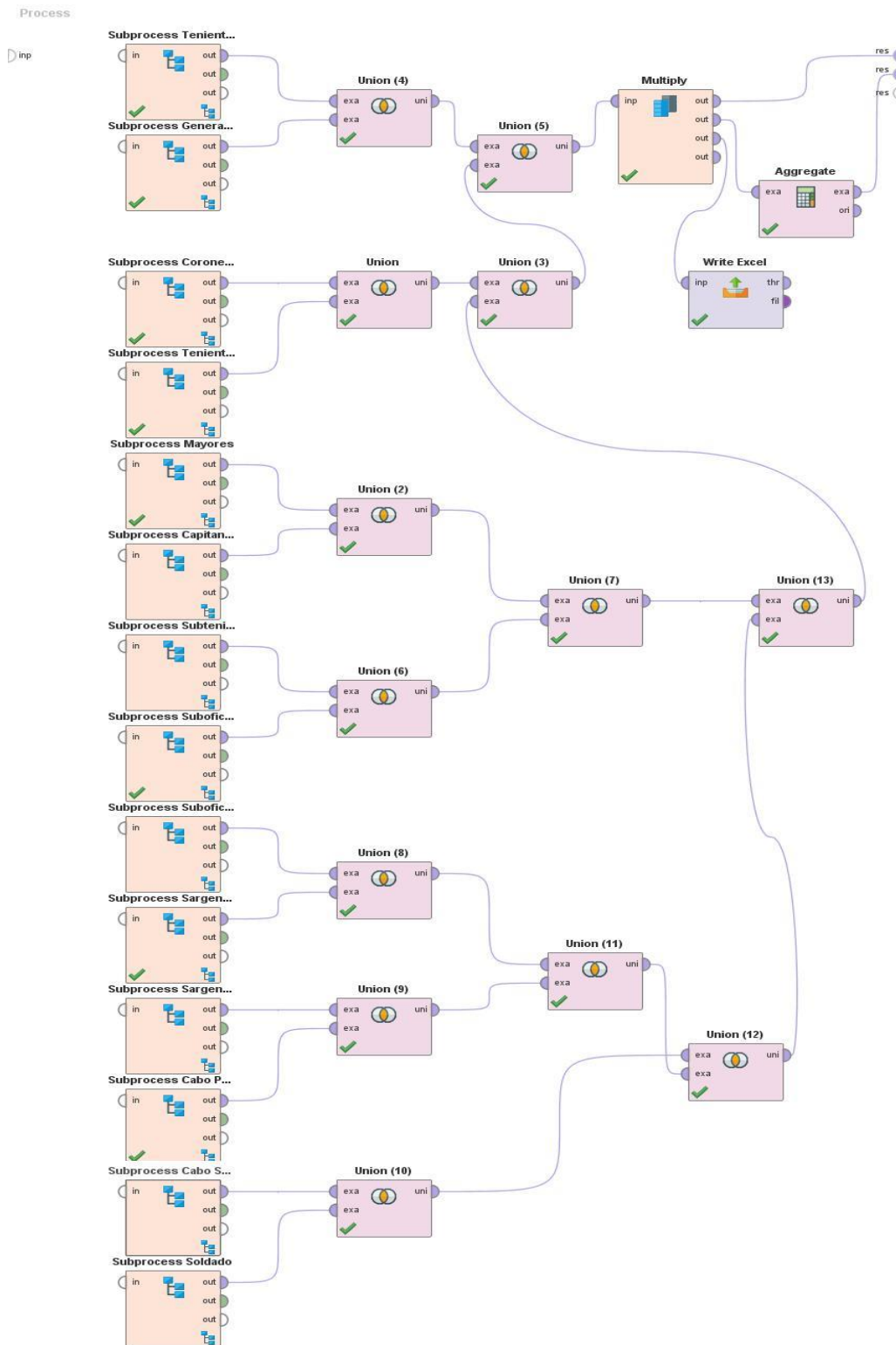


Figura 49: Subprocesos de todo el personal militar con redes neuronales

Para realizar este modelo se utilizó la base de datos con los registros del personal militar tomando en cuenta información de los años 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015; seguidamente se seleccionó los siguientes atributos año, codgradohist, men_cedula, salario, tiempogradohist considerados los atributos relevantes para la predicción del salario. Además se seleccionó la base de datos de predicción del personal militar para el año 2016; luego se eligió los siguientes atributos año, codgradohist, men_cedula, tiempogradohist y se aplicó el modelo de Redes Neuronales para la predicción del salario de cada personal militar y la suma del total de la predicción del salario necesario para el año 2016 a través del operador Aggregate.

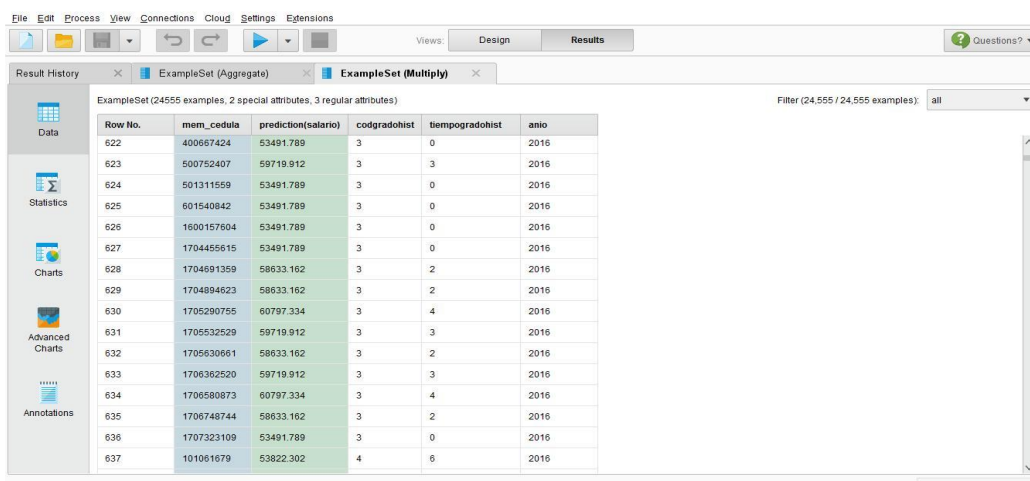
Seguidamente se muestra en la (figura 50) el total de la predicción generada utilizando redes neuronales.



Row No.	sum(prediction(salario))
1	403531940.236

Figura 50: Suma de predicción por grados con redes neuronales

Se puede visualizar en la figura (51) que la predicción individual por grados dentro de lo que es personal militar tiene un grado de error del $\pm 2\%$ que es un valor tolerable dentro de lo que son cálculos estadísticos.



Row No.	mem_cedula	prediction(salario)	codgradohist	tiempogradohist	año
622	400667424	53491.789	3	0	2016
623	500752407	59719.912	3	3	2016
624	501311559	53491.789	3	0	2016
625	601540842	53491.789	3	0	2016
626	1600157604	53491.789	3	0	2016
627	1704455615	53491.789	3	0	2016
628	1704691359	58633.162	3	2	2016
629	1704894623	58633.162	3	2	2016
630	1705290755	60797.334	3	4	2016
631	1705532529	59719.912	3	3	2016
632	1705630661	58633.162	3	2	2016
633	1706362520	59719.912	3	3	2016
634	1706580873	60797.334	3	4	2016
635	1706748744	58633.162	3	2	2016
636	1707323109	53491.789	3	0	2016
637	101061679	53822.302	4	6	2016

Figura 51: Predicción individual por grados con redes neuronales

A continuación se puede visualizar en la (figura 52) que el valor global de toda la predicción se tiene un error acumulado de $\pm 3\%$ que es un valor ya considerable porque se maneja cifras de millones de dólares americanos.

