VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSTGRADOS

MAESTRÍA EN SISTEMAS 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

DESARROLLO DE UN MODELO PREDICTIVO PARA DETERMINAR EL IMPACTO ECONÓMICO EN EL PAGO DEL SEGURO DE SALUD DEL ISSFA

AUTORES: ANASI SUNTASIG, KARINA ISABEL

CARRAZCO CÓNDOR, ANA MARÍA

DIRECTOR: MSC. DÍAZ ZUÑIGA, MAGI PAÚL

SANGOLQUÍ

2017

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA TECNOLÓGICA

CENTRO DE POSTGRADOS

MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, "DESARROLLO DE UN MODELO PREDICTIVO PARA DETERMINAR EL IMPACTO ECONÓMICO EN EL PAGO DEL SEGURO DE SALUD DEL ISSFA" realizado por las señoritas Ing. KARINA ISABEL ANASI SUNTASIG y la Ing. ANA MARÍA CARRAZCO CÓNDOR, 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 las Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar a las señoritas: KARINA ANASI SUNTASIG y ANA MARÍA CARRAZCO para lo sustenten públicamente:

Sangolquí, 20 de Noviembre del 2017

MSC. DIAZ ZUÑIGA MAGI PAÚL

DIRECTOR

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA TECNOLÓGICA

CENTRO DE POSTGRADOS

MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

AUTORÍA DE RESPONSABILIDAD

Nosotras, KARINA ISABEL ANASI SUNTASIG, con cédula de identidad N° 1721881330 y ANA MARÍA CARRAZCO CÓNDOR, con cédula de identidad N° 1716241086, declaramos que este trabajo de titulación "DESARROLLO DE UN MODELO PREDICTIVO PARA DETERMINAR EL IMPACTO ECONÓMICO EN EL PAGO DEL SEGURO DE SALUD DEL ISSFA" 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 responsables del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 20 de Noviembre del 2017

Karina Isabel Anasi Suntasig

C.C. 1721881330

Ana María Carrazco Cóndor

1716241086

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA TECNOLÓGICA

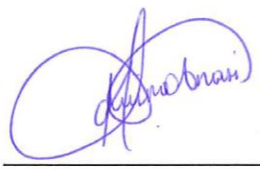
CENTRO DE POSTGRADOS

MAESTRÍA EN SISTEMAS DE GESTIÓN DE LA INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

AUTORIZACIÓN

Nosotras, KARINA ISABEL ANASI SUNTASIG y ANA MARÍA CARRAZCO CÓNDOR, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca virtual de la institución el presente trabajo de titulación "DESARROLLO DE UN MODELO PREDICTIVO PARA DETERMINAR EL IMPACTO ECONÓMICO EN EL PAGO DEL SEGURO DE SALUD DEL ISSFA" cuyo contenido, ideas y criterios son de nuestra autoría y responsabilidad.

Sangolquí, 20 de Noviembre del 2017



Karina Isabel Anasi Suntasig

C.C. 1721881330

Ana María Carrazco Cóndor

C.C. 1716241086

AGRADECIMIENTOS

A Dios, mi familia y mis amigos por ser parte de mi vida y estar conmigo en las victorias y dificultades.

Ana María

AGRADECIMIENTOS

Por un escalón más de la vida, agradezco a Dios por cada oportunidad que nos brinda día a día para poder cumplir nuestras metas y sueños, y que sin él no podríamos estar aquí.

A mi familia en especial a mi mamá, papá y mis hermanos Andrés y Carito de quienes he recibido su apoyo y amor incondicional, y de quienes he adquirido los valores que me forjan como persona y profesional.

A la ESPE y sus prestigiosos profesores, quienes han compartido sus conocimientos y experiencias para nuestro crecimiento profesional.

Finalmente, a mi compañera Any, quien ha sido un soporte para mí en los momentos más difíciles que he tenido que afrontar, y de quien he aprendido su fortaleza y ganas para seguir adelante a pesar de las circunstancias, admiro mucho su calidad profesional y personal.

Kari.

DEDICATORIA

A mi padre que me mira desde el cielo.

Ana María

DEDICATORIA

El esfuerzo de este objetivo se lo dedico a mi Papá Jorge Anasi que ahora ya no se encuentra junto a mí, pero que desde el cielo espero se sienta orgulloso y feliz por cada paso que doy; sé que ahora es un ángel y que aunque no lo vea, siempre estará a mi lado.

A la persona que Dios puso en mi camino para ser mi compañero de vida Rahul.

Kari.

ÍNDICE DE CONTENIDOS

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

RESUMEN

La Dirección del Seguro de Salud del ISSFA gestiona los procesos para la prestación del Servicios de Salud, uno de ellos corresponde a la Facturación médica, mediante el proceso de “Pertinencia y Liquidación de Servicios de Salud” se realizan los pagos a los prestadores de salud. En este contexto, se almacenan grandes volúmenes información que requieren análisis a fin de ofrecer valor agregado al ISSFA. Esta investigación presenta el desarrollo de un modelo predictivo para determinar el impacto económico en el pago del seguro de salud del ISSFA en enfermedades músculo - esqueléticas y conocer que factores son influyentes para que un afiliado sea propenso a padecerlas. Con la metodología de minería de datos CRISP, se identifican: los objetivos del negocio y de minería de datos, evaluación inicial de herramientas y técnicas; en la segunda fase se realiza la exploración de los datos y verificación de calidad; en la fase de preparación de los datos se selecciona el conjunto inicial de datos, criterios de exclusión; en la fase de construcción se realiza la integración de datos para definir el conjunto final. Con la herramienta RapidMiner se construyen los modelos: regresión lineal, máquina de vector de soporte y redes neuronales para la predicción del valor a pagar, mientras que para determinar la combinación de factores de una enfermedad se utiliza al árbol de decisión; se evalúan los modelos en función de los la tasa de error, porcentaje de precisión entre otros y finalmente, realiza el análisis de resultados contrastándolo con el presupuesto asignado por servicio de salud, resultando ser la técnica de Regresión lineal la más óptima.

Palabras clave:

- **MINERÍA DE DATOS**
- **MODELO PREDICTIVO**
- **REGRESIÓN LINEAL**
- **ÁRBOLES DE DECISIÓN**

ABSTRACT

The “Dirección del Seguro de Salud del ISSFA” manage the process of Health Services, one of them is Medical Billing. Through the process of “Relevance and liquidation of health services” the payments are made to Hospitals. In this context, there is a large volume of information stored that needs analytics for the purpose of give value added to ISSFA. This research shows the development of one predictive model to determine the financial impact in health services’s payments in health insurance of ISSFA in skeletal muscle diseases and identify which are the influential factors for an affiliate can get. With the methodology of data mining CRISP can identify: business and data mining objectives, initial evaluation of tools and data mining techniques; in second phase it shows the data exploration and quality verification; in data preparation data it selects the initial data set, exclusion criteria; in construction phase the data integration is made to define the final dataset. With the RapidMiner Tool the models: linear regression, support vector machine and neural net are built, for the prediction of value of payment, while to find the factors combination of one disease the decision tree is built; these models are evaluated and taking the error rate, accuracy percentage and others. Finally, the analysis of results is made contrasting with the allocated budget for healthy service, resulting the linear regressión as the more optimal.

Keywords:

- **DATA MINING**
- **PREDICTIVE MODEL**
- **LINEAR REGRESSION**
- **DECISSION TREE**

CAPÍTULO 1: GENERALIDADES

1.1. Introducción

En el pago por servicios de salud existe gran cantidad de información almacenada, que no es utilizada en forma oportuna y confiable. Frente a esta falencia, la tendencia es inclinarse por el uso de técnicas de minería de datos, que son procesos complejos para la obtención de conocimiento de grandes volúmenes de datos. La minería de datos ofrece modelos ampliamente utilizados en el campo académico, médico, financiero, especialmente en la prevención de enfermedades, predicción de pagos, detección de fraudes, entre otros.

El Ecuador cuenta con tres entidades públicas encargadas de la Seguridad Social de la ciudadanía. Entre las que se encuentra el Instituto de Seguridad Social de las Fuerzas Armadas ISSFA[1]. Una de sus áreas agregadoras de valor, es la Dirección del Seguro de Salud DSS[2], cuya misión es gestionar los recursos necesarios para la cobertura en servicios de salud al conglomerado militar en servicio activo, pasivo y sus dependientes. (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

Uno de los recursos del presupuesto es destinado al pago mensual de los servicios de salud facturados por hospitales, clínicas, policlínicos conocidos como prestadores. El departamento correspondiente dentro la DSS debe gestionar la asignación presupuestaria mensual y anual que generalmente requieren constantes modificaciones o reasignaciones presupuestarias.

Con este antecedente, se plantea la generación de modelos predictivos que permitan procesar la información existente de pagos por servicios facturados y proporcionar a los Directivos del ISSFA el conocimiento necesario para la toma de decisiones oportunas, frente a asignación presupuestaria. Adicionalmente, obtener conocimiento para la elaboración

[1] Instituto de Seguridad Social de las Fuerzas Armadas
[2] Dirección del Seguro de Salud

de planes de promoción y prevención de enfermedades que afectan a la población militar en actos de servicio, con el fin de reducir su índice de ocurrencia y paralelamente su pago.

El presente documento describe el proceso de desarrollo de un modelo efectivo para predecir el pago por servicios de salud de ISSFA en determinadas enfermedades. Con esta finalidad se aplican algoritmos de minería de datos, cuyos resultados se orientan a la selección el modelo de mayor precisión y se identifican las variables más influyentes que permitirán predecir el gasto realizado por el ISSFA para las enfermedades de tipo musculo –esqueléticas.

1.2. Motivación

Debido a la gran importancia que tiene la salud en la población, el tema de predicción de enfermedades ha sido abordado a nivel mundial. Algunos autores (Khemphila & Veera, 2010), (Jabbar, Deekshatulu, & Chndra, 2014) y (Kunwar, Khushboo, Sabitha, & Bansal, 2016) hacen uso de técnicas de minería de datos para la predicción de enfermedades de tipo cardiacas, renales así como de cáncer de mama que son las que más afectan a la población. En cuanto al ámbito de enfermedades músculo esqueléticas, ciertas investigaciones (Paliyawan, Nukoolkit, & Mongkolnam, 2014) se orientan a la obtención de datos acerca de posturas de los trabajadores y posteriormente realizan el análisis mediante algoritmos de clasificación para predecir enfermedades referentes a la columna.

Recientemente en el Ecuador, se realizaron estudios (Tapia, Pérez, & Pérez, 2016) que abordan la problemática de enfermedades ocupacionales de tipo muscular, donde se construye una herramienta para el análisis de variables que influyen en las dolencias trabajadores de industrias florícolas.

Considerando que no se ha encontrado ningún artículo orientado a la generación de modelos predictivos para la industria militar, y que según ciertas afirmaciones realizadas por (Tapia, Pérez, & Pérez, 2016), en el Ecuador no existen herramientas o técnicas capaces de realizar validación

del cálculo de padecimientos, por lo que se debe contar con la asesoría de un profesional médico ocupacional, se desarrolla un modelo que utilizará técnicas de minería de datos para la generación un modelo predictivo efectivo para determinar el impacto económico en los pagos de servicios de salud en función de variables más influyentes de los afiliados.

De esta manera, los resultados del presente proyecto serán de ayuda a la Dirección de Salud del ISSFA para elaborar medidas preventivas en relación al pago por servicios de salud.

1.3. Planteamiento del problema

El personal militar de las Fuerzas Armadas Ecuatorianas tiene como misión fundamental la defensa de la soberanía y la integridad territorial; la protección interna y el mantenimiento del orden público del país (Comando Conjunto de las Fuerzas Armadas, 2017) , para lo cual reciben preparación en actividades físicas, educativas, militares, entre otras. Estas actividades pueden ocasionar la adquisición de enfermedades propias de su trabajo, por lo que el Estado garantiza su derecho a la Seguridad Social. Según el Artículo 38 del código del Trabajo del Ecuador, el empleador debe hacerse responsable de las enfermedades que adquiere su personal durante las tareas que conlleva en la ejecución de su trabajo, por lo tanto una de estas responsabilidades es proporcionar el acceso al servicio de salud según la normativa vigente. (Ministerio de Trabajo y Empleo, 2015)

Es de importancia reconocer que para el cumplimiento de sus funciones, el personal militar realiza diferentes actividades que requieren un mayor esfuerzo físico o de alto riesgo, razón por la cual tiende a adquirir enfermedades de tipo músculo- esqueléticas a corto, mediano y largo plazo. Estos padecimientos se presentan con frecuencia generando pagos elevados del ISSFA a las diferentes Unidades de Salud a nivel nacional.

Para que el ISSFA realice el pago, la Dirección del Seguro de Salud lleva a cabo el procedimiento de “Pertinencia y Liquidación de Servicios de Salud”, que consiste en realizar una auditoría médica y financiera de la

facturación ingresada al sistema por prestadores, contemplando datos de diagnósticos, tratamientos, costos, fechas, servicios, entre otros.

En este contexto, la DSS[3] almacena grandes volúmenes de información de afiliados en temas de salud, sin embargo no existen modelos predictivos que permitan obtener conocimiento escondido en la situación económica actual dificultando la toma de decisiones para afrontar posibles escenarios o ejecutar planes preventivos. Al no existir un análisis de tipo preventivo de las lesiones que sufre una persona, ésta agrava su situación con el pasar del tiempo, dando como resultado enfermedades cada vez más difíciles de curar y por tal motivo más costosas.

Considerando que no se encontró ningún artículo orientado a la generación de modelos predictivos en el contexto de la industria militar, y que según algunas afirmaciones (Tapia, Pérez, & Pérez, 2016) que indican que en el Ecuador no existen herramientas o técnicas capaces de realizar la predicción del pago por enfermedades, se propone realizar la construcción de un modelo orientado a este objetivo.

En este propósito se plantea determinar el impacto económico en los pagos de servicios de salud específicos en función de variables más influyentes de los afiliados.

1.4. Formulación del problema a resolver

El problema a resolver abarca las siguientes preguntas de investigación:

- ¿Qué combinación de factores influyen en la adquisición de una enfermedad músculo esquelético en una persona?
- ¿Cuál es el modelo más preciso para predecir el pago por servicios de salud en enfermedades músculo esqueléticos?
- ¿Cuáles son las enfermedades en las que se predice mayor pago?

[3] Dirección de Seguro de Salud

1.5. Objetivo General

Desarrollar un modelo predictivo del pago en servicios de salud por enfermedades músculo-esqueléticas de los afiliados del ISSFA, utilizando técnicas de minería de datos.

1.6. Objetivos Específicos

- Realizar el levantamiento de requerimientos del negocio desde una perspectiva no técnica.
- Realizar la exploración y descripción de las características de la población de afiliados de acuerdo al pago en salud producido por enfermedades músculo - esqueléticas.
- Aplicar el proceso de extracción, limpieza e integración de datos.
- Construir un modelo para predecir el pago por servicios de salud utilizando técnicas de minería de datos.
- Evaluar el modelo predictivo en función de su tasa de error.
- Analizar los resultados obtenidos y su impacto en pago del seguro de salud del ISSFA.

1.7. Selección de la metodología

Para la selección de la metodología se decide evaluar dos opciones aceptadas nivel científico como son: CRISP-DM[4] y SEMMA[5], de acuerdo a los siguientes criterios:

- **Comprensión de los objetivos del negocio**: Fases o actividades de la metodología que permitan entender claramente el problema.
- **Metodología estructurada:** La metodología se subdivide en fases, las cuales tienen relación entre sí. Este punto permite ejecutar el proyecto de minería de datos sistemáticamente.
- **Especificación de las tareas:** Define claramente y a detalle las tareas que se deben desarrollar.

[4] Cross Industry Standard Process for Data Mining
[5] Sample, Explore, Modify, Model, Assess

- **Independencia de herramientas**: La metodología no es limitante en la selección de las herramientas para la construcción del modelo.
- **Aceptación en otros proyectos:** La metodología a utilizar debe ser aceptada ampliamente a nivel mundial, lo que demuestra su grado de madurez y buenos resultados en otros proyectos.
- **Información de soporte:** La información de ayuda sobre la metodología se encuentra disponible en libros, revistas, artículos, bibliotecas virtuales y sitios web oficiales.
- **Tiempo de desarrollo:** Las actividades que componen las fases de la metodología deben ejecutarse en un tiempo no mayor a seis meses.

La **Tabla** 1, muestra los resultados obtenidos de la evaluación de las metodologías propuestas.

Tabla 1
Comparativa de metodologías de Minería de Datos

Criterios	CRISP-DM	SEMMA
Comprensión de los objetivos del negocio	Si	No
Especificación de las tareas	Si	No, las tareas son a criterio del equipo
Independencia de herramientas	Si	No, la metodología está ligada a herramientas SAS
Metodología estructurada	Si	Si
Aceptación en otros proyectos	Si	Si
Información de soporte	Si	No
Tiempo de desarrollo	Si	Si

Fuente: (Moine & Haedo, 2016), (Tangient LLC, 2017)

Según los resultados obtenidos en la Tabla 1, la metodología que mejor se ajusta al objetivo del proyecto es CRISP –DM por las siguientes razones:

La metodología CRISP [6] según Chapman (Chapman, y otros, 2016) contiene un conjunto de procesos jerárquicos que se componen de fases, tareas genéricas, específicas e instancia de procesos.

De acuerdo a investigaciones (Gómez, Hernandez, & Martinez, 2016), CRISP tiene un conjunto de fases estructurales, que permite comprender más ampliamente los objetivos del negocio, a diferencia de otras metodologías.

La utilización de CRISP – DM, no se limita al uso de herramientas específicas o proveedores como el caso de SEMMA, puesto se califica como una metodología neutral de uso libre y gratuito. (Tangient LLC, 2017)

Según encuestas realizadas por el sitio web KDnuggets [7] (Piatetsky, 2017), la metodología CRISP es la primera en ser utilizada en proyectos de minería de datos en el ámbito de la salud (Huang, McGregor, & James, 2014) , educativo (Sailesh & Lu, 2016), marketing, planificación estratégica (Yun & Chen, 2014) , entre otras por su alto grado de madurez lo cual se muestra en la **Figura 1**.

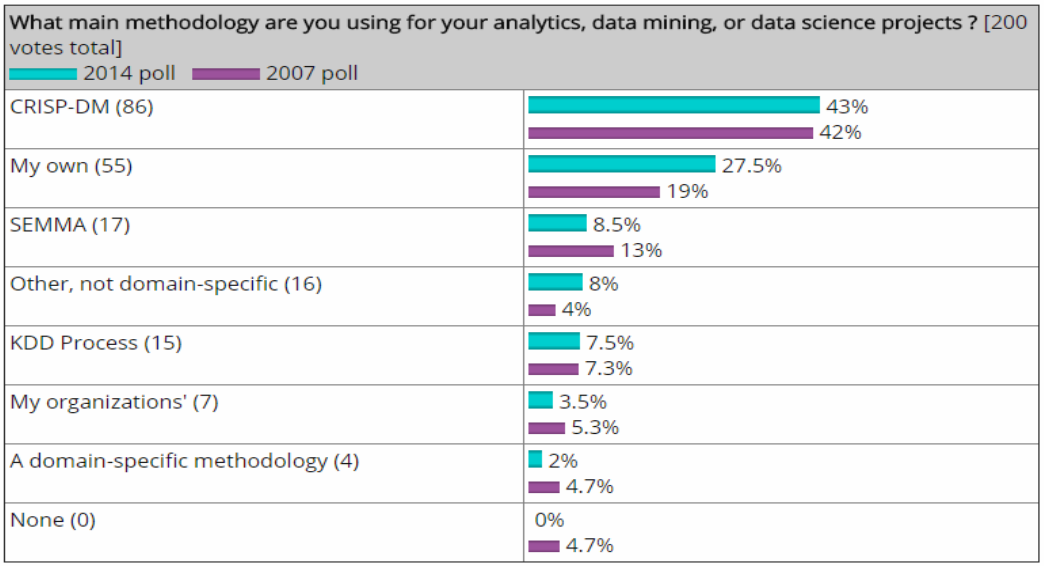


Figura 1 Utilización metodología en proyectos de minería de datos
Fuente: (Piatetsky, 2017)

[6] Cross Industry Standard Process for Data Mining
[7] Sitio Web que abarca temas de Analítica del negocio, Big Data, Minería de datos, aprendizaje de máquina, entre otros

Este factor ha contribuido a que proveedores como IBM la tomen como base para el desarrollo de sus herramientas, por ejemplo SPSS Modeler es una plataforma que permite realizar la creación de modelos predictivos tomando como base la metodología CRISP – DM (IBM, 2017), la cual brinda al usuario las directrices de las tareas que debe desarrollar en cada fase del proyecto (IBM, 2012).

CAPÍTULO 2: MARCO TEÓRICO

2.1. Estado del arte

A continuación se presenta el estado del arte tomando como fuente los siguientes artículos científicos:

Bontempi realiza un estudio de aplicación de métodos de selección de características en el campo de la Bioinformática (Bontempi, 2005), posteriormente otros autores (Venkataraman, Kubicki, Westin, & Golland, 2010) aplican el mismo procedimiento, como resultado de ambos estudios se evidencia que este tipo de técnicas ayuda a comprender los datos con más facilidad, reduce el tiempo de entrenamiento y mejora la precisión de la predicción.

En 2011, Jacob, Ramani y Nancy (Jacob & Ramani, Data Mining in Clinical Data Sets: A Review, 2012) se enfocan en la categorización de cáncer de mama mediante un estudio comparativo de 20 técnicas de clasificación, obteniendo como resultado un alto porcentaje de precisión cuando se utilizan los algoritmos RandomTree y C4.5 de Quinlan. Estos mismos resultados se obtienen en 2012 cuando Shomona G.J. R.Geetha Ramani, realizan la ejecución de varios algoritmos a fin de predecir el aparecimiento de tumores en mamas, enfermedades del corazón, afecciones en la piel el incluso ortopédicas. (Shomona & Ramani, 2012)

Sathyadevi realiza una investigación con el objetivo de mejorar el proceso de aprendizaje para determinar si un paciente tiene o no enfermedades hepáticas, mediante la utilización del algoritmo CART (Sathyadevi, 2011), una vez adquirido el conjunto de datos, se los introdujo en la herramienta Weka como datos de entrenamiento, se aplicó el algoritmo y se obtuvieron reglas de clasificación.

En 2012, Jacob y Ramani (Jacob & Ramani, Data Mining in Clinical Data Sets: A Review, 2012) realizan el análisis de 84 artículos científicos tomados de la base de datos bibliográfica MEDLINE (Medline, 2017), de la revisión se

concluye que las técnicas de minería de datos brindan apoyo a profesionales de la salud y permiten obtener conocimiento escondido partiendo de datos clínicos, los autores pudieron notar que las técnicas de predicción predominantes han sido Clasificadores Bayesianos, redes neuronales y máquinas de vector de soporte.

En este mismo año, ciertos autores (Shapiro, 2015) plantean un framework que se enfoca en incluir la detección de outliers antes de la aplicación de los algoritmos de clasificación para disminuir la complejidad computacional y remover el esparcimiento de los datos que no tengan relación de los pacientes. Con los resultados representados en las reglas de asociación se pueden identificar los síntomas de los pacientes. (Shomona & Ramani, 2012)

En estudios posteriores, Lukáčová, Babič, Paraličová y JánParalič, (Lukáčová, Babič, & Paraličová, 2015) utilizan métodos de aprendizaje de máquina en el contexto de Hepatitis. Los autores se apoyan en el método de particionamiento recursivo CHAID (Chi-squared Automatic Interaction Detector) para generar un árbol ampliamente visual y fácil de entender, en una fase posterior plantea el cálculo del costo beneficio para detectar la enfermedad en etapas tempranas.

En 2016, Tapia, Pérez y Pérez abordan la problemática de enfermedades ocupacionales en su estudio “Automating the analysis and evaluation of occupational risk factors accumulated in the flowerindustry” (Tapia, Pérez, & Pérez, 2016), donde los autores construyen una herramienta que analiza variables que influyen en las enfermedades ocupacionales de trabajadores de una florícola.

En 2016 Diz,Marreiros y Freitas (Diz, Marreiros, & Freitas, 2016) realizan un estudio comparativo a fin de encontrar las mejores técnicas de predicción en el ámbito de lesiones benignas y malignas de cáncer de mama. Para su investigación los autores utilizan los métodos de máquina de vector de soporte, árboles de decisión y Bayes, obteniendo altos índices de precisión

al utilizar máquina de vector de soporte, y Selvas Aleatorias (RandomForest). En este mismo año, utilizan una combinación de técnicas de minería de datos a fin de obtener un modelo altamente preciso en la predicción de cáncer de próstata (Tirumala & Narayanan, 2016), en la primera fase utilizan enfoques de selección de atributos para obtener las características más relevantes del conjunto de datos, con esta información pre procesada realizan una comparación al aplicar redes neuronales artificiales y Bayes. Como resultado experimental se evidencia que los modelos son más precisos cuando se utilizan técnicas de selección de atributos previamente.

Otro estudio en donde también se evidencian combinación de técnicas es (Roslina & Noraziah, 2010) en donde se enfocan en la predicción de diagnósticos de Hepatitis mediante el uso métodos Wrapper [8] para la selección de atributos relevantes y posteriormente se aplica el método de Vector de Soporte de Máquinas, en este artículo también se demuestra que el uso de métodos de selección de características incrementa la precisión del modelo, coincidiendo con otros autores (Durgadevi & Kalpana, 2017)

En el estudio reciente, “Chronic kidney disease analysis using data mining classification techniques” (Kunwar, Khushboo, Sabitha, & Bansal, 2016), los autores recaban información de 35 publicaciones científicas orientadas la detección de enfermedades haciendo uso de técnicas de minería de datos, producto de esta categorización, presentan datos estadísticos en donde se identifica el número de artículos por técnica de minería de datos.

Como se muestra en la Figura 2, la función más utilizada en los artículos estudiados, es la Clasificación, en segundo lugar se coloca el Agrupamiento o Clustering. Los autores mencionan que la técnica de Clasificación que más se utiliza son las redes neuronales ubicándose en segundo lugar los árboles de decisión que son modelos predictivos fáciles de entender.

[8] Métodos de envoltura

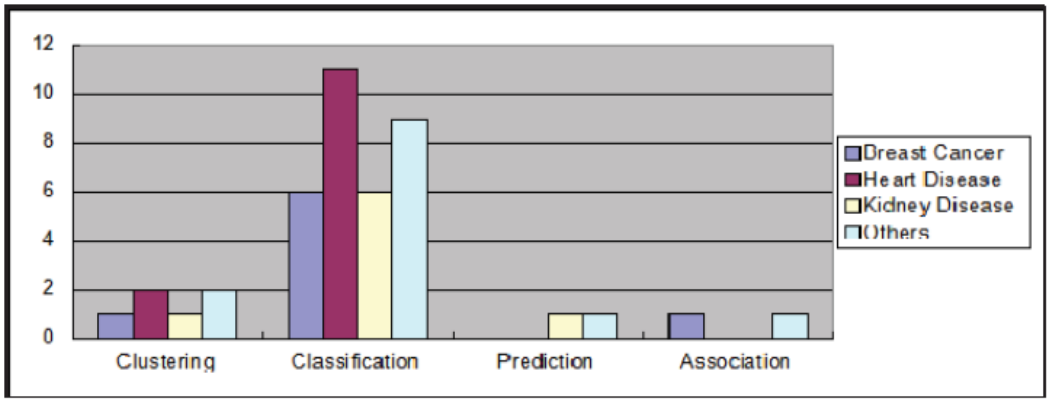


Figura 2 Técnicas de minería de datos para predicción de enfermedades

Fuente: (Kunwar, Khushboo, Sabitha, & Bansal, 2016)

No se ha encontrado ningún estudio que permita realizar la construcción de un modelo orientado al impacto económico en enfermedades típicas en la industria militar.

2.2. Técnicas de minería de datos

La Minería de datos, es un proceso que permite realizar el análisis de grandes cantidades de datos para la obtención de conocimiento escondido (Kunwar, Khushboo, Sabitha, & Bansal, 2016), el producto de este procesamiento son directrices, reglas y predicciones que dan al usuario el soporte adecuado para la toma de decisiones.

En el campo de la medicina, los modelos de predicción y detección temprana de enfermedades son ampliamente utilizados con el fin de que el profesional médico pueda enviar a tiempo tratamientos que combatan los síntomas de la enfermedad en las primeras etapas.

El modelamiento predictivo es un proceso de análisis que permite la generación de reglas de clasificación o predicción basada en un conjunto de datos de prueba y posteriormente realiza la predicción en un sub conjunto de datos con una respuesta conocida, estos son tomados para la validación. Este proceso de entrenamiento y validación se conoce como entrenamiento supervisado. Una vez que el modelo aprende y se valida su exactitud, se puede pronosticar resultados futuros tomando como fuente nuevos conjuntos de datos. (Gartner, 2017)

2.2.1. Algoritmos de Clasificación y Regresión

La Clasificación, es una técnica de minería de datos en la que cada objeto es catalogado en una clase, los miembros tienen una o varias características en común, a diferencia del Clustering en donde el agrupamiento no tiene una clase definida. Se basan en el descubrimiento de patrones y aplicación de algoritmos en grandes volúmenes de datos para predecir comportamientos futuros. (Gonzalez, 2014)

En los algoritmos de clasificación se toma como base la información que se extrae de los datos de entrenamiento ya clasificados. Son utilizados para la resolución de problemas en los cuales se conocen las clases para la categorización (ECC, IAA, & UTPL, 2008). El modelo de aprendizaje se compone de dos fases (Rodriguez, 2017):

- La fase de entrenamiento en donde se produce el aprendizaje y validación del modelo (generación de la regla de clasificación).
- La fase de validación en donde aplica el modelo para su clasificación con otras muestras.

Los algoritmos que pertenecen a esta categoría son árboles de decisión, redes neuronales y clasificación bayesiana.

2.2.1.1. Árboles de Decisión

Es un algoritmo de clasificación ampliamente utilizado. Sirve de herramienta para la toma de decisiones, pues presenta un conjunto de alternativas o condiciones secuenciales que se visualizan en forma de ramales y aportan conocimiento. Las entradas para el procesamiento de un árbol pueden ser objetos o atributos, y cada salida es la consecuencia de dichas entradas. En un árbol de decisión cada nodo realiza una decisión binaria, característica que lo distingue del resto.