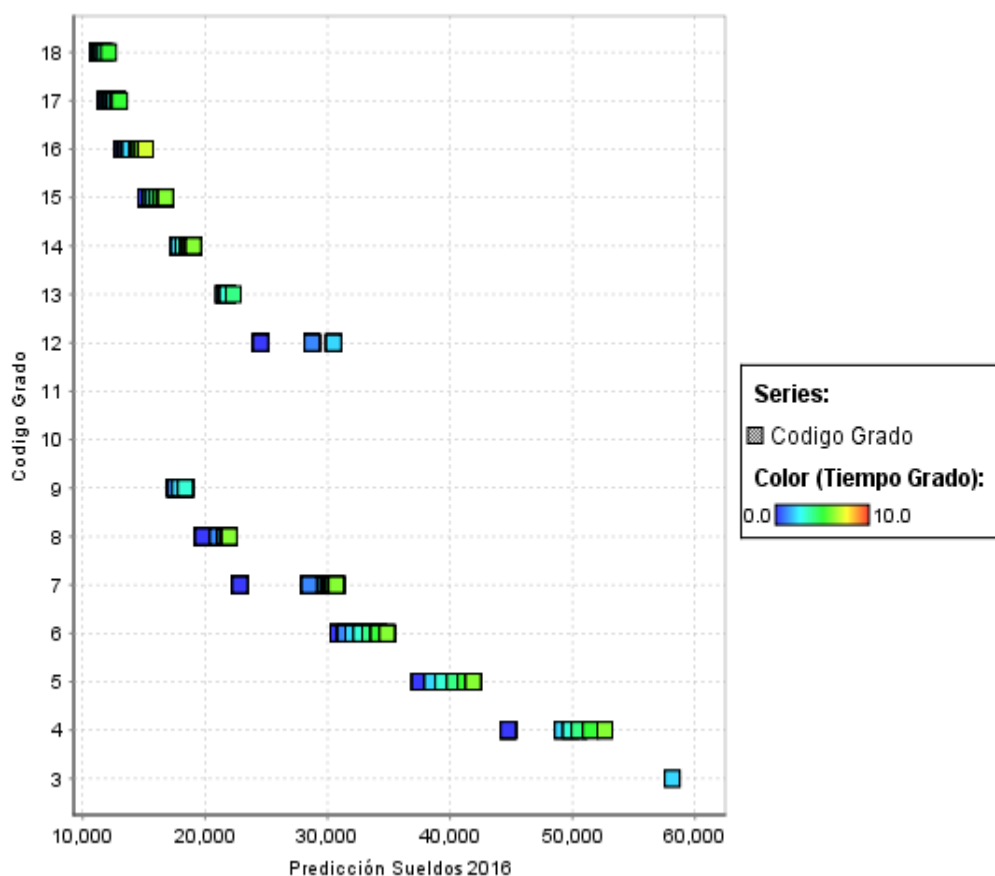


Figura 52: Predicción de sueldo al 2016 por grados

Se puede visualizar en la figura 40 que los sueldos más altos se concentran en los grados jerárquicos más altos, estos son los oficiales en grados de Generales y Coroneles que son relativamente pocos en relación a la gran mayoría que tiene los sueldos más bajos son los grados jerárquicos más bajos y donde se concentran el personal de voluntarios o tropa.

En la (figura 53) se puede visualizar el numérico del personal procesado en este modelo para la predicción del 2016.

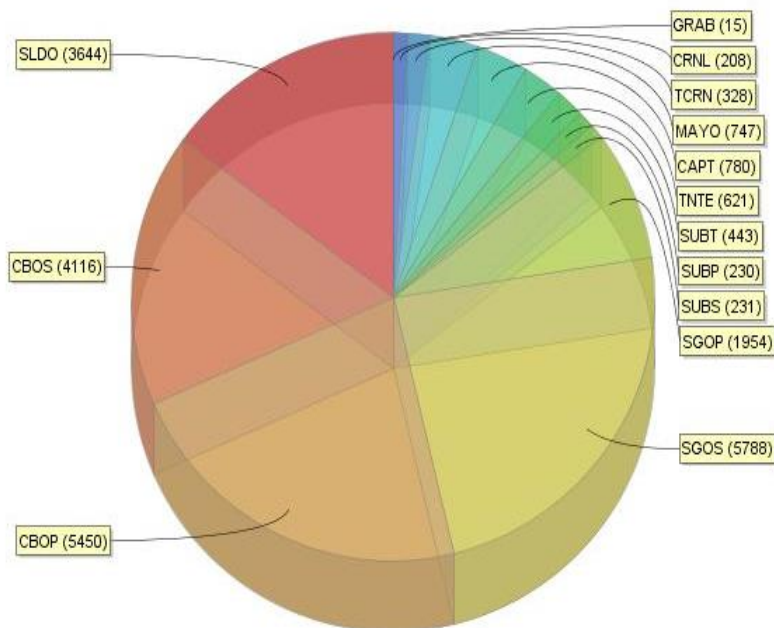


Figura 53: Numérico del personal militar por grados

Seguidamente se puede apreciar en la (figura 54) la proyección del total del salario necesario para el personal militar para el año 2016.

Row No.	sum(prediction(salario))
1	453879061.072

Figura 54: Suma de predicción por grados con regresión lineal

3.10.8.1 Determinación del modelo adecuado para solucionar el problema

Siguiendo los mismos pasos de la red neuronal y únicamente cambiando por el operado Linear Regression se obtuvo como predicción que se necesita para cubrir salarios del personal militar 453'879.061,072 en el año 2016; lo que realmente no se aproxima a la realidad; motivo por el cual se rechaza este modelo.

Seguidamente en la figura 55 se puede observar la proyección del salario para el año 2016 por cada una de las provincias.

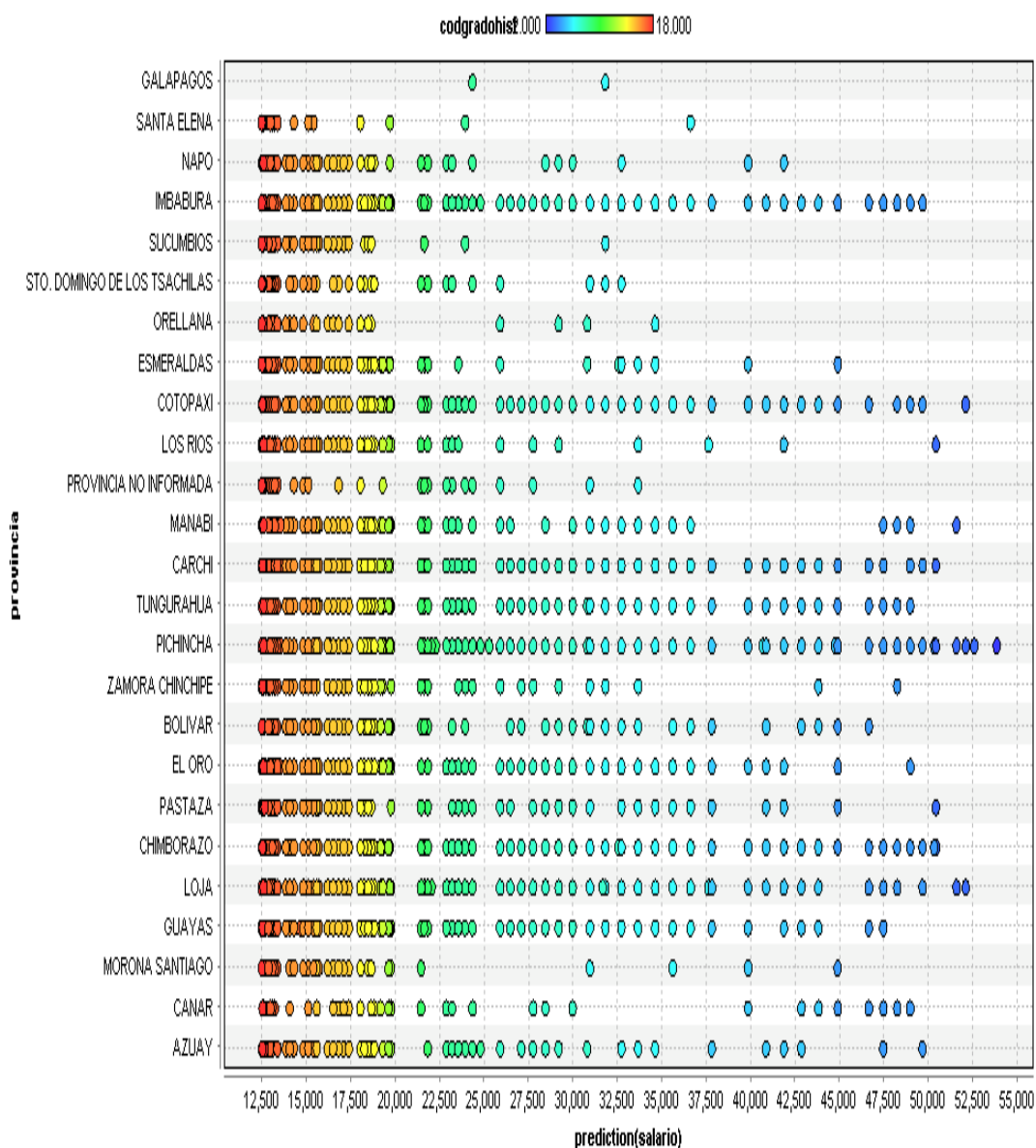


Figura 55: Proyección de salarios por provincias con n. n

Como se puede apreciar en el gráfico la provincia de Pichincha es donde más se gasta en salarios del personal militar llegando aproximadamente a 54.000 dólares anuales seguido de las provincias de Cotopaxi con 52.500 dólares, luego se tiene Chimborazo con 51.000 dólares, Guayas alcanza los 47.500 dólares corresponde a los grados más altos del (E.E). Galápagos es donde menos se gasta pero el personal militar que existe llegan aproximadamente a 32.000 dólares.

A continuación se visualiza en la (figura 56) la proyección de salario del año 2016 por estado civil utilizando redes neuronales sobre la totalidad de los datos del personal militar.