Para la construcción de árboles de decisión es necesario tener dos conjuntos de datos: de entrenamiento y de prueba, los primeros corresponden a la parte más grande del conjunto de datos recolectados y se

utilizan para la construcción del modelo, mientras que los segundos sirven para realizar la validación de la exactitud del árbol. (Bouza & Santiago, 2014)

Algunas variaciones de árboles de decisión son:

- **CART:** Es un algoritmo que permite mejorar el proceso de aprendizaje en los arboles de regresión y clasificación. Utilizan procedimientos de autovalidación en la búsqueda de patrones en las bases de datos con el fin de evitar el sobreajuste. Este algoritmo permite al usuario ver fácilmente la jerarquía de las variables y son fácilmente aplicables a nuevos datos. (Sathyadevi, 2011)
- **CHAID:** Detector de Interacción Automático Chi - Cuadrado o Chi-squared Automatic Interaction Detector. Es un método de particionamiento recursivo, que genera resultados ampliamente visuales y fáciles de interpretar. El algoritmo construye árboles no binarios, posteriormente utiliza la prueba de Chi - cuadrado para determinar la mejor partición en cada paso. (Lukáčová, Babič, & Paraličová, 2015)
- **Selvas aleatorias**[9]: Es un clasificador de alto rendimiento, utilizado cuando los tamaños de las clases difieren considerablemente o cuando los datos son fuertemente desequilibrados. (Diz, Marreiros, & Freitas, 2016) Es una combinación de árboles predictores, cada árbol es dependiente de los valores de un vector aleatorio. Se caracteriza por ser un algoritmo altamente preciso, puede ser utilizado con grandes conjuntos de datos y evidenciar los atributos importantes para la clasificación. (Breiman & Cutler, 2016)

2.2.1.2. Naive Bayes

Es un método de clasificación de tipo probabilístico, en donde se da un valor a la clase y se considera a todos los atributos independientes. Naive Bayes se caracteriza por un enfoque simple y facilidad de representación. (Diz, Marreiros, & Freitas, 2016)

[9] Random Forest

2.2.1.3. Redes Neuronales

Las redes neuronales son modelos predictivos de procesamiento automático y aprendizaje, establecen relaciones lineales o no lineales entre las salidas y entradas. Imitan el mismo principio de funcionamiento de las redes neuronales de seres vivos, en donde existe un conjunto de neuronas interconectadas entre sí, para procesar las entradas de información y obtener una salida. En base a la experiencia, las neuronas se auto organizan, realizan procesos de aprendizaje adaptativo, y generalizan situaciones que no fueron consideradas en su entrenamiento, con el objetivo de mejorar los resultados obtenidos. (Villada, Muñoz, & Henao, 2015)

Para el aprendizaje se deben conocer claramente ciertos fundamentos (Avellano, 2016) como:

- Paradigma de aprendizaje: información disponible
- Regla de aprendizaje: principios que gobiernan el aprendizaje.
- Algoritmo de aprendizaje: La forma en la que se lleva a cabo el ajuste de pesos

Las redes neuronales se componen de

- Nodos de entrada: Depende de la información disponible, correspondiente a los atributos seleccionados para el modelo.
- Capas ocultas: capa formada por los pesos asignados a los atributos y sus conexiones.
- Capa de salida: es el valor esperado como resultado.

La Figura 3, muestra una red neuronal con propagación hacia adelante:

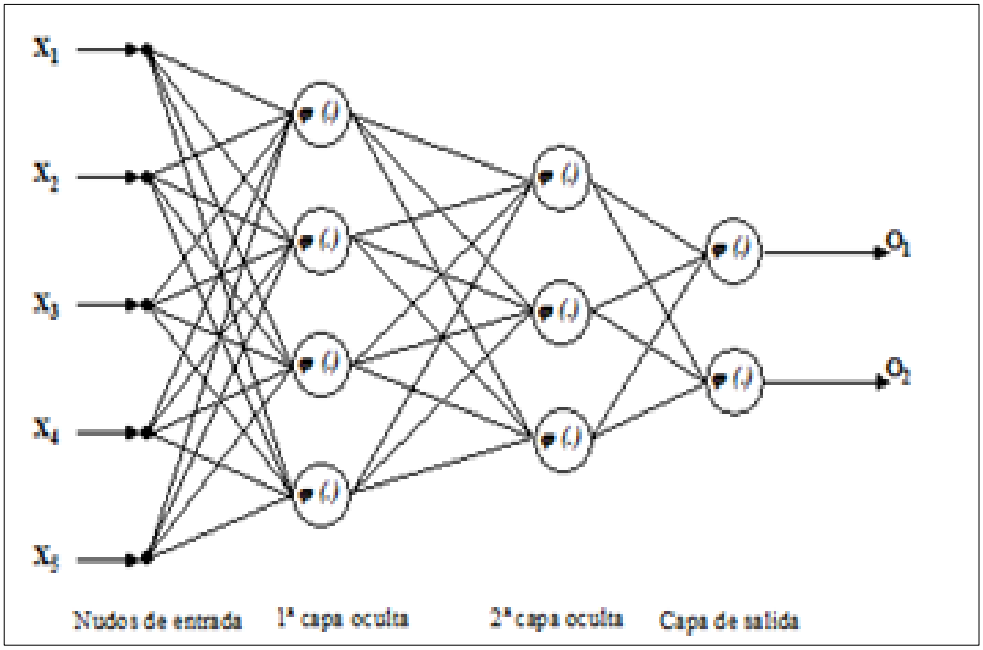


Figura 3 Red Neuronal con propagación hacia adelante
Fuente: (Avellano, 2016)

2.2.1.4. Regresión Lineal

Es un proceso utilizado para encontrar la ecuación lineal a la que mejor se ajusta un conjunto de datos y de esta forma explicar su comportamiento en función de una variable dependiente (y) y una o más independientes(x) (Frost, 2013). A continuación se puede observar la ecuación resultante del modelo, en donde Y es la variable dependiente, X la variable independiente y B los coeficientes a ser encontrados en el modelo:

$$\boldsymbol{y = \beta 0 + \beta 1 x 0 + \beta 2 x 1 \ldots \ldots \ldots + \beta n x n}$$

Para determinar la relación entre las variables se utiliza el coeficiente de correlación y determinación.

Según autores (Mora & Rodriguez , 2001), para la selección del modelo es necesario considerar criterios estadísticos como:

- P valor: Indica el nivel de redundancia entre las variables y funciona en relación a un valor crítico, que usualmente es 0,05. Un p valor mayor al valor crítico advierte que la información

proporcionada por una variable independiente puede estar también presente en las demás y debe ser descartada.

- Tolerancia: Indica la presencia de una variable como combinación lineal de las restantes, por lo que si este valor se aproxima a 0, la variable no se debe considerar para el modelo.

2.2.1.5. SVM

El algoritmo SVM[10], llamado también Máquina de Vector de Soporte, es uno de los métodos de predicción más utilizado. Son modelos de aprendizaje supervisado que realizan el análisis de datos destinados a clasificación y análisis de regresión. (Diz, Marreiros, & Freitas, 2016)

SVM[11] un clasificador binario, permite ubicar a un conjunto de puntos en una de dos clases, para lo cual determina el hiperplano en el espacio de características que separa los puntos y maximiza la distancia de cada clase, separándolas lo mayor posible. (Pérez, Guevara, Silva, Ramos, & Loureiro, 2014). Un atributo corresponde a la variable predictora y las características son aquellas que se utilizan para la construcción del hiperplano, para ello debe realizar un proceso previo de selección de características, los puntos más cercanos al hiperplano forman un vector conocido como vector de soporte. Para separar los planos, es necesario utilizar funciones Kernel o núcleo, las cuales sirven para trasladar el espacio de entradas X a un nuevo espacio de características de mayor dimensionalidad. (Cristianini & Shawe-Taylor , 2017)

Las funciones núcleo más conocidas son:

- Lineal: Es la función más utilizada por su simplicidad, está definida de la forma k(x,y)=(x*y). La Figura 4, muestra la división de ejemplos en el hiperplano con este tipo de núcleo.

[10] Support Vector Machine o Máquina de Vector de soporte
[11] Support Vector Machine o Máquina de Vector de soporte

- Polinomial: Un núcleo polinomial utiliza una función de la forma k(x,y)=(x*y+1)^n, en donde n es el grado del polinomio. Suele utilizarse en datos normalizados.
- Neural: El núcleo neural se compone de una red neuronal de dos capas tal que tanh(a x*y+b) en donde a es el coeficiente de x y b es el intercepto.

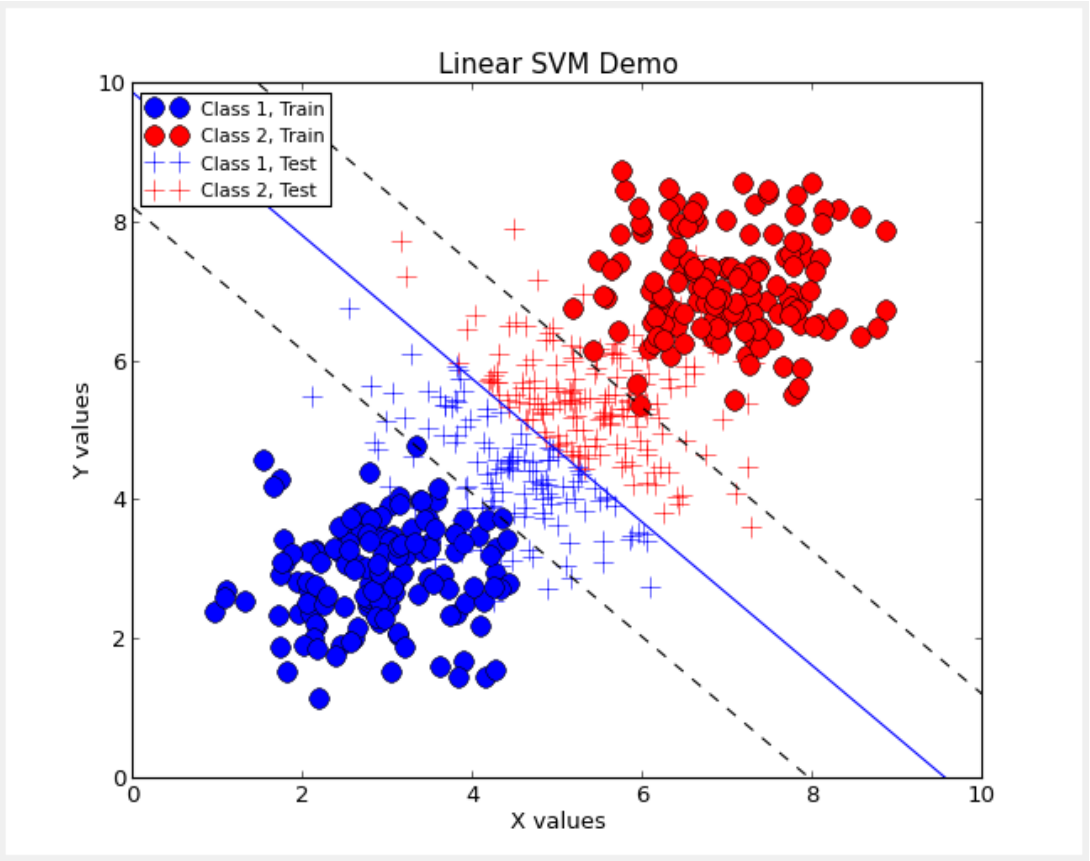


Figura 4 Clasificación SVM con función lineal.
Fuente: (Random Forests, 2014)

En este tipo de algoritmo, el entrenamiento y la clasificación son eficientes, tienen un alto índice de precisión en problemas típicos y son muy robustos para generalizaciones. Sin embargo en la fase de validación del modelo es necesario considerar el valor de sesgo – varianza conocido también como bias ya que según (Stuart, Bienenstock, & Doursat, 1992) un valor de sesgo bajo proviene de un modelo complejo que podría contener ruido haciendo que sus predicciones sea menos precisas y llevando a la

generalización, del otro lado un valor alto que proviene de un modelo simple puede ser indicativo de sobreajuste.

2.2.2. Algoritmos de Clusterig

Los algoritmos de Clustering se caracterizan por agrupar o segmentar una colección de datos en subconjuntos, grupos o clúster, de manera que cada elemento dentro del grupo tenga características similares pero diferentes características en relación a los otros grupos. Estos métodos dividen un grupo de N elementos en un número K de grupos de objetos similares y la elección de sus respectivos centroides, los mismos que son el centro geométrico del clúster. La distribución de los objetos a cada clúster se calcula dependiendo de la distancia entre objetos y su centroide. (Rousseeuw & Kaufman, 1990)

Los pasos básicos de este método son (García & Gómez, 2014):

- Seleccionar los K centroides iniciales: dividir el conjunto de datos en grupos o clústeres y elección del centroide para cada uno.
- Clasificación de cada dato al clúster más cercano de acuerdo al cálculo de las distancia respecto al centroide, este cálculo se puede hacer en base a una distancia euclidiana, de Manhattan, de Chebyshev o de Minkowski
- Recalculo de centroides: para cada clúster se vuelve a calcular los centroides.
- Verificación de convergencia: Finalizar si no hay reasignación de los puntos o si la reasignación satisface a una regla de parada (satisfacer la convergencia, hasta que los centroides no se muevan) de lo contrario ir al paso 2.

La diferencia entre los tipos de métodos particionados radica en los procedimientos para obtener los centroides iniciales y las reglas para la reasignación de datos, así como el cálculo de sus distancias. Entre los métodos particionados más comunes se tiene:

- K-means: Selecciona el centroide en base al valor medio
- K-medianas: Selecciona el centroide inicial y sustituye el valor de promedios por el vector de medianas del grupo de datos y utiliza la distancia Manhattan como la medida de similitud.
- K-modes: Utiliza modas para realizar el agrupamiento

2.3. Métodos de selección de características

La selección de características implica una de las principales tareas de minería de datos. Los modelos de selección de características permiten obtener los atributos más importantes de grandes conjuntos de datos, de esta manera es posible reducir tiempo y esfuerzo en el procesamiento de datos en problemas de clasificación o regresión. (Jacob, Ramani, & Nancy, Feature Selection and Classification in Breast Cancer Datasets through Data Mining Algorithm, 2011) En algunos casos ha demostrado ser útil para incrementar la precisión del modelo, sin embargo la reducción del número de características también puede llevar a perder la capacidad de discriminación. (Jain , Duin, & Mao, 2000)

El mejor conjunto de características depende del criterio tomado para la selección, existen ciertos algoritmos aplicados al total de atributos a fin de obtener un subconjunto m, sin embargo en varias investigaciones (Jacob, Ramani, & Nancy, Feature Selection and Classification in Breast Cancer Datasets through Data Mining Algorithm, 2011), (Pudil, Novovicová, & Kittler, 1994), (Tirumala & Narayanan, 2016) se evidencia que el más utilizado es el Método de búsqueda flotante secuencial.

Según un estudio realizado por Pudil, Novovicová y Kittler (Pudil, Novovicová, & Kittler, 1994), los métodos de selección de características conocidos como SFSM[12] por su siglas en inglés, son las técnica de selección más efectivas y pueden ser de dos tipos: hacia adelante SFFS[13] o hacia

[12] Sequential Floating Search Methods
[13] Sequential Forward Floating Selection

atrás SFBS[14]. A continuación se describe brevemente el funcionamiento de cada tipo:

- **SFFS**: El algoritmo comienza con un conjunto vacío de características, en las iteraciones se seleccionan los atributos no utilizados del conjunto de datos de entrada. En cada atributo seleccionado se evalúa el rendimiento de acuerdo a una función de criterio, aquel que tenga mejor rendimiento es añadido a la selección actual. En los siguientes pasos se desea mejorar el conjunto actual ya sea incluyendo o excluyendo alguno de los atributos considerados en pasos anteriores.
- **SFBS**: El algoritmo parte del total de características del conjunto de datos y en cada paso busca mejorar conjunto actual mediante la disminución de atributos.

Una mejora de los métodos de búsqueda flotante secuencial, son los de búsqueda flotante adaptativa conocidos como AFSM[15] por sus siglas en inglés, en donde difiere el número de atributos que se puede incluir o excluir en cada paso del algoritmo. (Pudil, Novovicová, & Kittler, 1994)

2.4. Modelos de detección de elementos atípicos

Son modelos que representan detección de ruido, novedades o elementos que se desvían del comportamiento común de los miembros en una muestra (outlier). Sirven para reducir la complejidad computacional y remover los datos esparcidos (Jacob, Ramani, & Nancy, Feature Selection and Classification in Breast Cancer Datasets through Data Mining Algorithm, 2011). Además son considerados como detalles de información significativos para obtener información importante. (Agrawal, Imielinski, & Swami, 1993) Existen varios métodos para la detección de elementos atípicos, algunos de ellos son:

[14] Sequential Floating Backward Selection
[15] Adaptive Floating Search Methods

- **Diagramas de caja y bigotes:** Son gráficos utilizados para representar la distribución de datos de una variable numérica mediante cuartiles. Como se muestra en la Figura 5, los elementos atípicos se ubican fuera de los límites admisibles del diagrama (Muñoz & Amón, 2013), que se determinan mediante la fórmula:
  Rango Intercuartil (IQR) = Cuartil 3 (Q3) – Cuartil 1 (Q1)
  Límite superior = Q3 + 1.5 IQR
  Límite inferior = Q1 - 1.5 IQR

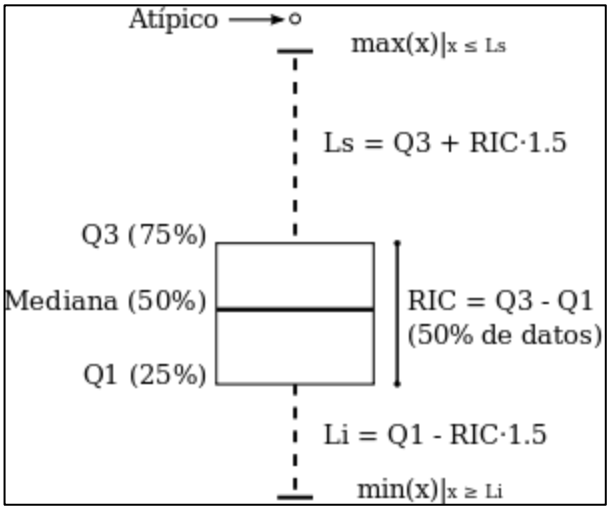


Figura 5 Diagrama de Caja
Fuente (Wikipedia, 2017)

- Factor atípico local: Algoritmo conocido como LOF por sus siglas en inglés Local Outlier Factors, que identifica los elementos aislados en base a la medida de la desviación local de un ejemplo con respecto a sus vecinos. (Breuning, Kriegel, & Sander, 2000) Esta distancia se utiliza para estimar su densidad, se consideran como elementos atípicos, a aquellos ejemplos que presentan baja densidad. Según estudios realizados por (Motaz, Saad, & Nabil, 2009), este algoritmo es altamente eficiente y no implica gran cantidad de costo computacional.
- Factor atípico de clase: Algoritmo conocido como COF por sus siglas en inglés Class Outlier Factor, que determina la existencia de elementos atípicos en función del grado de pertenencia a una

clase aislada otorgado por la instancia de cada ejemplo del conjunto de datos en referencia a tres factores: la desviación de instancias de la misma clase, la probabilidad de la clase entre sus vecinos y la distancia a sus vecinos más cercanos. (Motaz, Saad, & Nabil, 2009)

2.5. Tipos de análisis de información

Existen diferentes tipos de análisis de la información, a continuación se detallan los más utilizados.

2.5.1. Análisis descriptivo

Este tipo de análisis también se conoce como tradicional. El análisis descriptivo permite entender los acontecimientos actuales y pasados, para la obtención de alternativas. Para la empresa consultora Lantares (Lantares, 2016), las posibilidades que oferta el análisis descriptivo no dan el soporte necesario para la toma de decisiones y son poco confiables. (Gartner, 2017)

Este tipo de análisis contesta a la pregunta: ¿Qué ha sucedido o que sucede?

2.5.2. Análisis Predictivo

El análisis predictivo es utilizado para realizar pronósticos mediante la generación de modelos. Este tipo de análisis realiza comparaciones de la influencia de factores en ciertas situaciones y describe las posibles consecuencias, generando valor agregado desde los datos. (Gartner, 2017)

Este análisis responde a las preguntas ¿Qué sucederá? ¿Qué sucederá si?

Da lugar al descubrimiento de un conjunto de escenarios probables en donde los beneficios e impactos no son determinantes para su selección, al contrario del análisis prescriptivo. Según estudios realizados (Lantares, 2016), el análisis predictivo se orienta a la toma de decisiones y existen dos aplicaciones primordiales en este análisis:

- Pronósticos de los fracasos en un proyecto, de esta manera es posible descartar aquellas opciones que no garanticen la consecución de un objetivo, lo que es conocido como un fracaso evitable, y como consecuencia un costo evitable.
- Creación de estrategias para nuevas opciones de negocios y optimización de recursos.

2.5.3. Análisis causal

Según Piffaut (Piffaut, 2015), el análisis causal tiene el objetivo de identificar las causas raíces que originan ciertos problemas o patrones, se caracteriza por la determinación de factores causa – efecto, descartando las asociaciones. Este tipo de análisis trata de eliminar todas las excusas y justificaciones que pueden obstaculizar en el cumplimiento de un objetivo, sino que se orienta a la identificación de razones o causas válidas que expliquen los resultados obtenidos, ya sean buenos o malos.

2.6. Metodología

A continuación se presenta la caracterización de la metodología seleccionada tomando como base el documento Step by step Data Mining Guide escrito por el autor de CRISP. (Chapman, y otros, 2016)

CRISP-DM[16] o Proceso Estándar de Trans-Industria para la Minería de Datos, puede ser entendida desde dos enfoques: desde el modelo o el proceso.

El enfoque del modelo representa una visión general de las fases y las tareas que la componen, mientras que el proceso CRISP representa en forma detallada los pasos para llevar a cabo la tarea (Chapman, y otros, 2016), lo antes mencionado se representa en la Figura 6.

[16] Cross-Industry Standard Process for Data Mining

Así mismo, los niveles superiores las tareas son más generales, mientras que en los inferiores las tareas se detallan para describir las acciones a realizar en un escenario establecido

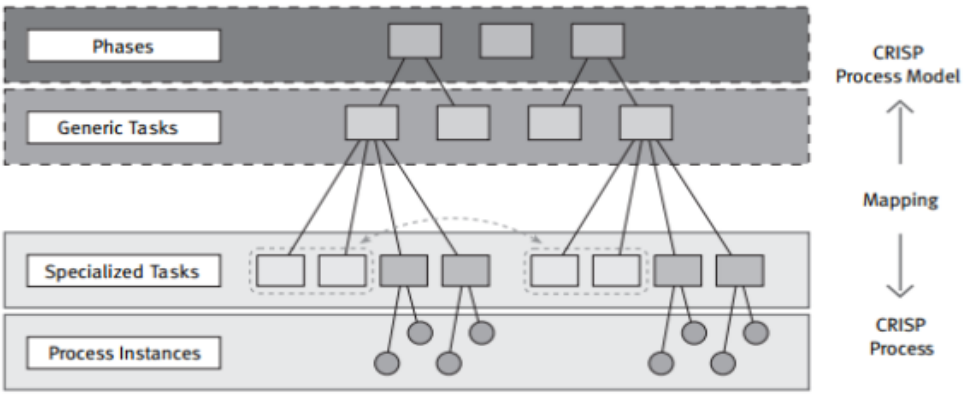


Figura 6 Cuatro niveles de la metodología CRISP-DM
Fuente: (Chapman, y otros, 2016)

En los primeros niveles existe una visión general del ciclo de vida de un proyecto de minería de datos, lo que se denomina fases. En la Figura 7 se pueden observar las 6 fases que componen la metodología y sus relaciones entre sí:

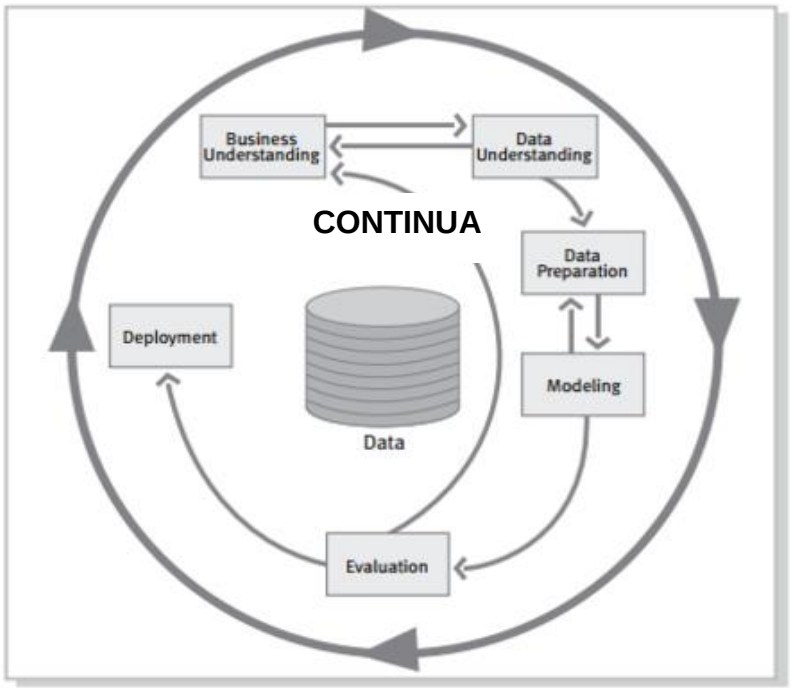


Figura 7 Fases del modelo de referencia CRISP-DM
Fuente: (Chapman, y otros, 2016)

En la metodología el escenario ideal es que las tareas se desarrollen en secuencia, sin embargo según los autores (Chapman, y otros, 2016) muchas tareas se pueden desarrollar en un orden diferente e incluso repetirlas, de acuerdo a las necesidades. Las flechas representan las dependencias que se producen con mayor frecuencia, se considera el proceso cíclico porque una vez que se despliega una solución puede ser que se presenten más preguntas del negocio con mayor enfoque y especificidad.

A continuación de describen las fases que componen la metodología (Chapman, y otros, 2016) :

2.6.1. Entendimiento del negocio

Es la fase inicial de la metodología, el propósito es entender el problema y los objetivos desde una perspectiva del negocio, se redefine el problema en términos de minería de datos. Como salida se tiene un plan que permita la consecución de objetivos.

Las tareas genéricas de la fase son (Chapman, y otros, 2016):

- **Determinar los objetivos de negocio:** Describir el problema en términos del negocio. El papel del analista es sacar a la luz algunos factores que podrían influir en los resultados del proyecto y determinar lo que realmente quiere el cliente. Como salidas se tiene:
  - Fondo o antecedentes, es lo que se conoce de la organización antes de la implementación del proyecto de minería de datos
  - Objetivos del negocio.
- **Evaluar la situación actual**: En esta tarea se detallan más a fondo algunos factores como, recursos, posibles riesgos y suposiciones que se deberán plasmar en el plan de proyecto y serán una entrada necesaria para el entendimiento de los datos. Como salidas se tiene:
  - Inventario de recursos (Humanos, hardware, software y datos)
  - Supuestos y restricciones
  - Terminología en el contexto del negocio y en el ámbito de minería de datos.

- **Establecer objetivos de minería de datos:** Realizar la conversión de los objetivos del negocio y resultados del proyecto en términos técnicos que sean aplicables a la minería de datos. Su salida son los objetivos de minería de datos.
- **Producir un plan de proyecto:** Esta tarea incluye la elaboración de un plan que incluya las tareas, entradas, salidas, dependencias tiempos y recursos que se deben cumplir en el resto de fases del proyecto, el establecimiento de un plan adecuado incrementa la posibilidades de éxito en la consecución de objetivos de minería de datos y por tal los objetivos del negocio. Como salidas se tiene:
  - Plan de proyecto: Es un documento dinámico en donde se revisan las tareas cumplidas al final de cada fase, y puede ser actualizado si existe la necesidad.
- **Evaluación inicial de herramientas y técnicas:** incluye un listado de las herramientas disponibles y las que posiblemente serán utilizadas en el proyecto de minería de datos, junto con un match hacia los requerimientos.

2.6.2. Entendimiento de los datos

En esta fase existe una familiarización con los datos y su calidad, descubriendo algunas directrices que podrían conducir a la formulación de hipótesis sobre ciertas partes de la información (Chapman, y otros, 2016).

Las tareas en esta fase son:

- **Recolectar datos iniciales:** Esta tarea incluye el proceso de extracción o acceso a los datos que serán utilizados para el logro de los objetivos del proyecto. En la tarea se podría incluir la integración de múltiples fuentes de datos o la carga de la data en alguna herramienta específica que brinde las facilidades para su entendimiento. Su salida es la colección inicial de datos.
- **Describir los datos:** Se realiza un perfilamiento de datos, es decir se describe superficialmente los datos recolectados. Su salida es una

descripción de datos, incluye el número de registros, tablas y sus campos, el tipo y formato de datos y cuales datos satisfacen a los requerimientos del proyecto.

- **Explorar los datos:** Es un estudio de perfilado de los datos, en esta tarea se debe identificar cuáles serán los atributos clave y como se relacionan entre sí, realizando análisis estadísticos no complejos para la identificación de patrones. Su salida es la exploración de datos.
- **Verificar la calidad de los datos:** En la tarea se pretende confirmar aspectos como la completitud e integridad de los datos, errores o valores perdidos. Se debe describir cómo y con qué frecuencia se presentan estas novedades. Su salida son las observaciones sobre la calidad de los datos.

2.6.3. Preparación de datos

En esta fase se realizan las actividades necesarias para limpiar y transformar el conjunto de datos inicial hasta un conjunto final, que será el insumo para la fase de modelado (Chapman, y otros, 2016).