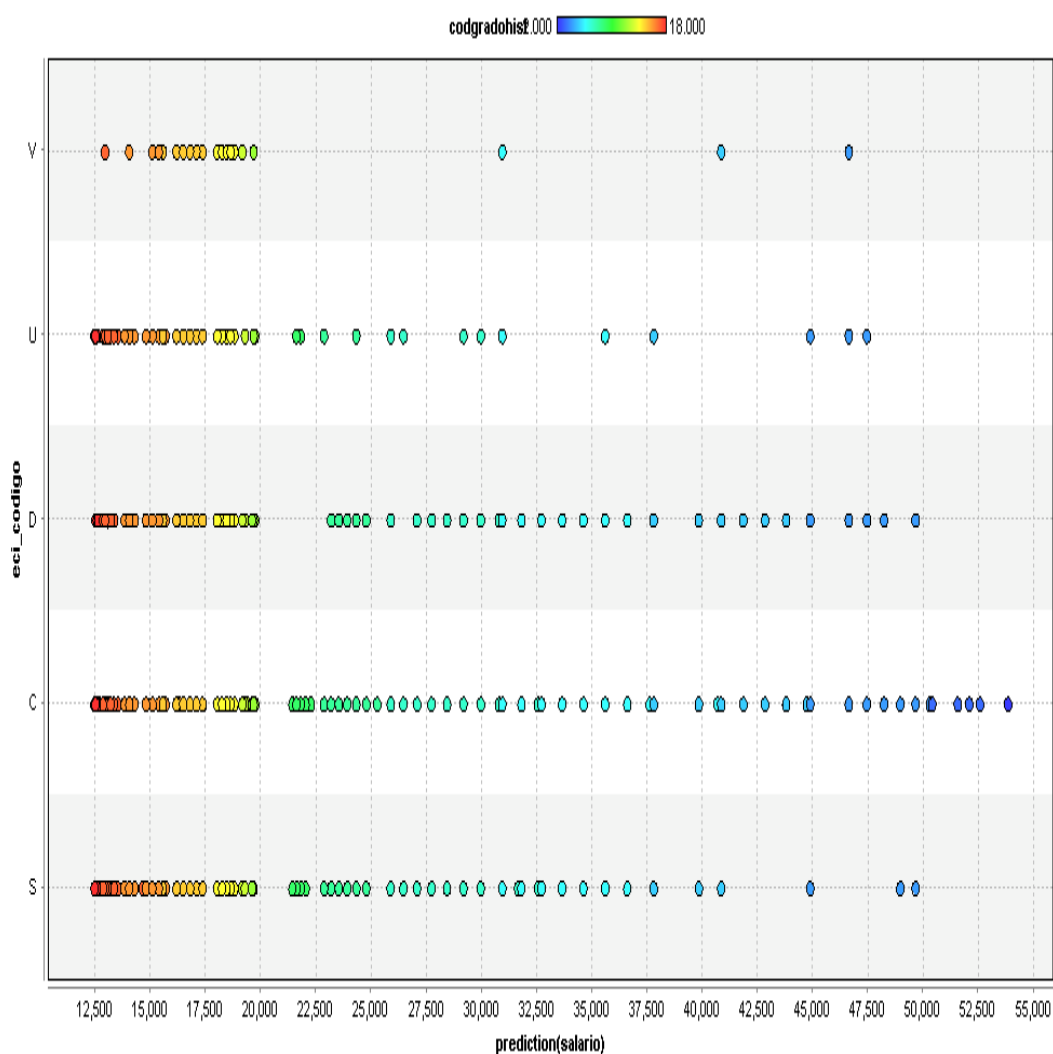


Figura 56: Proyección de salarios por estado civil con n. n

Una vez generado la figura se determinó que el personal militar de estado civil casado es quien más existe y que alcanza los 54.000 dólares, seguido de los solteros y divorciados con 50.000 dólares aproximadamente, algo que llama la atención es la cantidad de divorciados que existen por lo que se deberá realizar un estudio a fin de determinar las causas y poder encontrar los problemas para dar soluciones al mismo; lo cual será otro tema de investigación.

A continuación se puede apreciar en la (figura 57) la proyección de salario del año 2016 por fecha de estado utilizando redes neuronales sobre la totalidad de los datos del personal militar.

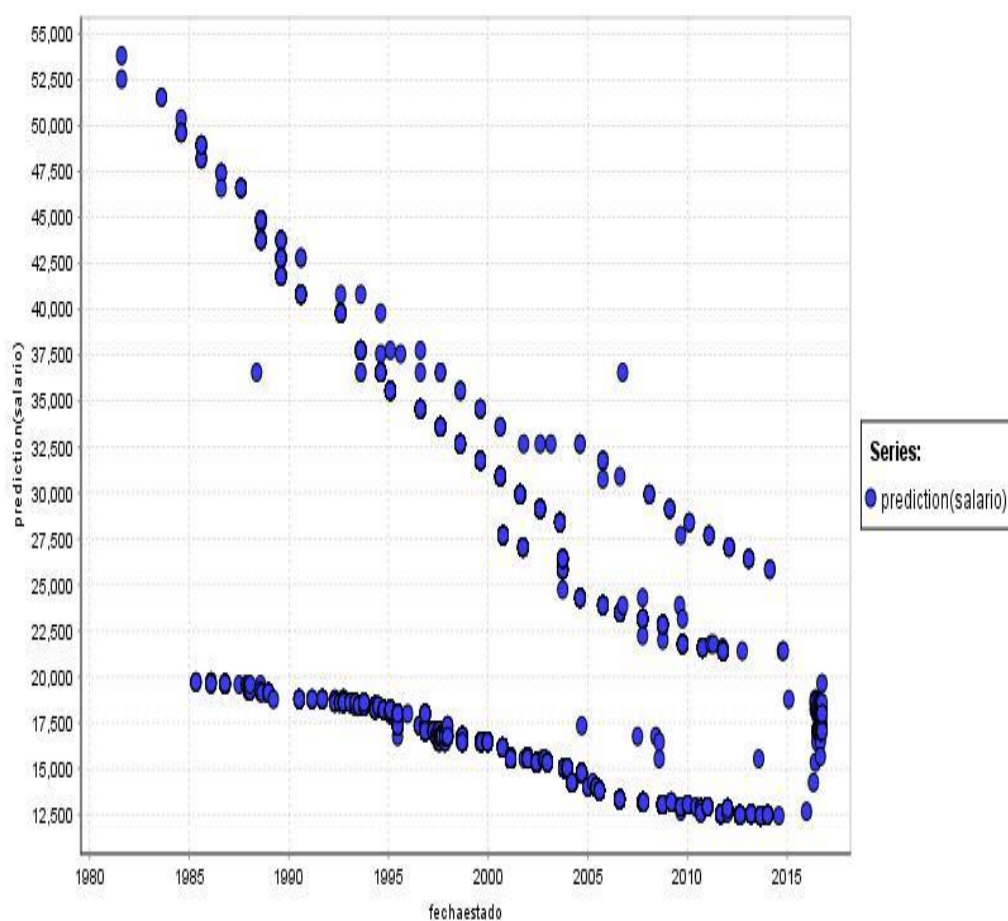


Figura 57: Predicción del salario por fecha de estado

Una vez generado la figura se pudo determinar que el personal militar con fecha de estado mayor o igual a 1980 y menor o igual a 1985; es cuando obtuvieron los salarios más altos que oscilaban entre 50.000 y 54.000 dólares en cambio cuando bajaron a salarios entre 15.000 y 20.000 dólares fueron la fecha de estado superior al 2015.

Seguidamente en la (tabla 2) se puede observar la diferencia en dólares con relación a total real del año 2016 con regresión lineal y red neuronal.

3.10.9 Total proyección con redes neuronales y regresión lineal

Tabla 2:

Diferencia proyección neural net y regresión lineal

Total por grados con lo real (enlaces)	
Proyectado 2016	415'623.952,220
Neural net	403'531.940,236
Diferencia	12'092.011,984

Total por grados con lo real (enlaces)	
Proyectado 2016	415'623.952,220
Regresión lineal	453'879.061,072
Diferencia	38'255.108,852

Una vez realizado el análisis de los valores con redes neuronales con tasas de aprendizajes de 0.4 e impulsos de 0.3 con un ciclo de entrenamiento de 500 y con regresión lineal comparado con los valores reales de predicción del año 2016 se determina que la red neuronal es la que genera mejor predicción comparado con los datos reales.

3.10.10 Propuesta de modelo de remuneraciones

Se ha creído necesario manejar la opción del modelo de predicción completo en redes neuronales para predecir el valor casi exacto de forma global porque el margen de error es muy bajo $\pm 0.17\%$ que al manejar cifras de millones de dólares es casi ínfima, a continuación se presenta el modelo completo de redes neuronales sin filtros.

Seguidamente se puede observar en la (figura 58) los procesos utilizando todos los datos del personal militar a través de redes neuronales, donde intervienen los atributos que son necesarios en nuestro proyecto de investigación.