Las tareas que se llevan a cabo son:

- **Seleccionar los datos:** En la tarea se escoge la información que sea relevante y de calidad para el cumplimiento de los objetivos de minería de datos. Específicamente se seleccionan tablas, atributos específicos y registros que serán utilizados para la construcción del conjunto de datos. Su salida es:
  - Base de inclusión y exclusión: Lista de datos y los motivos por los que fueron excluidos del proyecto o fueron seleccionados
- **Limpiar los datos:** Esta tarea está destinada al incremento de la calidad de los datos, se enfoca en selección de conjuntos de datos que estén limpios, incluir valores por defecto en el caso de faltantes. En esta tarea se pueden aplicar técnicas más complejas a fin de estimar algunos valores faltantes. Su salida es la definición de criterios para limpieza de datos.

- **Construir los datos:** Es la ejecución de operaciones sobre los datos, algunas pueden ser obtención de nuevos registros, o inclusión de campos transformados o derivados. Sus salidas son:
  - Atributos Derivados: Atributos construidos de los ya existentes
  - Nuevos Registros generados.
- **Integrar los datos:** Se realiza la combinación de datos de múltiples fuentes para la obtención de nuevos atributos o registros. Su salida es:
  - Datos fusionados: Conjunto de datos de diferentes fuentes pero de un mismo objeto o contexto

2.6.4. Modelado

En esta fase se seleccionan y aplican técnicas de modelado. Un problema de minería de datos podría ser resuelto con más de una técnica. (Chapman, y otros, 2016) . Las actividades que se desarrollan son:

- **Seleccionar técnica de Modelado:** En esta tarea se elige una o varias técnicas específicas de minería de datos.
  Sus salidas son:
  - Técnicas de modelado
  - Suposiciones de modelado: En el caso de que se necesite, por ejemplo la distribución de los datos, valores nulos no permitidos, entre otras.
- **Generar Diseño de Pruebas:** Es construir un diseño de pruebas que servirá para realizar la evaluación de la precisión y calidad del modelo que se construirá. Su salida es:
  - Diseño de pruebas: En donde se especifican los métodos a ser utilizados y los datos destinados al entrenamiento, pruebas y evaluación del modelo
- **Construir el Modelo:** Es la utilización de la herramienta seleccionada con el conjunto de datos preparado a fin de obtener uno o varios modelos.

2.6.5. Validación

En esta fase se realiza una evaluación del modelo construido en la fase anterior, lo cual asegura que sea de buena calidad y cumpla con los objetivos del negocio. El objetivo de la fase es decidir el uso que se le dará a los resultados obtenidos de la minería de datos. A continuación de describen las tareas de la fase:

- **Evaluar los resultados:** Se busca evaluar el grado en el cual el modelo permite cumplir con los requerimientos del negocio. Se descubrirán cuáles son los modelos más apropiados relacionando las necesidades y cuáles deben ser descartados, e incluso descubrir directrices para proyectos futuros. Las salidas esperadas son:
  - Evaluación de los resultados de minería de datos con respecto a los criterios de éxito del negocio: Listar los objetivos e indicar el cumplimiento por los modelos desarrollados
  - Modelos Aprobados: Son los modelos que cumplen con los criterios de éxito del negocio

2.6.6. Despliegue

La fase se refiere a incrustar el modelo construido en el proceso de toma de decisiones de una organización, de tal manera que los resultados puedan ser visibles y útiles para el usuario. Dependiendo de la necesidad de la empresa el despliegue puede ser la implementación de un reporte simple a una implementación compleja de minería de datos en diferentes áreas (Chapman, y otros, 2016).

A continuación se describen las tareas genéricas de cada fase:

- **Planear el despliegue:** En la tarea se debe desarrollar una estrategia de despliegue del modelo.
- **Planear Monitoreo y Mantenimiento:** Crear una estrategia de mantenimiento y monitoreo que permita prevenir que a futuro existan resultados inesperados.

- **Producir reporte final:** Es la escritura de un reporte final, el cual puede ser orientado a la documentación de las experiencias adquiridas en el desarrollo del proyecto o una presentación de los resultados obtenidos.
- **Revisar el proyecto:** Es evaluar las tareas ejecutadas en el desarrollo del proyecto, especificando lo que se llevó a cabo de manera satisfactoria y lo que se debería mejorar.

CAPÍTULO 3: DESARROLLO DE LA PROPUESTA

El presente capitulo muestra el desarrollo de las tareas especificadas en la guía a: “CRISP-DM 1.0 Step by step data mining guide” (Chapman, y otros, 2016) descritas en el capítulo 2:

3.1. Fase de Entendimiento del Negocio

3.1.1. Determinación de los objetivos del negocio

3.1.1.1. Fondo del negocio

El Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, es una Institución pública que administra sus propios reglamentos, normas y leyes; brinda prestación de servicios públicos obligatorios para la Seguridad Social del colectivo militar en servicio activo, pasivo, a sus dependientes, aspirantes a oficiales, tropa, conscriptos y pensionistas. Una de las prestaciones principales es el Seguro de Enfermedad y Maternidad que es administrado por la Dirección de Salud (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

El proyecto de minería de datos se desarrolla en la Dirección de Salud, un área agregadora de valor, que se encuentra bajo la Subdirección General como se muestra en la Figura 8.

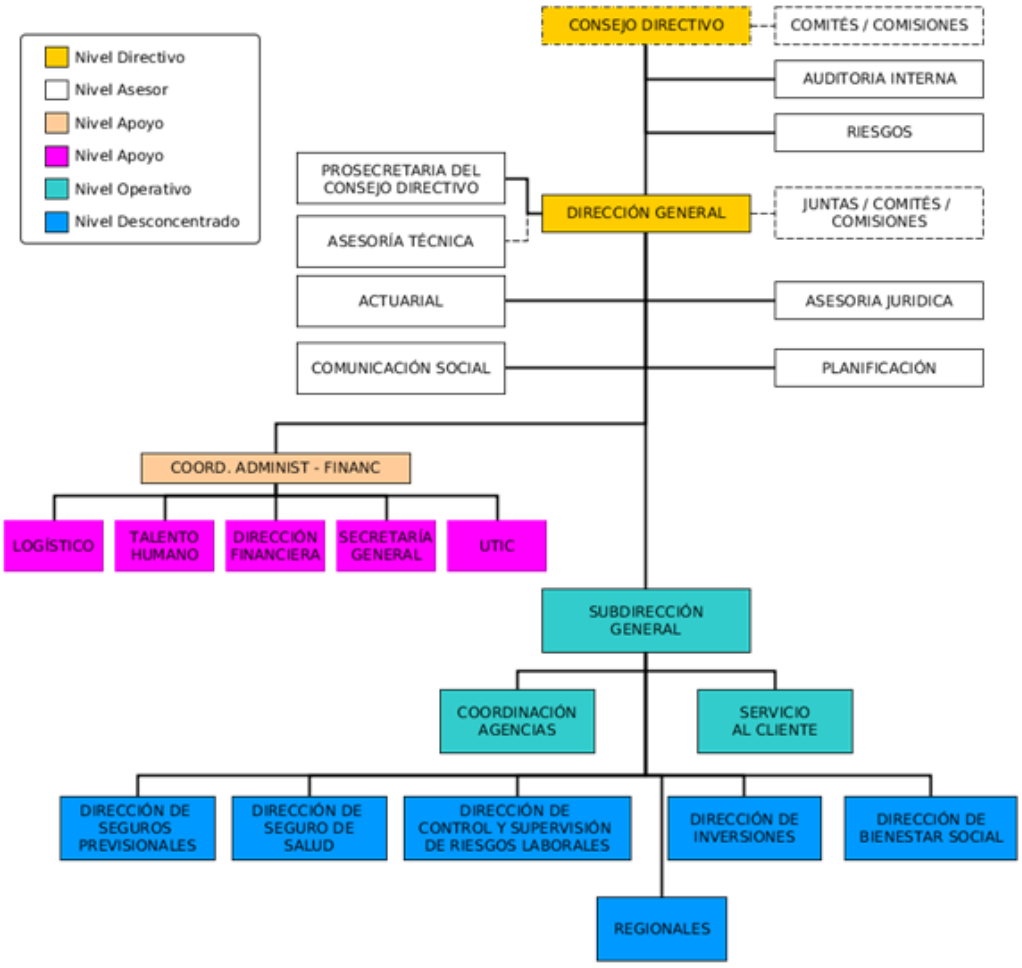


Figura 8 Estructura Organizacional del ISSFA
Fuente: (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

La Dirección del Seguro de Salud es el área encargada de gestionar el proceso de cobertura de servicios de salud brindados a los asegurados del ISSFA, tales como (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017):

- Medicina preventiva
- Asistencia clínica y quirúrgica
- Asistencia obstétrica
- Asistencia odontológica
- Rehabilitación, órtesis y prótesis
- Asistencia farmacológica
- Hospitalización
- Emergencia

- Consulta Externa
- Servicios de Diálisis

Al ser la misión de la Dirección del Seguro de Salud: “Administrar eficientemente el seguro de enfermedades y maternidad, para garantizar la entrega equitativa, oportuna y efectiva de los beneficios a los que tiene derecho el afiliado y su familia, a fin de coadyuvar a la solución de sus necesidades de salud” (Dirección del Seguro de Salud del ISSFA, 2016), tiene definidos **procedimientos** que rigen las actividades y productos a entregarse.

El proyecto de Minería de Datos toma como insumo el procedimiento de “Pertinencia y Liquidación de Servicios de Salud”, cuyo objetivo principal consiste en: “Normar el procedimiento para efectuar la pertinencia documental, médica y financiera de la facturación de las prestaciones de salud en los servicios de: consulta externa, exámenes y procedimientos especiales por consulta externa, emergencia, hospitalización y odontología entregadas a los asegurados con referencia a protocolos y guías de práctica clínica, en base a los criterios de aplicación del Tarifario de prestaciones para el Sistema Nacional de Salud y demás normativa”. (Dirección del Seguro de Salud del ISSFA, 2016)

De esta manera, el punto de partida son las planillas de cobro emitidas por el Prestador de Servicio de Salud, posteriormente realiza la ejecución de la pertinencia documental, médica y financiera hasta la emisión de la orden de gasto para el respectivo pago.

Según (Dirección del Seguro de Salud del ISSFA, 2016), para que el ISSFA a través de la Dirección de Salud realice el pago al Prestador de Salud, se deberá generar una solicitud al finalizar cada mes, la solicitud es correspondiente al tipo de Servicio de Salud y contiene el conjunto de planillas que se han facturado. Una planilla representa las atenciones brindadas a los asegurados por los servicios de salud que los Prestadores de Salud facturan.

Cada planilla contiene información correspondiente a:

- Afiliado
- Fecha de Atención
- Diagnóstico Primario, Diagnóstico Secundario
- Valor por servicio
- Servicio de Salud

Cabe señalar que para que la Dirección de Salud del ISSFA pueda realizar el pago por las atenciones facturadas, existe un presupuesto definido, el mismo que se encuentra asignado a una cuenta correspondiente al tipo de Servicio de Salud, es decir existe una relación directa del tipo de Servicio con la cuenta presupuestaria asignada, a continuación en la Tabla **2** se muestra la relación existente entre los servicios y las cuentas:

Tabla 2
Cuentas y Servicios de Salud

Tipo de Servicio	Cuenta
Atenciones Médicas por Consulta Externa	7.4.1.03.01.05.11 7.4.1.03.01.10.09
Exámenes y Procedimientos (Consulta Externa)	7.4.1.03.01.05.12 7.4.1.03.01.10.01
Emergencia	7.4.1.03.01.05.10 7.4.1.03.01.10.08
Hospitalización y Reposición Gastos Hospitalarios	7.4.1.03.01.05.09 7.4.1.03.01.10.07

Fuente (ISSFA, 2017)

Es importante recalcar, que al inicio del año se asigna un presupuesto para cada servicio de salud, sin embargo se presentan reasignaciones presupuestarias constantes debido a la falta de un modelo que permita realizar una proyección adecuada del valor para cada servicio, a continuación la Tabla 3 muestran los datos correspondientes al valor presupuestado, valor ejecutado, porcentaje de cumplimiento y el número de reasignaciones de los servicios en los años 2015 al 2017.

Tabla 3

Valor presupuestado vs. Valor ejecutado Porcentaje de cumplimiento y el Número de reasignaciones

Año	Tipo de Servicio	Valor Proyectado	Valor Ejecutado	Saldo	Porcentaje de cumplimiento	Número de reasignaciones
2015	Atenciones Médicas por Consulta Externa	$ 3.018.000,00	$ 1.849.840,37	$ 1.168.159,63	61,29%	236
	Emergencia	$ 4.348.938,00	$ 2.975.364,75	$ 1.373.573,25	68,42%	232
	Exámenes y Procedimientos	$ 26.264.000,00	$ 21.462.222,78	$ 4.801.777,22	81,72%	282
	Hospitalización	$ 32.212.177,00	$ 28.726.002,13	$ 3.486.174,87	89,18%	242
2016	Atenciones Médicas por Consulta Externa	$ 1.152.549,00	$ 1.037.073,74	$ 115.475,26	89,98%	155
	Emergencia	$ 2.520.000,00	$ 1.513.488,18	$1.006.511,82	60,06%	109
	Exámenes y Procedimientos	$ 10.700.000,00	$ 10.563.597,46	$ 136.402,54	98,73%	150
	Hospitalización	$ 35.900.000,00	$ 24.444.951,84	$ 11.455.048,16	68,09%	136
2017	Atenciones Médicas por Consulta Externa	$ 1.052.549,00	$ 732.203,74	$ 320.345,26	69,56%	130
	Emergencia	$ 2.520.000,00	$ 1.012.126,33	$ 1.507.873,67	40,16%	105
	Exámenes y Procedimientos	$ 10.800.000,00	$ 6.078.012,83	$ 4.721.987,17	56,28%	91
	Hospitalización	$ 31.776.000,00	$ 14.901.208,68	$ 16.874.791,32	46,89%	107

Fuente (Dirección Financiera y Departamento de presupuesto del ISSFA, 2017)

Se puede apreciar en la Figura 9, que la asignación de presupuestos definidos para los servicio de salud en los años 2015 al 2016 no se ejecutan al 100%.

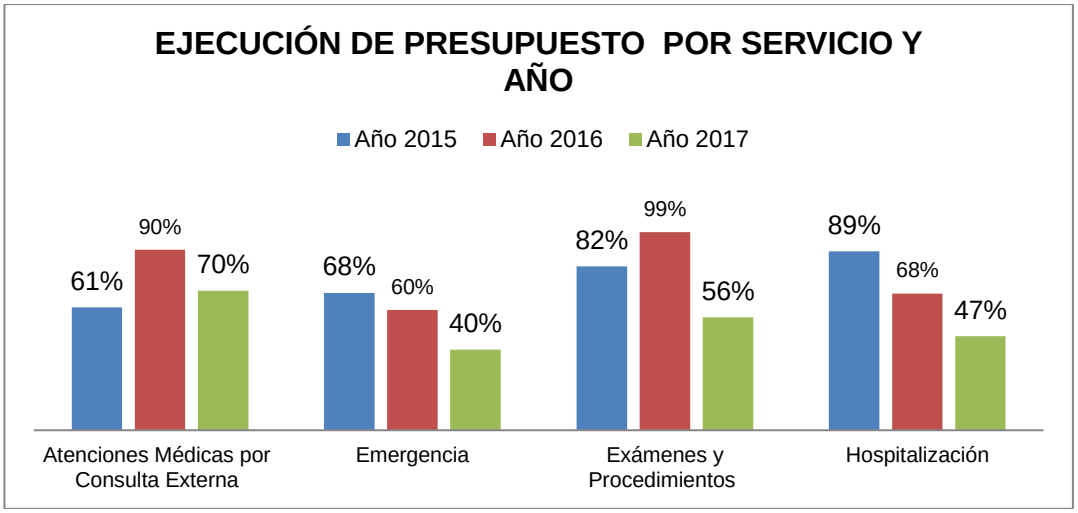


Figura 9 Ejecución de Presupuesto por servicio y año

Este comportamiento da paso a las reasignaciones presupuestarias, las cuales se presentan con frecuencia, en la Figura 10, se aprecia que en el servicio que se dan menos reasignaciones es en el servicio de Emergencia a diferencia de los demás servicios.