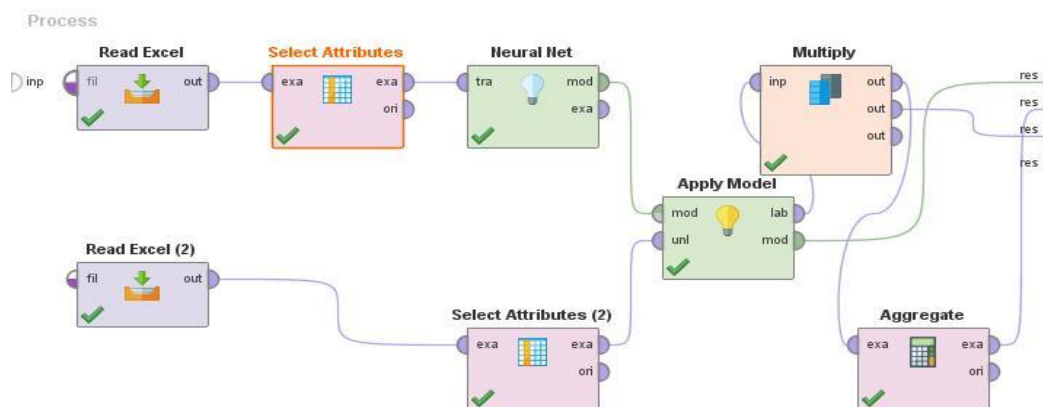


Figura 58: Modelo de predicción completo con redes neuronales

Se puede observar en la (figura 59) los resultados de este modelo que los valores individuales ya tienen errores de $\pm 6\%$ que es un error considerable si se toma los valores individuales pero este modelo compensa en algunos grados

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

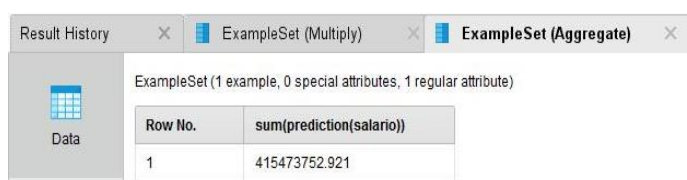
Result History x ExampleSet (Multiply) x ExampleSet (Aggregate) x ImprovedNeuralNet (Neural Net) x

ExampleSet (24560 examples, 2 special attributes, 3 regular attributes) Filter (24,560 / 24,560 examples): all

Row No.	mem_cedula	prediction(salario)	codgradohist	tiempogradohist	anio
1	101061679	49526.237	4	6	2016
2	101505493	42853.109	5	5	2016
3	101661023	39664.021	5	2	2016
4	101719987	46441.394	4	2	2016
5	101751105	19462.289	12	2	2016
6	101789089	49526.237	4	6	2016
7	101876928	29146.454	7	5	2016
8	101943314	48063.526	4	4	2016
9	102022670	19462.289	12	2	2016
10	102114683	39664.021	5	2	2016
11	102133576	37635.240	5	0	2016
12	102144987	17013.204	15	5	2016
13	102157328	39664.021	5	2	2016
14	102245479	37635.240	5	0	2016
15	102253218	37635.240	5	0	2016
16	102286226	47271.697	4	3	2016

Figura 59: Predicción total por grados con redes neuronales

Al compensar los valores perdidos en unos grados y ganados en otros se puede observar en la (figura 60) que el valor total es muy similar al valor real que se gastara en el año 2016 únicamente con un error de $\pm 0.17\%$



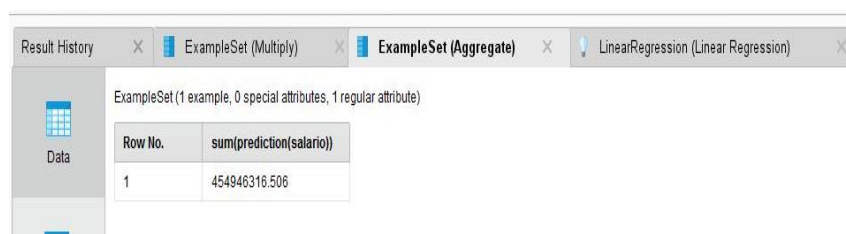
Row No.	sum(prediction(salario))
1	415473752.921

Figura 60: Suma de predicción total con redes neuronales

En cambio con la predicción utilizando Regresión Lineal se puede observar en la (figura 61) que en los valores individuales son de $\pm 3.5\%$ que es un valor ya muy grande y en el valor total se tiene por lo tanto un valor mayor de error acumulado de $\pm 9\%$.

Es pertinente mencionar el modelo con redes neuronales que se ha realizado con subprocessos específicos para cada grado militar permite poder llegar a resultados con un grado de error muy bajo y de esta manera tener una predicción más acercada a la realidad.

Ningún modelo de predicción es 100% libre de errores pero el modelo de redes neuronales es el que más se aproxima a valores reales tomando en cuenta que son millones de dólares.



Row No.	sum(prediction(salario))
1	454946316.506

Figura 61: Suma de predicción total con regresión lineal

Siguiendo los mismos pasos de la red neuronal y únicamente cambiando por el operado Linear Regression se obtuvo como predicción que se necesita para cubrir salarios del personal militar 454'946.316,506 en el año 2016; lo que realmente no se aproxima a la realidad; motivo por el cual se rechaza este modelo.

Seguidamente en la (tabla 3) se puede observar y analizar la diferencia que se ha generado utilizando redes neuronales y regresión lineal considerando el total de datos en la presente investigación para el año 2016.

3.10.11 Total proyección con redes neuronales y regresión lineal

Tabla 3:

Diferencia entre neural net, regresión lineal con datos reales

Proyectado 2016	415'623.952,220
Neural net	415'473.752,921
Diferencia	150.199,299

Proyectado 2016	415'623.952,220
Regresión lineal	454'946.316,506
Diferencia	39'322.364,286

Una vez realizado el análisis de los valores con redes neuronales con tasas de aprendizajes de 0.4 e impulsos de 0.3 con un ciclo de entrenamiento de 500 y con regresión lineal comparado con los valores reales de predicción del año 2016 se determina que la red neuronal es la que arroja mejor predicción comparado con los datos reales. Además se debe aplicar sobre el total de las bases de datos sin separar por grados debido a los ascensos y salidas que se producen en el personal militar; cuando interviene el total de los datos con las variables independientes necesarias en el proyecto de investigación; generan la mejor proyección comparada con el total del presupuesto del salario para el año 2016 de acuerdo a los datos reales que facilitó el departamento de remuneraciones del personal de la fuerza terrestre.

CAPÍTULO 4

RESULTADOS

4.1 Informe de resultados

4.1.1 Antecedentes

El Ejército Ecuatoriano cuenta con un sistema informático robusto y una base de datos en Oracle, en la que se encuentra la información relacionada a todo su manejo incluido las remuneraciones del personal militar y servidores públicos, pero no cuentan con herramientas ni personal especializado en inteligencia de negocio, es por esta razón que se procedió a desarrollar este tema de tesis para abrir una ventana con la que se pueda iniciar la capacitación del personal para el manejo de herramientas para inteligencia de negocios y de esta manera los mandos de la institución puedan tomar decisiones más acertadas en beneficio de la institución.

4.1.2 Investigación

Dentro de la investigación realizada para poder generar los modelos de predicción para remuneraciones, se pudo observar las diferentes formas de cómo debe ser visualizada la información base para poder lograr una proyección óptima de los resultados.

Se procedió a realizar la investigación probando dos modelos, la primera fue regresión lineal y la segunda redes neuronales.

4.1.3 Desarrollo

4.1.3.1 Regresión Lineal

Dentro de la herramienta RapidMiner se utilizó la función linear regression para proyectar los sueldos básicos unificados del personal militar pero al realizar las pruebas con datos reales se obtiene un error mayor a lo esperado en el cálculo por grados militares y al momento de generar la información con todos los grados el error acumulado aumenta, es por esta razón que no se optó por seleccionar esta herramienta para la proyección definitiva.

A continuación en la (figura 62) se presenta el modelo de regresión lineal para el grado de coronel.

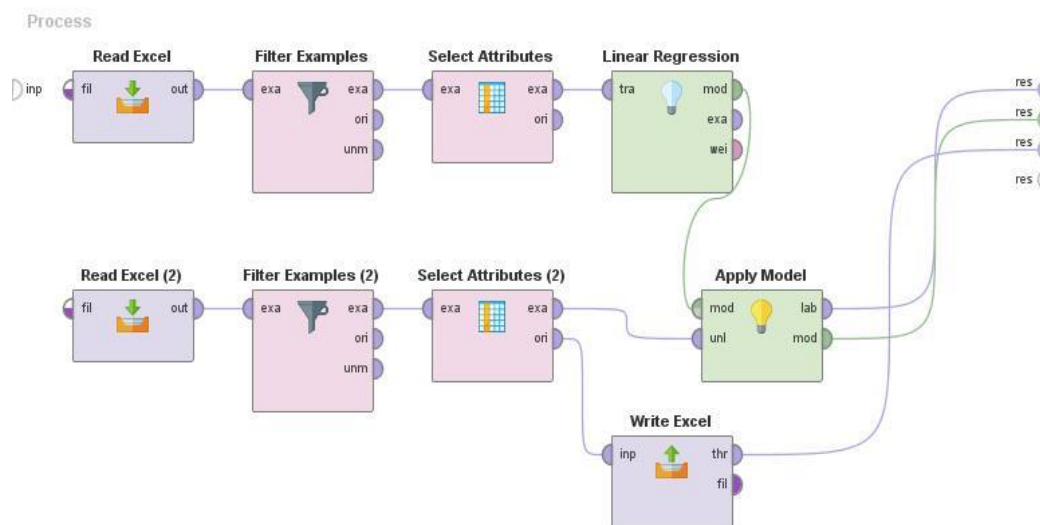


Figura 62: modelo de regresión lineal coroneles proyección 2016

A continuación en la (figura 63) se puede observar la proyección de salarios de los coroneles en forma individual.