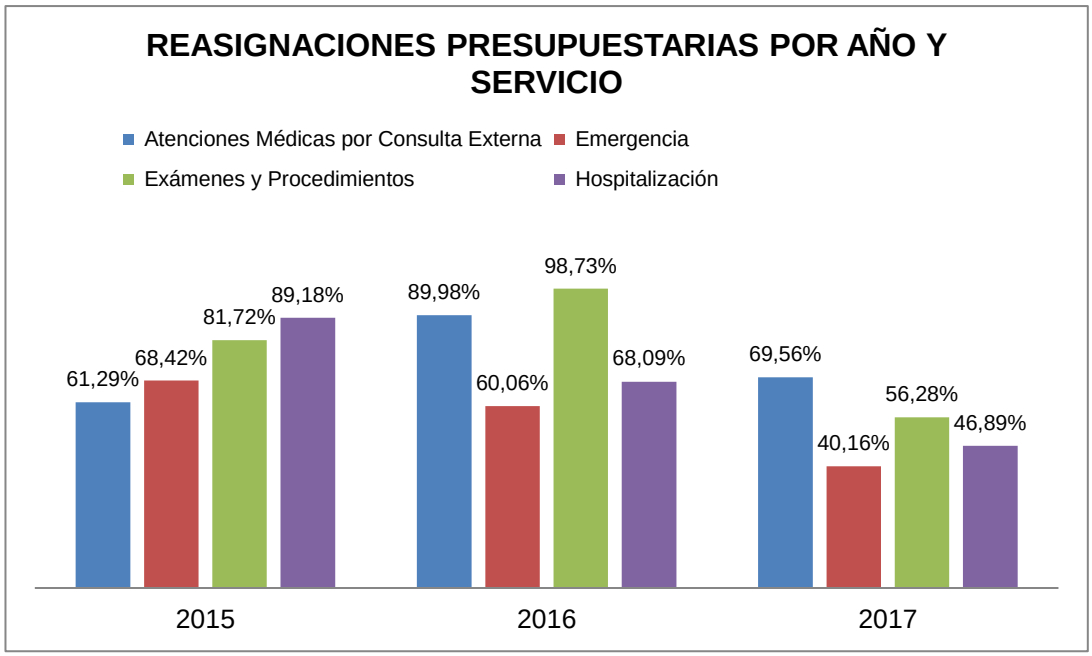
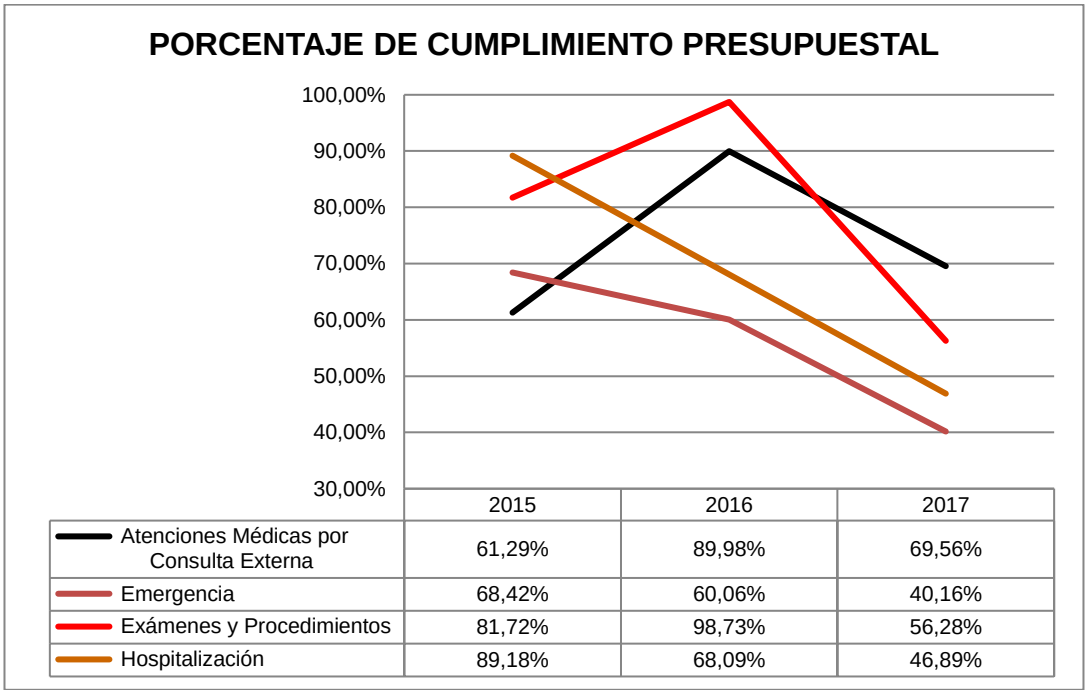


Figura 10 Reasignaciones presupuestarias por año y servicio

La Figura 11 la permite conocer el comportamiento del cumplimiento presupuesta por servicio y año, en los servicios de Hospitalización y Emergencia se muestra que el cumplimiento del presupuesto del 2015 al 2017 es decreciente, este comportamiento se replica a la fecha de corte 30 de Septiembre del 2017, lo que fácilmente indica que el porcentaje de este año no se cumplirá al 100%. Por otro lado en los servicios de Atenciones Médicas por Consulta Externa y en Exámenes y Procedimientos el cumplimiento es incremental del 2015 al 2016, mientras que desde el 2016 al corte de al 30 de septiembre del 2017 el cumplimiento disminuye a solo un 56% siendo la mitad del periodo del año.



	2015	2016	2017
Atenciones Médicas por Consulta Externa	61,29%	89,98%	69,56%
Emergencia	68,42%	60,06%	40,16%
Exámenes y Procedimientos	81,72%	98,73%	56,28%
Hospitalización	89,18%	68,09%	46,89%

Figura 11 Porcentaje de Cumplimiento Presupuestal

Es en este punto, donde el proyecto de minería de datos debe implementar un modelo para predecir el impacto económico del gasto que realiza el ISSFA por las enfermedades de tipo Musculo Esqueléticas, con el fin de reducir el número de reasignaciones presupuestarias.

La Dirección de Salud se encuentra dividida en cuatro departamentos mostrados en la Figura 12:

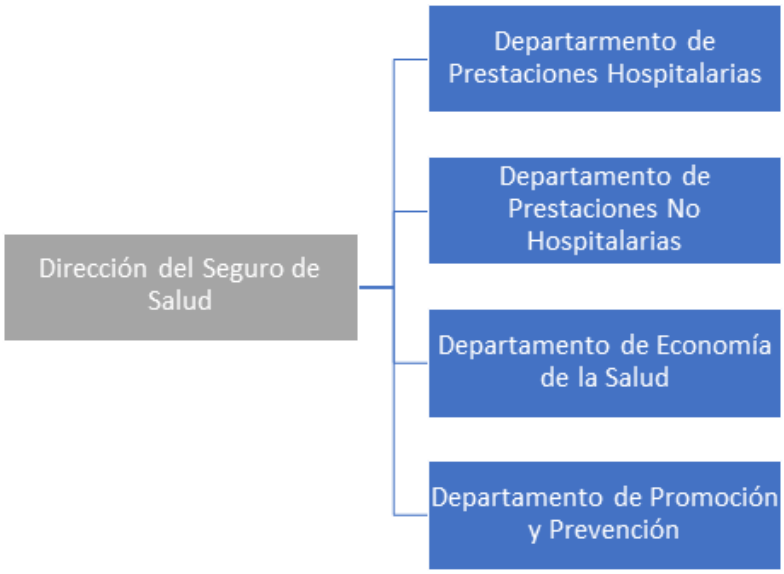


Figura 12 Estructura de la Dirección del Seguro

A continuación, se describe brevemente cada departamento de la Dirección del Seguro de Salud:

Departamento de Prestaciones Hospitalarias: Según (Dirección del Seguro de Salud del ISSFA, 2016) Se encarga de gestionar la cobertura en servicios de Salud tales como Hospitalización, Exámenes y procedimientos, emergencia, atenciones médicas por consulta externa, asistencia odontológica, diálisis y atención pre hospitalaria brindados en Prestadores de salud como policlínicos , clínicas y hospitales.

Departamento de Prestaciones No Hospitalarias: Se responsabiliza de gestionar la cobertura en servicios de Salud tales como son rehabilitación, entrega de órtesis y prótesis, medicina de pacientes crónicos y reposición de gastos por alquiler de concentradores de oxígeno. (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

Departamento de Economía de la Salud: Es el área encargada de coordinar, organizar y programar los procesos y procedimientos que permitan optimizar el presupuesto asignado al Seguro de Enfermedad y Maternidad para la generación de los diferentes productos y servicios de economía de la salud. (Dirección del Seguro de Salud del ISSFA, 2016)

Departamento de Promoción y Prevención: El departamento de Promoción y Prevención de la Salud toma como insumo el estudio descriptivo realizado por el departamento de Economía de Salud, para la generación y ejecución de planes ,proyectos o medidas preventivas que contribuyan a cuidar la salud de los afiliados y sus dependientes. (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

A continuación, se describen los Stakeholders del negocio y sus responsabilidades relacionadas con la ejecución del proyecto:

- **Director del Seguro de Salud:** Es el CEO (ChiefExecutiveOfficer) de la Dirección del Seguro del ISSFA, sus actividades esenciales son (Dirección del Seguro de Salud del ISSFA, 2016):
  - Administrar los recursos disponibles del seguro de enfermedad y maternidad para otorgar prestaciones médicas a los afiliados conforme a la normativa vigente.
  - Gestionar los procesos del seguro de enfermedad y maternidad, para la entrega oportuna y efectiva de los servicios de salud a los afiliados.
  - Dirigir el cumplimiento de los planes, programas y proyectos de la Dirección del Seguro de Salud, a fin de cumplir objetivos institucionales.
  - Suscribir la autorización de los pagos a las Unidades de Salud calificadas, para cubrir las atenciones médicas entregadas a los afiliados.
  - Gestionar la evaluación de los objetivos, metas e indicadores de la Dirección del Seguro de Salud, para establecer acciones correctivas y de mejora continua.

- **Jefe de Prestaciones Hospitalarias:** Es el COO (ChiefOperatingOfficer) de la Dirección de Salud, responsable del “Procedimiento de Pertinencia y liquidación de Servicios de Salud”.

Sus actividades esenciales son (Dirección del Seguro de Salud del ISSFA, 2016):

- Autorizar el pago por servicios de salud a unidades calificadas.
- Legalizar y autorizar prestaciones de servicios de salud a afiliados en condiciones normales y especiales. (Dirección del Seguro de Salud del ISSFA, 2016)
- Realizar el control en órdenes de gastos.

- **Jefe de la Sección de Liquidación y Pagos:** Es la persona encargada de realizar un control sobre la ejecución de pagos por servicios de salud que se brindan a los asegurados del ISSFA.
- **Médico Auditor:** Es el funcionario encargado de la pertinencia médica que consiste en revisar y validar que los servicios y procedimientos médicos fueron entregados a los afiliados de forma oportuna. Algunas de sus actividades esenciales son (Dirección del Seguro de Salud del ISSFA, 2016):
  - Realizar los ajustes necesarios entre el número de prestaciones y la cantidad de procedimientos solicitados por el prestador y los definidos por la pertinencia.
  - Aplicar objeciones o débitos de acuerdo a la evaluación realizada.
- **Jefe de Economía de la Salud:** Es el CFO (ChiefFinancialOfficer) de la Dirección del Seguro de Salud. Sus actividades esenciales son (Dirección del Seguro de Salud del ISSFA, 2016):
  - Coordinar y organizar los procesos de economía de la salud, para el manejo de estimación de presupuestos, políticas de salud y modelos financieros de salud aplicables.
  - Programar y evaluar las acciones para verificar la calidad de modelos económicos de salud entregados.
  - Diseñar y programar los proyectos para el mejoramiento de la gestión de la economía de la salud.
  - Evaluar el cumplimiento de la gestión, presupuesto y proyectos del área y establecer las acciones de mejora correspondiente

- Implantar técnicas y procedimientos para el desarrollo de la gestión de la economía de la salud, a fin de cumplir con los objetivos institucionales.

- **Analista de Presupuesto**: Es la persona encargada de realizar estudios económicos que permitan determinar el monto que debe ser asignado al presupuesto de cada área de la Dirección del Seguro de Salud para el desarrollo de sus actividades.
- **Jefe del departamento de Promoción y Prevención:** Es la persona encargada de coordinar y ejecutar planes, programas y proyectos en beneficio de la salud de los afiliados y sus dependientes. (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

La Figura 13 muestra la estructura jerárquica de los StakeHolders de la Dirección de Salud, en donde distinguen al COO, CFO, Jefe de Liquidación y Pagos y Médico Auditor como expertos del negocio o usuarios clave.

Figura 13 Estructura jerárquica de los StakeHolders Dirección de Salud

Estado actual del proyecto

La Dirección de Salud está en la capacidad de realizar únicamente análisis descriptivos con la herramienta Oracle Discoverer, para lo cual emplea Reportes construidos que muestran información como la que se describe en los siguientes ejemplos:

- Número de atenciones médicas brindadas en el año 2015

- Montos pagados ($) a prestadores de salud en servicios de emergencia por prestaciones brindadas a militares en servicio activo.

Se identifica que la Dirección de Salud no cuenta con herramientas que permitan realizar Análisis Predictivo.

3.1.1.2. Objetivos del negocio

- Determinar la combinación de factores que influye en la adquisición de una enfermedad músculo - esquelética en un afiliado
- Identificar el pago de enfermedades músculo esquelético por tipo de servicio y el porcentaje de consumo a nivel del presupuesto.
- Utilizar los resultados obtenidos del modelo como un insumo para la elaboración de planes preventivos.

3.1.1.3. Criterios de éxito del negocio

Se mencionan los criterios de éxito considerados para el negocio:

- Obtención de información confiable y consistente para la elaboración de programas de prevención de trastorno músculo esqueléticos.
- Disponibilidad de un informe gerencial que permita conocer la tendencia del pago por servicios de salud

3.1.2. Evaluación de la situación

3.1.2.1. Inventario de recursos

Las fuentes de datos que se utilizarán para el desarrollo del proyecto son principalmente la información transaccional almacenada en la base de datos del ISSFA, y algunos archivos excel que puedan aportar conocimiento adicional. El detalle de las fuentes de datos se muestra en la Tabla 4.