ExampleSet (208 examples, 2 special attributes, 3 regular attributes) Filter (208 / 208 examples):

Row No.	MEM_CEDULA	prediction(S...	CODGRADO...	TIEMPOGRA...	ANIO
1	101061679	55816.795	4	6	2016
2	101719987	52068.853	4	2	2016
3	101789089	55816.795	4	6	2016
4	101943314	53942.824	4	4	2016
5	102286226	53005.838	4	3	2016
6	200938223	52068.853	4	2	2016
7	200955573	52068.853	4	2	2016
8	200987600	50194.882	4	0	2016
9	300743028	50194.882	4	0	2016
10	300832706	52068.853	4	2	2016
11	300879665	53942.824	4	4	2016
12	400637757	52068.853	4	2	2016
13	400646642	55816.795	4	6	2016
14	400675146	54879.809	4	5	2016
15	400694220	54879.809	4	5	2016
16	400696894	52068.853	4	2	2016

Figura 63: Proyección sueldo coroneles individual 2016 con r. 1

Un señor coronel que tiene de 6 a 7 años en el grado normalmente gana un sueldo básico unificado de 52593 USD al año, la predicción por el mismo coronel utilizando el modelo de regresión lineal da como resultado 55816,795 que da una diferencia de 3223.795 y esto da un porcentaje de error de $\pm 5.77\%$ que es un error alto para una predicción de millones de dólares que maneja el Ejército Ecuatoriano.

4.1.3.2 Redes Neuronales

Según las pruebas realizadas en el modelo de predicción con redes neuronales los resultados obtenidos son los más cercanos a la realidad es decir el error obtenido no es muy alto por lo tanto este modelo es el seleccionado para la proyección real de los sueldos del personal militar para los años 2016, 2017 y 2018 que se ha creído pertinente realizarlo ver (figura 64).

Se señala que cada uno de los atributos relevantes después de realizados los filtros correspondientes para nuestra investigación han sido conectadas a una entrada de la red neuronal; se ha de tomar en cuenta que estos corresponden a los datos históricos de los salarios del personal militar desde el año 2007 al 2015; utilizando un perceptrón multicapa luego pasan a varias capas ocultas para ser procesada la proyección del salario de cada personal militar generando la proyección en una sola capa de salida en la que se puede observar en forma individual el resultado correspondiente de cada uno de ellos.

Modelo matemático que representa una neurona es:

$$y = g(b + \sum_{i=1}^n w_i x_i)$$

Fuente: (Raveane & López , 2014)

Este modelo sería otro proyecto de investigación futura.

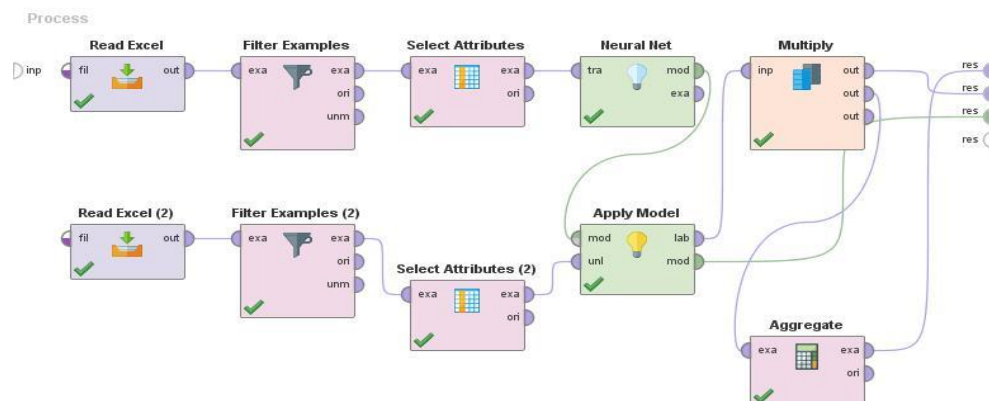


Figura 64: Modelo de redes neuronales de coroneles proyección 2016

Seguidamente en la (figura 65) se puede observar la proyección de salarios del personal militar con el grado de coronel para el año 2016.

ExampleSet (208 examples, 2 special attributes, 3 regular attributes) Filter (208 / 208 examples): all

Row No.	MEM_CEDULA	prediction(S...	CODGRADO...	TIEMPOGRA...	ANIO
1	101061679	52644.439	4	6	2016
2	101719987	49177.491	4	2	2016
3	101789089	52644.439	4	6	2016
4	101943314	50528.170	4	4	2016
5	102286226	49846.497	4	3	2016
6	200938223	49177.491	4	2	2016
7	200955573	49177.491	4	2	2016
8	200987600	44748.418	4	0	2016
9	300743028	44748.418	4	0	2016
10	300832706	49177.491	4	2	2016
11	300879665	50528.170	4	4	2016
12	400637757	49177.491	4	2	2016
13	400646642	52644.439	4	6	2016
14	400675146	51451.799	4	5	2016
15	400694220	51451.799	4	5	2016
16	400696894	49177.491	4	2	2016
17	400702064	51451.799	4	5	2016

Figura 65: Proyección neural net de sueldo coroneles 2016

De igual manera que en el modelo anterior se tiene a un señor coronel de 6 a 7 años en el grado y tiene un sueldo básico unificado de 52593 USD en un año y la proyección de sueldos según el modelo de redes neuronales es de 52644.439 USD en un año se tiene una diferencia de 51,439 que relativamente es un error de $\pm 0.09\%$ y es bastante bajo, es por esta razón que se tomó este modelo como definitivo para la proyección de sueldos del Ejército Ecuatoriano.

4.1.3.3 Redes Neuronales con todos los grados militares.

Es necesario especificar que para cada una de las proyecciones de sueldos prácticamente se utiliza el mismo modelo de redes neuronales que se usó para los señores coroneles pero realizando un proceso para cada uno de los grados por medio de una herramienta de filtrado que proporciona RapidMiner, esto es necesario porque si se realiza un solo modelo sin filtrado los resultados no son iguales tienen una

variación considerable en el valor calculado de los sueldos justamente por la diferencia de valores que existe entre oficiales y voluntarios e inclusive dentro de esos grupos existen diferencias considerables entre los grados militares, es por esta razón que se ha optado por utilizar el operador unión entre procesos para que al final obtener un solo archivo de predicción de sueldos con valores más aproximados a la realidad y en conjunto.

4.1.3.4 Proyección para el año 2016.

Para realizar la proyección de sueldos de todo el personal militar fue necesario realizar un modelo por cada grado militar dentro de un proceso y posteriormente unirlos por medio de un operador unión que al final da como resultado la proyección de todo el personal militar que labora en el Ejército Ecuatoriano.

A continuación en la (figura 66) se procede a analizar el resultado obtenido con el modelo de redes neuronales completo:

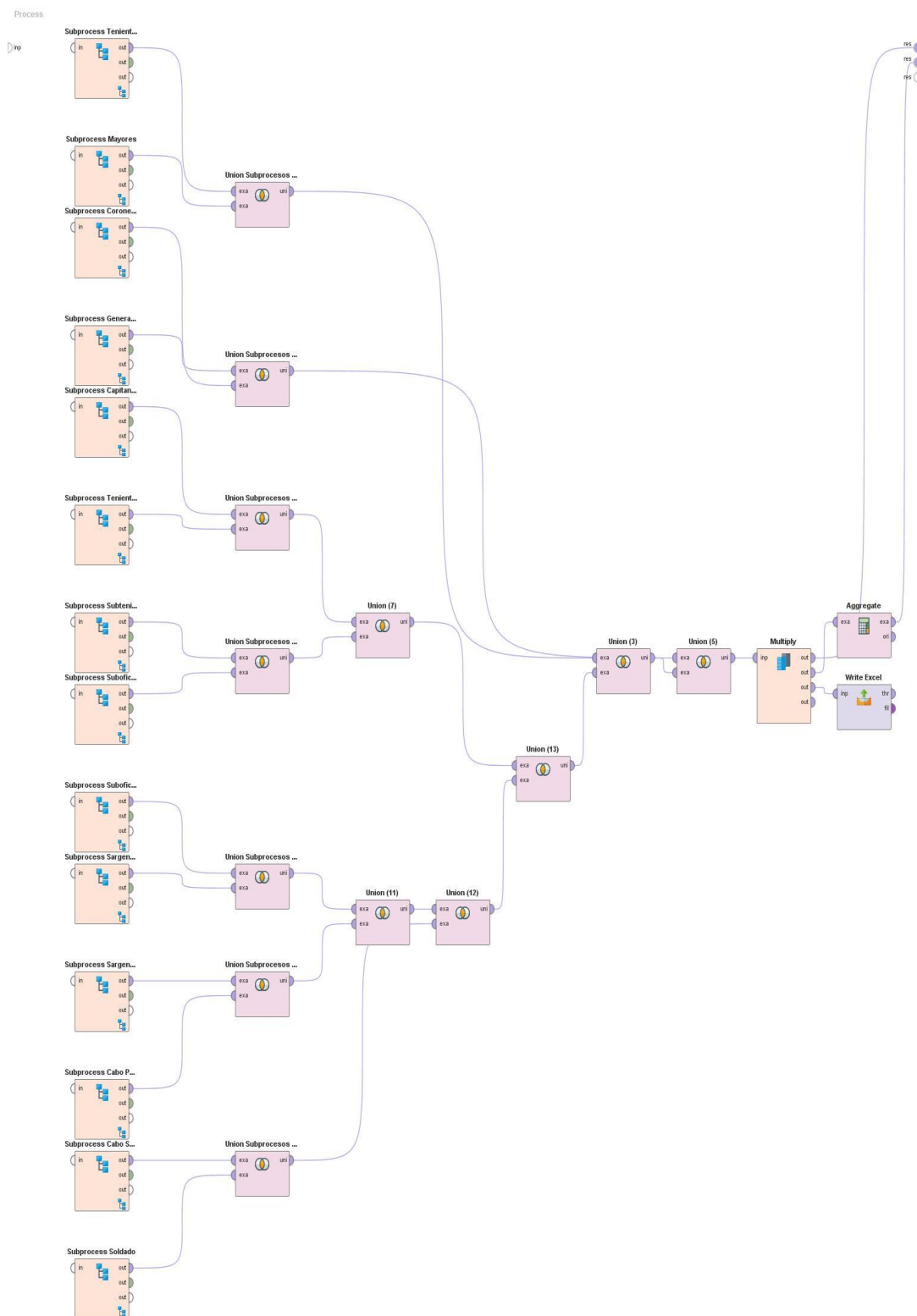


Figura 66: Modelo de proyección neural net de sueldo completo 2016

A continuación en la (figura 67) se puede observar la proyección de salarios para el año 2016.

ExampleSet (24555 examples, 2 special attributes, 12 regular attributes) Filter (24,555 / 24,555 examples): all

Row No.	MEM_CEDULA	prediction(S...	SEX_CODIGO	CODGRADO...	NOMGRADO...	TIEMPOGRA...	TIEMPOSER...	EDAD	GRA
1	400667424	55546.452	M	3	GRAB	0	32	53	GEN
2	500752407	59328.602	M	3	GRAB	3	34	57	GEN
3	501311559	55546.452	M	3	GRAB	0	32	52	GEN
4	601540842	55546.452	M	3	GRAB	0	32	53	GEN
5	1600157604	55546.452	M	3	GRAB	0	32	54	GEN
6	1704456615	55546.452	M	3	GRAB	0	32	53	GEN
7	1704691359	58152.566	M	3	GRAB	2	33	59	GEN
8	1704894623	58152.566	M	3	GRAB	2	33	53	GEN
9	1705290755	60503.224	M	3	GRAB	4	35	55	GEN
10	1705532529	59328.602	M	3	GRAB	3	34	54	GEN
11	1705630661	58152.566	M	3	GRAB	2	33	53	GEN
12	1706362520	59328.602	M	3	GRAB	3	34	54	GEN
13	1706580873	60503.224	M	3	GRAB	4	35	55	GEN
14	1706748744	58152.566	M	3	GRAB	2	33	52	GEN
15	1707323109	55546.452	M	3	GRAB	0	32	54	GEN
16	101061679	52644.439	M	4	CRNL	6	29	62	COR

Figura 67: Resultados de proyección neural net de sueldo completo 2016

Para analizar los datos del modelo de redes neuronales para todos los grados militares durante el año 2016 se toma como referencia el sueldo de un general de brigada de 0 a 1 año en el grado es de 54498 USD anual utilizando nuestro modelo propuesto se observa que el sueldo del mismo general proyectado al 2016 es de 55546.452 USD obteniendo una diferencia de 1047,452 en porcentaje da un error del 1.88% que es un valor tolerable para la proyección de sueldos del personal militar para el año 2016.

4.1.3.5 Proyección para el año 2017

Para realizar la proyección de sueldos del personal militar correspondiente al año 2017 se debe añadir a los datos iniciales los datos obtenidos en la proyección de sueldos del 2016 debido a que con los datos del sueldo para el año 2016 se completaría la información para realizar la proyección de mencionado año, es necesario también señalar que el modelo de proyección es el mismo utilizado para la proyección del año 2016, la única diferencia son los datos puesto que en los datos de origen se aumentaron

los datos del año 2016 y en los datos de proyección se actualizaron los grados y tiempo en el grado de cada una de las personas que intervienen en esta proyección, a continuación se detalla en la (figura 68) el análisis de este proceso:

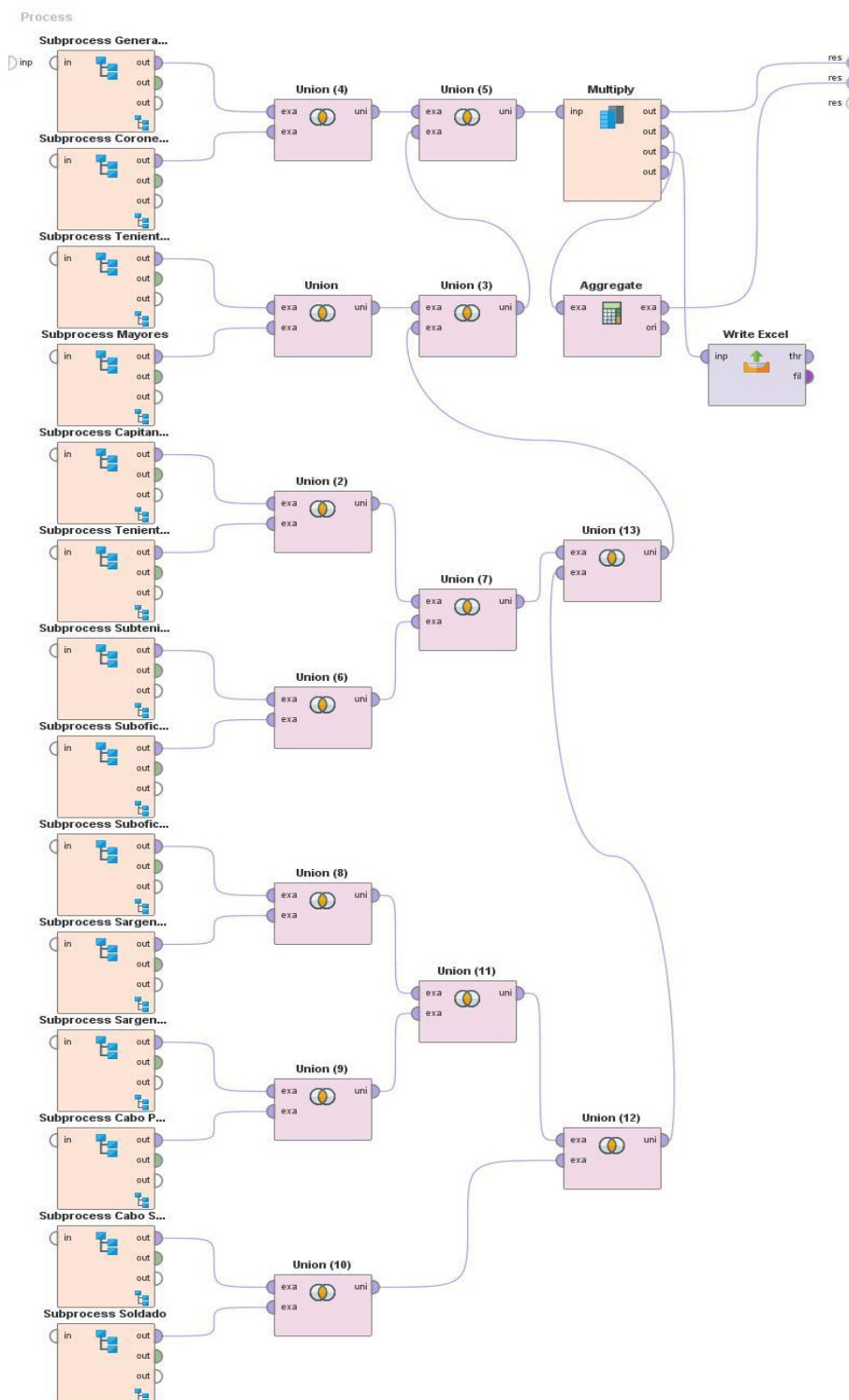


Figura 68: Modelo de proyección neural net de sueldo completo 2017

A continuación se puede observar en la (figura 69) los salarios del personal militar para el año 2017.