Tabla 4
Inventario de fuentes de datos

Fuente de datos	Tipos de Fuente	Herramienta para extracción	Información Relevante
Base de datos del ISSA	Transaccional	Oracle Discoverer	Afiliados, Costo por enfermedad, Información de planillaje de atenciones por servicios de salud
Información generada por humanos	Archivos planos	Hoja de cálculo	Informes médicos
Internet	Web	Hoja de cálculo	Patologías músculo esqueléticas

En los recursos humanos disponibles para la ejecución del proyecto se cuenta con personal técnico así como expertos en el dominio del negocio. El detalle del recurso disponible se visualiza en la Tabla 5:

Tabla 5
Detalle del recurso humano disponible para el proyecto

Rol Desempeñado	Persona
Expertos en minería de datos	Ing. Karina Anasi Ing. Ana María Carrazco
Expertos en el dominio	Jefe del Departamento de Prestaciones Hospitalarias Jefe del Departamento de Economía de la Salud Jefe de la sección de Liquidación y Pago Auditor Médico

3.1.2.2. Supuestos y restricciones

A continuación se detallan los supuestos considerados para el desarrollo del proyecto:

- Se descartará cualquier tipo de datos heterogéneos como notas de voz, imágenes, videos, e información física como recetas o fichas médicas escritas a mano.
- Los datos empleados para la construcción del modelo corresponderán al periodo de tiempo de 2012 a julio 2017.

3.1.3. Determinación los objetivos de minería de datos

3.1.3.1. Objetivos de minería de datos

- Clasificar la enfermedad del tipo Musculo esquelética a la que es propenso un afiliado de acuerdo a sus características como género, fuerza, categoría, grado, grupo etario.
- Predecir el pago de enfermedades músculo esqueléticas facturadas por servicios de salud.
- Identificar las características de los afiliados en los que se paga mayor cantidad de dinero.

3.1.4. Generación el plan de proyecto

3.1.4.1. Plan de proyecto

A continuación en la Tabla 6, se muestra el plan de proyecto:

Tabla 6 Plan de Proyecto

PLAN DEL PROYECTO		
1. Entendimiento del negocio	20	20
2. Entendimiento de los datos	20	
2.1. Colección inicial de los datos		5
2.2. Describir los datos		5
2.3. Explorar los datos		5
2.4. Verificar calidad de datos		5
3. Preparar los datos	30	
3.1. Selección de datos		6
3.2. Limpieza de datos		6
3.3. Construcción de datos		6
3.4. Integración de datos		6
3.5. Formateo de datos		6
4. Modelado	20	
4.1. Seleccionar la técnica de modelamientos		5
4.2. Generar diseño de pruebas		5
4.3. Construir el modelo		5
4.4. Evaluar el modelo		5
5. Evaluacion	12	
5.1. Evaluacion de resultados		4
5.2. Proceso de revisión		4
5.3. Determinar pasos siguientes		4
6. Despliegue	12	
6.1. Plan de despliegue		3
6.2. Plan de monitoreo y mantenimiento		3
6.3. Producir el reporte final		3
6.4. Revisión del proyecto		3
TOTAL DÍAS	**114**	**114**

3.1.4.2. Evaluación inicial de herramientas y técnicas

Del estado del arte desarrollado en el capítulo I y de investigaciones (Divya & Agarwal , 2013) y (Chandra Pandey & Allahabad, 2017), se evidencia que la tarea de clasificación se utiliza generalmente en el ámbito de la salud, para realizar la predicción de los pagos de tratamientos de diferentes enfermedades [8]. Con la finalidad de conocer el modelo con mayor precisión y menor tasa de error, se emplearán las siguientes técnicas: árboles de decisión, redes neuronales, SVM y regresión lineal.

Para la selección de estas técnicas de minería de datos se ha considerado el estudio comparativo publicado en el año 2017 “Data mining techniques for medical data: A review” y los criterios de la sección Estado del Arte del capítulo II.

La Tabla 7 muestra la evaluación de criterios mencionados:

Tabla 7
Criterios para selección de técnicas

Criterio	Árboles de decisión	Redes Neuronales	SVM	Regresión Lineal
Manejo de datos heterogéneos	Si	Si	Si	Si
Fácil de interpretar	Si	No	No	Si
Maneja valores faltantes y ruido de forma apropiada	Si	Si	Si	Si
Soporta gran cantidad de registros	Si	Si	Si	Si
Ampliamente utilizado	Si	Si	Si	Si
Soporta gran cantidad de variables de entrada	Si	Si	Si	Si
Requiere adecuado tiempo y costo de procesamiento computacional	Depende del conjunto de datos	Generalmente el procesamiento es alto	Depende del conjunto de datos	No
Es un clasificador altamente preciso	Depende del conjunto de datos	Depende del conjunto de datos	Si en comparación de los mostrados en la	Si en comparación de los mostrad

			tabla	os en la tabla
Supera inconvenientes de sobreajuste	Depende del conjunto de datos	Si	Si	Si

Fuente: (Chandra Pandey & Allahabad, 2017)

En la selección de las herramientas se ha considerado la investigación realizada por los analistas que elaboran el reporte del Cuadrante Mágico de Gartner para Plataformas de Ciencia de los Datos cuyo último estudio corresponde al 14 de Febrero del 2017, esta investigación se enfoca en las herramientas que tienen alta demanda en la funcionalidad de predicciones.

Gartner define las Plataformas de Ciencia de los Datos como: “Una aplicación de software cohesivo que ofrece una mezcla de esenciales bloques de construcción básicos para crear todo tipo de soluciones de Ciencia de Datos y para incorporar esas soluciones en procesos del negocio, rodeado de infraestructura y productos”, adicionalmente que los bloques construidos son módulos que se integraran dentro de una sola plataforma. (Gartner, 2017)

Se consideran las siguientes características de las herramientas que están ubicadas en el cuadrante de Líderes

- Acceso a los datos
- Preparación de los datos
- Visualización y exploración de los datos
- Automatización
- Interface del Usuario
- Aprendizaje de máquina
- Otra analítica avanzada
- Flexibilidad, extensibilidad y apertura
- Rendimiento y Escalabilidad
- Desarrollo
- Plataforma y administración del proyecto

- Manejo del modelo
- Soluciones precisas
- Colaboración
- Coherencia.

En función de estas características, en la Figura 14, se muestran los proveedores de las herramientas en el cuadrante mágico de Gartner.

Figura 14 Cuadrante de Gartner para herramientas de minería de datos
Fuente: (Gartner, 2017)

Dentro del cuadrante de Líderes, se ha escogido al proveedor **RapidMiner**, y su programa Rapid Miner Studio 7.6 en su versión Educational, a continuación se muestra una descripción de sus ventajas (Gartner, 2017):

Ventajas:

- Amplia gama de algoritmos, capacidad de modelamiento flexible, integración de varias fuentes de datos y preparación de datos
- Fácil aprendizaje y utilizar, rápido desarrollo del modelo.
- Comunidad de usuarios que brindan soporte.
- Versión educacional sin costo y con manejo de más de 10000 filas.
- Pre diseños de modelos que resuelven casos comunes en el negocio.
- Experiencia de las Autoras.

3.2. Fase de Entendimiento de los Datos

3.2.1. Colección inicial de los datos

La base transaccional del ISSFA será la única fuente de extracción de datos, los datos corresponden al periodo del año 2012 al julio del 2017. Existen atributos en las tablas y vistas seleccionadas que no representan significancia para la generación del modelo.

En la

Tabla **8**, se describe la información necesaria, su disponibilidad y las características seleccionadas:

Tabla 8
Recolección Inicial de Datos

Información Necesaria	Disponibilidad	Características Seleccionadas
AFILIADOS Y DEPENDIENTES Información que describe al afiliado y sus dependientes.	Base de Datos: ISSFA Esquema: Afiliación Disponible: Sí	Categoría Edad Sexo Fuerza Rango Provincia y ciudad

PLANILLAJE Información sobre los pagos que los diferentes Prestadores de Salud facturan al ISSFA por los servicios brindados a los afiliados.	Base de Datos: ISSFA Esquema: UPM Disponible: Sí	Diagnóstico Servicio Atenciones Valor Pagado Valor Solicitado Fecha de Atención Prestador de Salud
PRESTADORES DE SALUD Información sobre Cada Unidad de Salud: Hospitales privados, públicos, de la RPIS	Base de Datos: ISSFA Esquema: UPM Disponible: Sí	Nivel Tipo Provincia Región Administrado por

La Figura 15 se muestra el diagrama físico de las tablas involucradas en el proceso de recolección de datos.

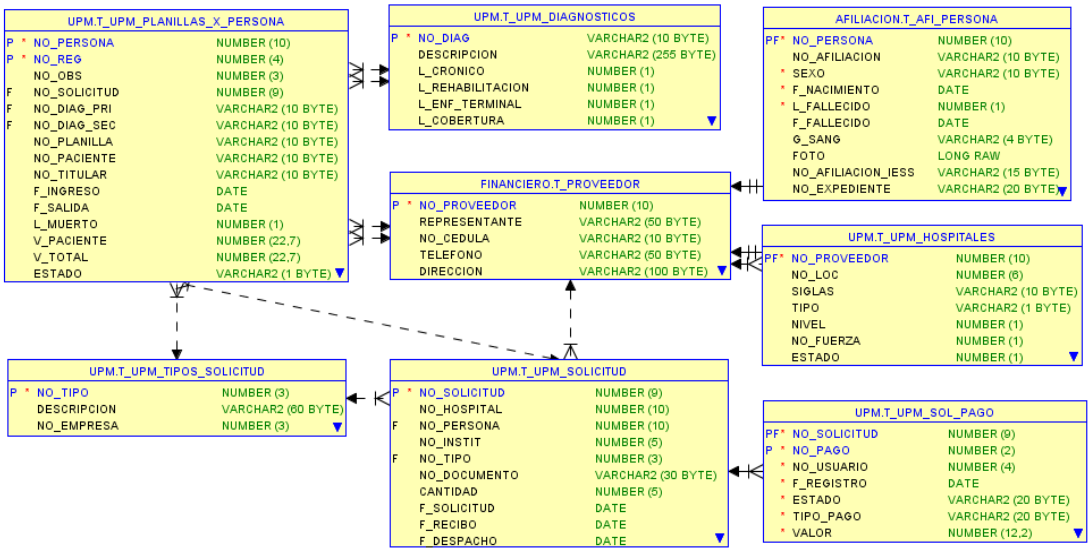


Figura 15 Diagrama físico de las tablas para la recolección de datos
Fuente: Base de datos ISSFA

3.2.2. Descripción de los datos

Partiendo de la recolección inicial de los datos, se determina como necesarios los siguientes objetos de la Base de Datos:

V_AFI_PERSONA

V_DIS_SOLICITUD

T_UPM_PLANILLAS_X_PERSONA

A continuación se muestra la descripción de la tabla o vista y los atributos que se utilizarán para la construcción del modelo:

V_AFI_PERSONA

Este objeto es una vista de base de datos que contiene información personal del afiliado. Cuenta con 233930 registros, se consideraron importantes 10 atributos. En la Tabla 9 se muestra la descripción de los datos

Tabla 9
Descripción de la vista V_AFI_PERSONA

Nombre del campo	Descripción	Tipo de dato	Valores Nulos	Rangos/ Número de registros (regs)
NO_PERSONA	Clave primaria (PK) y código identificador de un afiliado. El código se genera automáticamente cuando se crea un nuevo registro	Numérico	0 = 0%	500000regs
SEXO	Género del afiliado	Nominal	0 = 0%	MASCULINO= 129490 regs FEMENINO= 104440 regs
L_FALLECIDO 1 = Persona viva 0= Persona fallecida	Indica si un afiliado ha fallecido	Booleano	0 = 0%	
EMPRESA	Fuerza a la que pertenece un afiliado.	Nominal	145960 =62%	FUERZA TERRESTRE = 58459 regs FUERZA NAVAL = 17250 regs FUERZA AÉREA = 12261

				regs
GRADO	Grado Militar del afiliado.	Nominal	145960 =62%	Treinta y seis (36) grados militares
PAIS	País de residencia del afiliado.	Nominal	67496 = 29%	Treinta y seis (35) países, en Ecuador se encuentran 165356 registros
PROVINCIA	Provincia de residencia del afiliado.	Nominal	142523 = 61%	Cuarenta y dos (42) provincias o estados. Prevalece Pichincha con 31188 registros
TS_MESES	Meses de servicio en la fuerza	Numérico	0=0%	Min = 0 Máx = 1235 Media = 78.807
TIPO_GRADO	Indica el tipo de grado del afiliado	Nominal	0=0%	TROPA = 74998 regs OFICIAL = 11908 regs ASPIRANTE = 1063 regs ? = 145961 regs

V_DIS_SOLICITUD

Vista que contiene información de una solicitud, en el contexto del negocio, una solicitud representa al conjunto de planillas facturadas durante un mes por Prestador de Salud, es decir es el número de trámite bajo el cual se ejecutará el proceso de Liquidación y Pertinencia Médica. La vista contiene 16 atributos y 28679 registros.

En la **Tabla 10**, se especifican los atributos a considerar en la vista.

CONTINUA

Tabla 10
Descripción de la vista V_DIS_SOLICITUD

Nombre del campo	Descripción	Tipo de dato	Valores perdidos	Rango / Número de registros (regs)
NO_SOLICITUD	Clave primaria (PK) autogenerada, identificador de la solicitud	Numérico	0=0%	
TIPO_SOLICITUD	Descripción del tipo de Servicio de la Solicitud	Nominal	0=0%	Once (11) once tipos de servicio. Predomina Re posición de Gastos Hospitalarios con 10215 registros
HOSPITAL	Nombre del prestador de Salud	Nominal	12486 = 38%	Doscientos sesenta y cinco (265) prestadores de salud. Predomina el Hospital de Especialidades de las Fuerzas Armadas N. 1 con 561 registros
FECHA_DESDE	Fecha de inicio de la cobertura médica	Fecha	0 = 0%	Min = 01/01/2012 Máx = 01/06/2017
FECHA_HASTA	Fecha de fin de la cobertura médica	Fecha	3 = 0%	Min = 01/01/2012 Máx = 01/06/2017
F_DESPACHO	Fecha de despacho de la solicitud	Fecha	5229 = 16%	Min = 01/01/2012 Máx = 30/06/2017
ESTADO_ORIGINAL **D = Despachada** **N = Negada** **T= En proceso** **P= Pendiente**	Estado de la solicitud	Nominal	0 = 0%	D = 28062 regs N = 820 regs T= 2602 regs P= 57 regs R= 736 regs G= 796 regs

R= Recibida G= Generada				
ESTADO_SOLICITUD	Descripción del estado de la solicitud	Nominal	0 = 0%	Despachada = 28062 regs Negada = 820 regs En proceso= 2602 regs Pendiente= 57 regs Recibida= 736 regs Generada= 796 regs
V_SOLICITADO	Monto solicitado ($) por el prestador de salud por los servicios de salud brindados	Numérico	0 = 0%	Min = 0 Máx = 1719304.07 Media = 8501.4
V_PAGADO	Monto pagado ($) pagado por el ISSFA al prestador de salud	Numérico	0 = 0%	Min = 0 Máx = 1439826.72 Media = 4769.041
NUMERO_AFILIADOS	Cantidad (#) de afiliados atendidos durante un periodo	Numérico	0 = 0%	Min = 0 Máx = 14378 Media = 56.502
NUMERO_ATENCIONES	Cantidad (#) de atenciones realizadas de los afiliados durante ese periodo	Numérico	0 = 0%	Min = 0 Máx = 37992 Media = 117.457

T_UPM_PLANILLAS_X_PERSONA:

Es una tabla que almacena datos sobre el planillaje que los Prestadores de Salud facturan al ISSFA. Esta información ingresa al sistema en 2 formas:

Sistema de Planillaje: es una aplicación web que permite a los Prestadores de Salud registrar ON-LINE la facturación por los servicios de salud brindados

Archivo Plano: Los prestadores de Salud que no utilizan el Sistema Web de planillaje, envían los datos de su facturación en un archivo plano, éste es cargado por los usuarios de la Dirección de Salud en la Base de Datos del ISSFA.