ExampleSet (24510 examples, 2 special attributes, 12 regular attributes) Filter (24,510 / 24,510 examples): all

Row No.	MEM_CEDULA	prediction(S...	SEX_CODIGO	CODGRADO...	NOMGRADO...	TIEMPOGRA...	TIEMPOSER...	EDAD	GRA
1	101061679	55787.314	M	3	GRAB	0	30	63	COR
2	101789089	55787.314	M	3	GRAB	0	33	55	COR
3	400646642	55787.314	M	3	GRAB	0	33	55	COR
4	400667424	57067.857	M	3	GRAB	1	33	54	GEN
5	500752407	60235.655	M	3	GRAB	4	35	58	GEN
6	501146336	55787.314	M	3	GRAB	0	33	56	COR
7	501311559	57067.857	M	3	GRAB	1	33	53	GEN
8	601050305	55787.314	M	3	GRAB	0	33	55	COR
9	601540842	57067.857	M	3	GRAB	1	33	54	GEN
10	601569783	55787.314	M	3	GRAB	0	33	57	COR
11	1001143336	55787.314	M	3	GRAB	0	33	55	COR
12	1001302288	55787.314	M	3	GRAB	0	33	55	COR
13	1001354289	55787.314	M	3	GRAB	0	33	55	COR
14	1101833893	55787.314	M	3	GRAB	0	33	54	COR
15	1102111406	55787.314	M	3	GRAB	0	33	53	COR
16	1600157604	57067.857	M	3	GRAB	1	33	55	GEN

Figura 69: Resultados de proyección neural net de sueldo completo 2017

Para analizar los datos del modelo de redes neuronales para todos los grados militares durante el año 2017 se toma como referencia el sueldo de un general de brigada de 0 a 1 año en el grado es de 54498 USD anual utilizando nuestro modelo propuesto se observa que el sueldo del mismo general proyectado al 2017 es de 55787.314 USD obteniendo una diferencia de 1289,314 en porcentaje da un error del 2.31% que es un valor tolerable para la proyección de sueldos del personal militar para el año 2017.

Si bien se tomó como referencia a un señor General de Brigada para los ejemplos de proyección del 2016 y 2017 se puede observar que existe una diferencia entre valores calculados, esto es debido a que en los grados de generales varió el numérico porque algunos coroneles fueron ascendidos, por lo tanto los valores no van a ser iguales.

4.1.3.6 Proyección para el año 2018

Para realizar la proyección de sueldos del personal militar correspondiente al año 2018 se debe añadir a los datos iniciales los datos obtenidos en la proyección de sueldos del 2016 más la proyección de sueldos del 2017 debido a que con los datos del sueldo para el año 2017 se completaría la información para realizar la proyección de mencionado año, es necesario también señalar que el modelo de proyección es el mismo utilizado para la proyección de los años 2016 y 2017, la única diferencia son los datos puesto que en los datos de origen se aumentaron los datos de los años 2016 y 2017, en los datos de proyección se actualizaron los grados y tiempo en el grado de cada una de las personas que intervienen en esta proyección, a continuación se detalla el análisis de este proceso.

Seguidamente se puede ver en la (figura 70) el modelo utilizado para la proyección de salarios del año 2018.

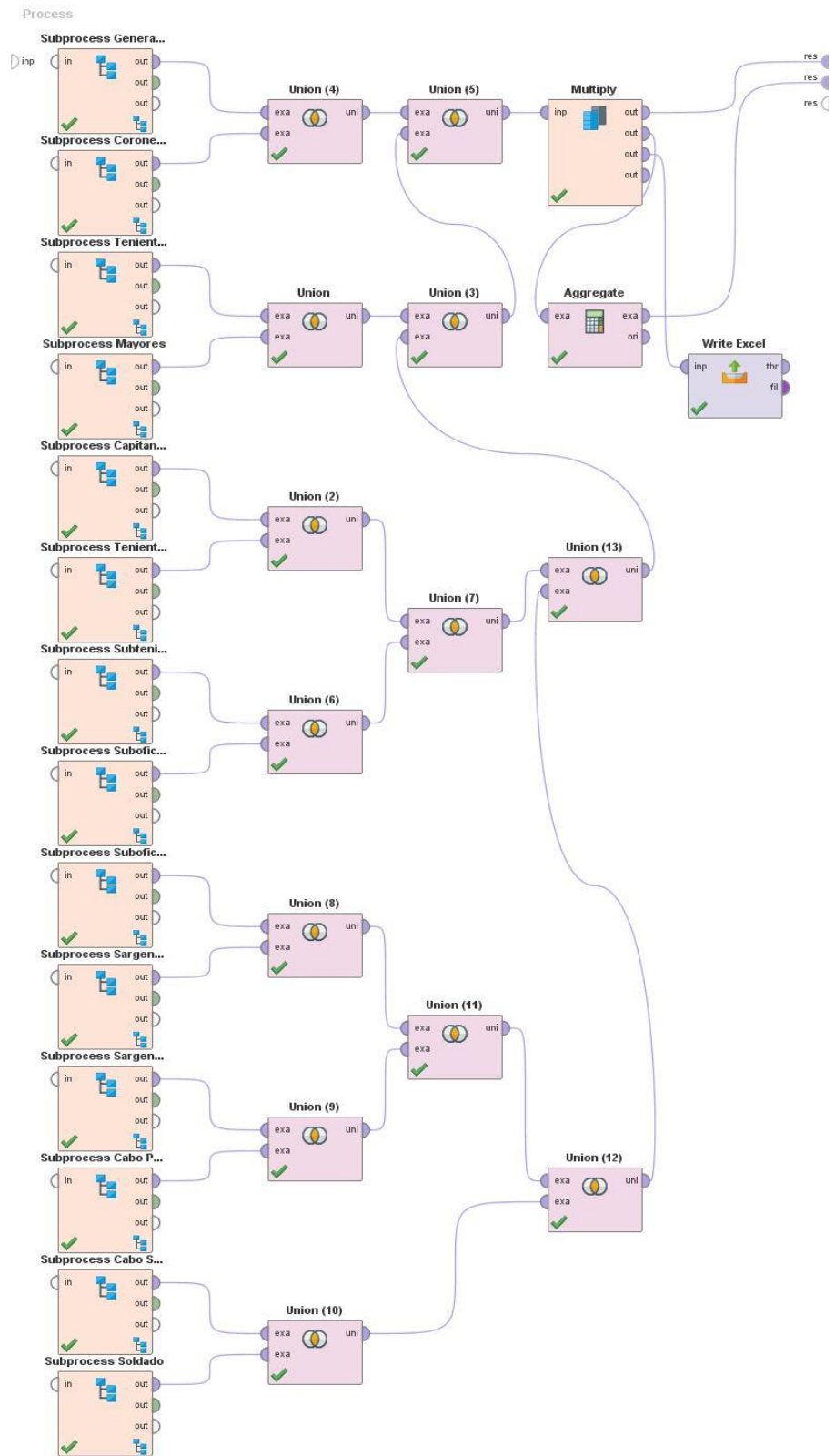


Figura 70: Modelo de proyección neural net de sueldo completo

2018

Row No.	MEM_CEDULA	prediction(S...	SEX_CODIGO	CODGRADO...	NOMGRADO...	TIEMPOGRA...	TIEMPOSER...	EDAD	GRA
1	101061679	57088.734	M	3	GRAB	1	31	64	COR
2	101789089	57088.734	M	3	GRAB	1	34	56	COR
3	400646642	57088.734	M	3	GRAB	1	34	56	COR
4	400667424	58250.310	M	3	GRAB	2	34	55	GEN
5	400675146	55851.205	M	3	GRAB	0	33	55	COR
6	400694220	55851.205	M	3	GRAB	0	33	55	COR
7	400702064	55851.205	M	3	GRAB	0	33	55	COR
8	501146336	57088.734	M	3	GRAB	1	34	57	COR
9	501304216	55851.205	M	3	GRAB	0	33	55	COR
10	501311559	58250.310	M	3	GRAB	2	34	54	GEN
11	601050305	57088.734	M	3	GRAB	1	34	56	COR
12	601540842	58250.310	M	3	GRAB	2	34	55	GEN
13	601569783	57088.734	M	3	GRAB	1	34	58	COR
14	601876972	55851.205	M	3	GRAB	0	33	54	COR
15	1001143336	57088.734	M	3	GRAB	1	34	56	COR
16	1001302288	57088.734	M	3	GRAB	1	34	56	COR

Figura 71: Resultados de proyección neural net de sueldo completo 2018

Para analizar los datos del modelo de redes neuronales para todos los grados militares durante el año 2018 se toma como referencia el sueldo de un general de brigada de 0 a 1 año en el grado es de 54498 USD anual utilizando nuestro modelo propuesto se observa que el sueldo del mismo general proyectado al 2016 es de 55851.205 USD obteniendo una diferencia de 1353,205 en porcentaje da un error del 2.42% que es un valor tolerable para la proyección de sueldos del personal militar para el año 2018.

Si bien se tomó como referencia a un señor General de Brigada para los ejemplos de proyección del 2016, 2017 y 2018 se puede observar que existe una diferencia entre valores calculados, esto es debido a que en los grados de generales varió el numérico porque algunos coroneles fueron ascendidos, por lo tanto los valores no van a ser iguales.

CAPÍTULO 5

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

- Se realizó la depuración de la base de datos remuneraciones a fin de contar con información realmente fiable que ayudó en la investigación; mediante el uso de una de las varias herramientas que existe para minería de datos como es RapidMiner con sentencias de consultas y filtros se logró determinar las variables independientes que permitieron realizar proyecciones de los salarios con un error mínimo.
- Mediante el uso de técnicas de minería de datos en la base de datos remuneraciones; se realizó un estudio profundo de los diferentes modelos generados hasta llegar a determinar que los algoritmos de regresión lineal y redes neuronales son los que más se aproximan a los salarios reales del personal militar.
- Posteriormente se comparó en forma individual los salarios proyectados con las dos técnicas de minería de datos como son regresión lineal y redes neuronales; llegándose a determinar que las redes neuronales son las que proyectan las remuneraciones con un error que va entre un 0.5% y un 2%; por lo que este modelo es el que se propone para futuras proyecciones; ya que permitirá solicitar presupuestos a tiempo y de esta forma ayudará en la toma de decisiones.
- La base de datos depurada se exportó a Microsoft Excel lo que permitió utilizar la información de una forma más sencilla; mediante el uso de la herramienta RapidMiner se pudo desarrollar el modelo de proyección más óptimo para obtener los salarios proyectados de los años 2016 al 2018.
- Se generó una serie de gráficos estadísticos como por ejemplo de dispersión lo que permitió visualizar proyecciones de salarios según los atributos; entre ellas se analizó las provincias que más presupuesto necesitan para cubrir los salarios, así como el presupuesto del personal militar por estado civil entre otros.

- Una ventaja importante en nuestra propuesta es el modelo de predicción del sueldo básico unificado para el personal militar, se puede aplicar para simular un número indeterminado de años en adelante a partir de un año específico.

5.2 Recomendaciones

- Antes de aplicar minería de datos debe haber realizado una investigación a profundidad sobre el tipo de datos en el que desea realizar el trabajo; de esta manera se podrá encontrar un modelo que permita realizar las proyecciones con un error que vaya del 0.5% al 2%.
- En este tipo de investigación se deberá utilizar una base de datos que permita comprobar la predicción generada por una regresión lineal o red neuronal.
- Los atributos seleccionados en la investigación deben permitir llegar a los objetivos establecidos y deben ser eliminados los que no son necesarios.
- Se recomienda por tanto utilizar minería de datos para realizar la proyección de todos los ítems presupuestarios que a lo largo de un periodo entrarán en el cálculo de nómina para saber el monto del gasto que se va a necesitar en uno o varios años determinados.
- Utilizar la información obtenida por medio de la minería de datos para la toma de decisiones a nivel de mandos medios y altos en el Ejército Ecuatoriano.
- Utilizar el estudio de esta investigación como base para futuros análisis como por ejemplo la proyección de sueldos de los Servidores Públicos que tienen una forma de cálculo diferente al personal militar.

REFERENCIAS BIBLIOGRÁFICAS

(s.f.).

Aluja, T. (13 de 2 de 2013). La minería de datos entre la estadística y la inteligencia artificial. Obtenido de <http://upcommons.upc.edu/bitstream/handle/2099/4162/article.pdf?sequence=4&isAllowed=y>

Amores, B., & Cruz, D. (7 de Abril de 2012). Aplicación de data minig para establecer patrones de comportamiento. Obtenido de <http://bibdigital.epn.edu.ec/handle/15000/1041>

AMORES, B., & DIEGO, C. (MARZO de 2013). TESIS EPN MINERIA CLIMA QUITO. Obtenido de <http://bibdigital.epn.edu.ec/handle/15000/1041>

Anónimo. (2014). Data warehouse vs data marts . Obtenido de <http://bi-insider.com/posts/data-warehouse-edw-vs-datamarts/>

Anónimo. (2014). Imágenes de gráficos de un data warehouse en minería de datos. Obtenido de https://www.google.com.ec/search?q=graficos+de+un+data+warehouse+en+mineria+de+datos&biw=1366&bih=623&source=lnms&tbn=isch&sa=X&ved=0ahUKEwiRgqiAoKjOAhVPfiYKHfxBC4IQ_AUIBigB&dpr=1#imgrc=wVM3glLBYmbHdM%3A

C.G.E.E. (10 de 5 de 2015). Manual de remuneraciones. Quito, Pichincha, Ecuador.

Calderón, V. (18 de 7 de 2014). Implantación de un proyecto de Knowledge . Obtenido de <http://hdl.handle.net/10609/442>

- Carranza, S., Naranjo, M., & Rosley, A. (12 de 11 de 2014). Modelo basado en las técnicas de minería de datos. Obtenido de <http://dspace.esPOCH.edu.ec/handle/123456789/3545#sthash.linPSZEv.dpuf>
- Castillo, J., & Palomino, I. (2013). Implementación de un Datamart como una solución de Inteligencia de Negocios. *Revista de Información de Informática y Sistemas*, 53-63.
- Constante, G., & Karla, P. (5 de 11 de 2014). Diseño del sistema de remuneraciones empresa coheco. Obtenido de <http://dspace.ups.edu.ec/handle/123456789/7435>
- Dandretta, G. (Julio de 2013). Implementando técnicas de data mining en un servidor web. Obtenido de <http://repositorio.ub.edu.ar:8080/xmlui/handle/123456789/369>
- Gutierrez, E. (10 de 01 de 2014). **DESCUBRIMIENTO DE CONOCIMIENTOS EN LA BASE DE DATOS ACADÉMICA DE LA UNIVERSIDAD AUTÓNOMA DE MANIZALES APLICANDO REDES NEURONALES.** Obtenido de [http://repositorio.autonoma.edu.co/jspui/bitstream/11182/345/1/MSG-JGutierrez-Informe\(V2\).pdf](http://repositorio.autonoma.edu.co/jspui/bitstream/11182/345/1/MSG-JGutierrez-Informe(V2).pdf)
- Gutierrez, J. (5 de 11 de 2012). Descubrimiento de conocimiento en bases de datos. Obtenido de [http://repositorio.autonoma.edu.co/jspui/bitstream/11182/345/1/MSG-JGutierrez-Informe\(V2\).pdf](http://repositorio.autonoma.edu.co/jspui/bitstream/11182/345/1/MSG-JGutierrez-Informe(V2).pdf)
- Hernández Orallo, J. (2014). Minería de datos 2. El proceso KDD. Obtenido de <http://users.dsic.upv.es/~jorallo/master/dm2.pdf>

- Luzón, V. (18 de 06 de 2014). Implantación de un proyecto de Knowledge Center con una herramienta comercial Synera. Obtenido de <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/442/1/27899tfc.pdf>
- Mamani, A. (5 de 01 de 2014). Inteligencia Artificial en Medicina . Obtenido de <http://www.revistasbolivianas.org.bo/pdf/rits/n1/n1a08.pdf>
- MICROSOFT. (07 de 05 de 2014). DEFINICIÓN DE MINERÍA DE DATOS SEGÚN MICROSOFT . Obtenido de <https://msdn.microsoft.com/es-es/library/ms174949.aspx>
- Moine, J. (Abril de 2013). Metodologías para el descubrimiento de conocimiento en base de datos: un estudio comparativo. Obtenido de http://sedici.unlp.edu.ar/bitstream/handle/10915/29582/Documento_completo.pdf?sequence=1
- Molina, K. (3 de 2012). APLICACIÓN DE LAS TÉCNICAS DEL DATA MINING EN LA WEB. Obtenido de <http://repositorio.puce.edu.ec/handle/22000/1104>
- OOCITES. (30 de 06 de 2014). DEFINICIÓN DE MINERÍA DE DATOS SEGÚN OOCITES . Obtenido de http://www.oocities.org/es/mineria.datos/definicion_tecnicas_mineria_datos.pdf
- Raveane, W., & López , D. (29 de Octubre de 2014). Redes neuronales. Obtenido de <https://www.youtube.com/watch?v=yN3bPRHLd5s>
- SINNEXUS. (25 de 05 de 2013). DEFINICIÓN DE MINERÍA DE DATOS SEGÚN SINNEXUS . Obtenido de http://www.sinnexus.com/business_intelligence/datamining.aspx

Velasco López , J. E. (2013). Uso de técnicas de web mining. Obtenido de

[http://eprints.sim.ucm.es/25812/1/Trabajo%20Fin%20Master%20Jorge%20Velasco%20\(1\).pdf](http://eprints.sim.ucm.es/25812/1/Trabajo%20Fin%20Master%20Jorge%20Velasco%20(1).pdf)