En la Tabla 11 se encuentran 4268567 registros y 12 atributos que se describen a continuación:

Tabla 11
Descripción de la tabla T_UPM_PLANILLAS_X_PERSONA

Nombre del campo	Descripción	Tipo de dato	Valores perdidos	Rango / Número de registros (regs)
NO_PERSONA	Clave primaria de la tabla (PK I), que identifica a un afiliado	Numérico	0 = 0%	Min=3 Max=613483
NO_SOLICITUD	Número de solicitud a la que corresponde una planilla.	Numérico	2119 = 0%	Min=138 Max=74813
DIAGNOSTICO_PRIMARIO	Codificación del diagnóstico principal registrado en la atención de un afiliado	Nominal	3 = 0%	CIE 10 (Clasificación Internacional de Enfermedades) 4464 registros distintos 993 registros únicos
DIAGNOSTICO_SECUNDARIO	Codificación del diagnóstico secundario registrado en la atención de un afiliado	Nominal	16897 = 1%	CIE 10 (Clasificación Internacional de Enfermedades) 4369 registros distintos 977 registros

				únicos
NO_PLANILLA	Número de la planilla	String	28 = 0%	
F_INGRESO	Fecha inicio de atención del afiliado	Fecha	0 = 0%	
F_SALIDA	Fecha final de atención del afiliado	Fecha	73 = 0%	
V_TOTAL	Monto total a pagar ($) por la atención recibida	Numérico	0 = 0%	Min = 0 Max=136781.18 Media=56.883
ESTADO **C = Cancelada** **N= Negada** **P = Pendiente** **F = Facturada** **E = Devuelta**	Estado del proceso de auditoría de una planilla	Nominal	0 = 0%	C = 2714411 regs N= 202101 regs P = 1328273 regs F = 20659 regs E = 3123 regs
CATEGORIA_FECHA	Categoría del afiliado a la fecha de atención	Nominal	0 = 0%	Diez y siete (17) categorías distintas de afiliados

Atributos irrelevantes

Se han considerado como datos irrelevantes los campos que tienen relación con información de contacto del afiliado como números de teléfono, nombres de calles, números de casa, así como su grupo sanguíneo, número de cédula y afiliación.

3.2.3. Exploración los datos

En función del conjunto de datos, se puede extraer la siguiente información realizando una exploración de los datos:

En el ISSFA, los servicios que mayor facturación reciben por enfermedades musculo esqueléticos es el de Hospitalización, seguido de Exámenes y Procedimientos Especiales, mientras que en el que menos

facturación existe es en Atenciones Médicas por Consulta Externa (ver Figura 16).

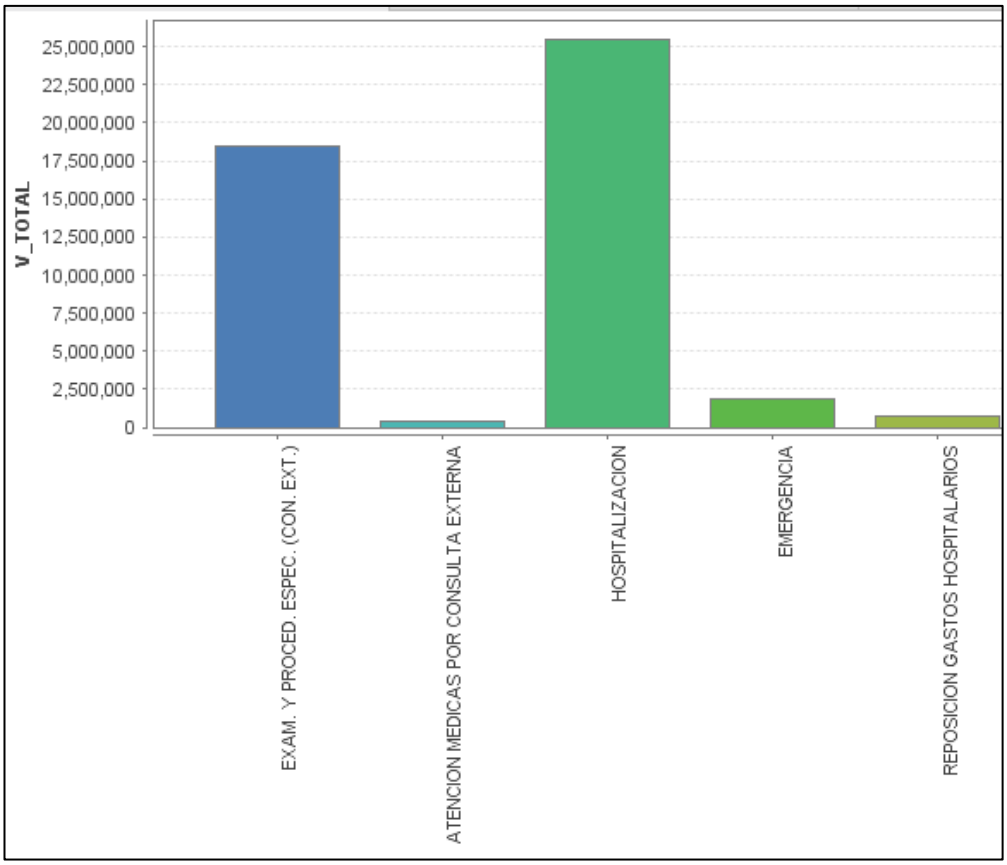


Figura 16 Valor pagado por tipo de servicio

CONTINUA

La Figura 17 muestra que la categoría Activos son los que más facturan enfermedades musculo esqueléticas, a diferencia de la categoría Montepío Hermana:

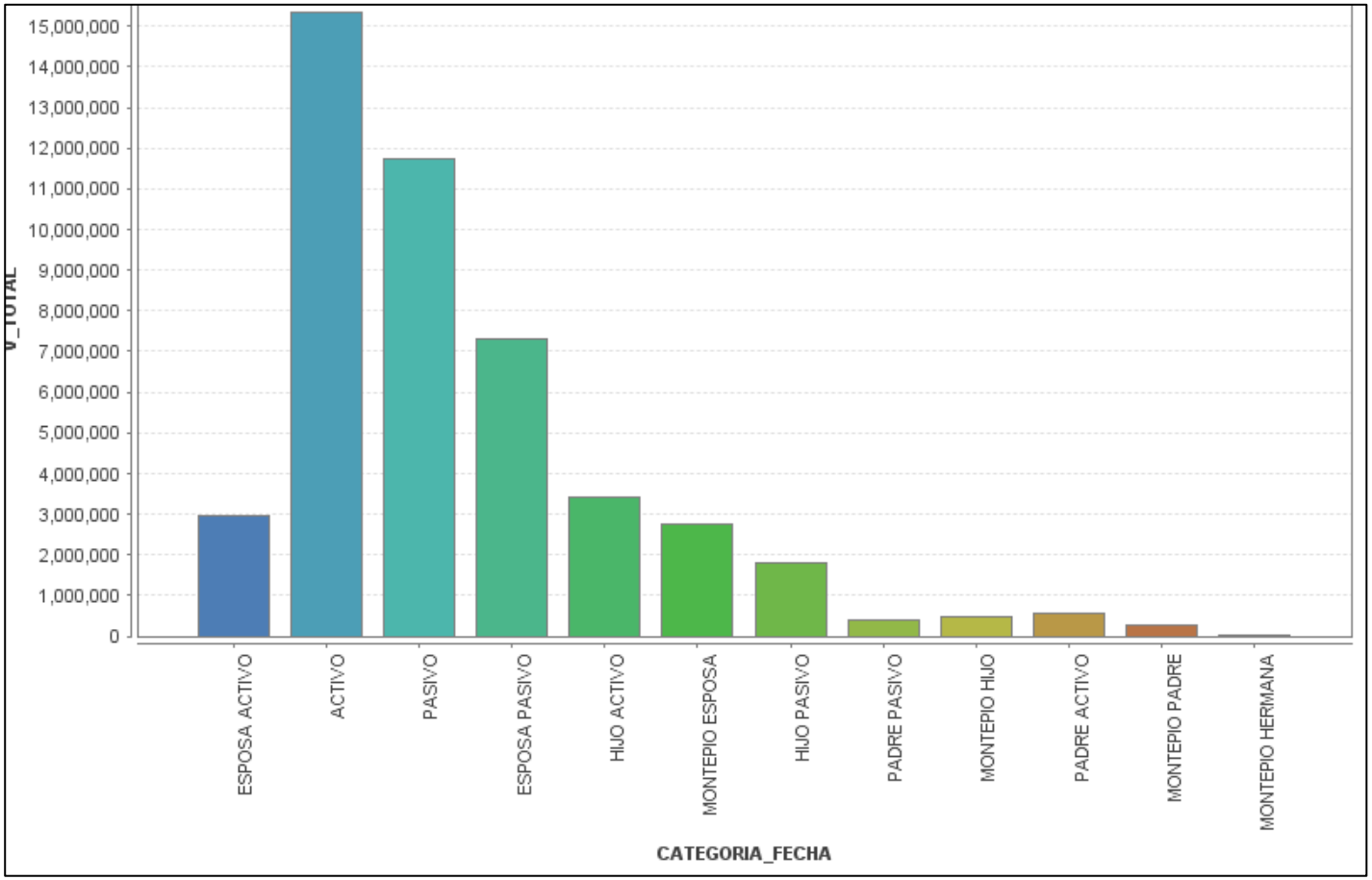


Figura 17 Valor pagado por categoría de afiliado

En la Figura 18, se observa que el ISSFA paga más por enfermedades musculo esqueléticas en dependientes como Esposas, hijos, padres, seguida de la fuerza Terrestre a diferencia de la fuerza aérea que es en la que se paga menos.

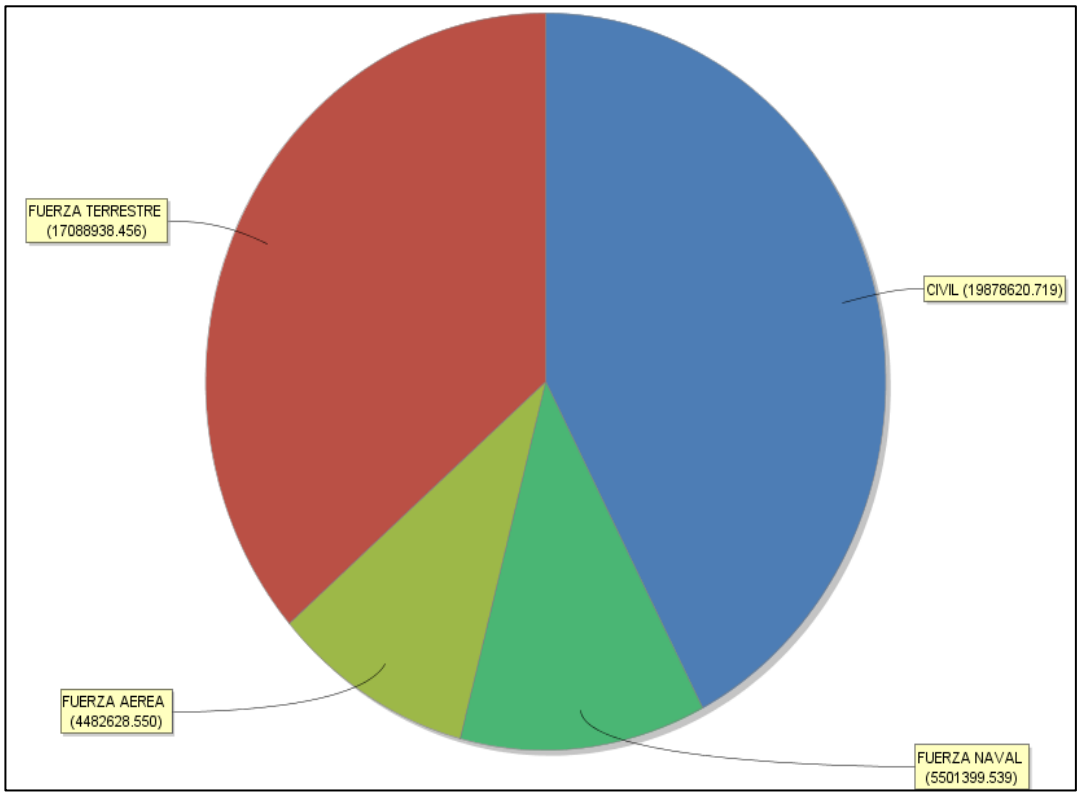


Figura 18 Valor pagado por fuerza
Fuente: Elaboración Propia

El grado que mayor monto paga al ISSFA es el correspondiente a CIVIL, seguido del grado de Sargento primero, Sargento Segundo, Cabo Segundo (ver Figura 19)

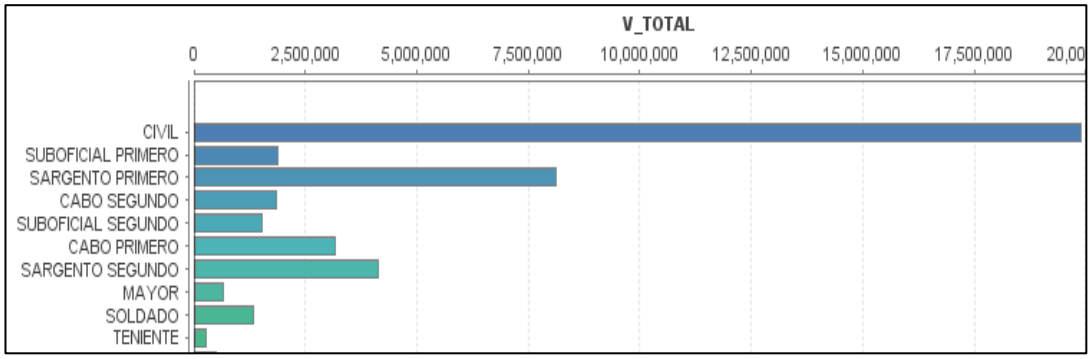


Figura 19 Valor pagado por grado

De acuerdo a la ubicación geográfica, se puede observar que los pagos se concentran en prestadores correspondientes a la provincia de Pichincha. En segundo lugar se encuentran prestadores de la provincia de Guayas, todo esto debido a que existen prestadores de Salud con mayores coberturas y servicios en estas provincias (ver Figura 20):

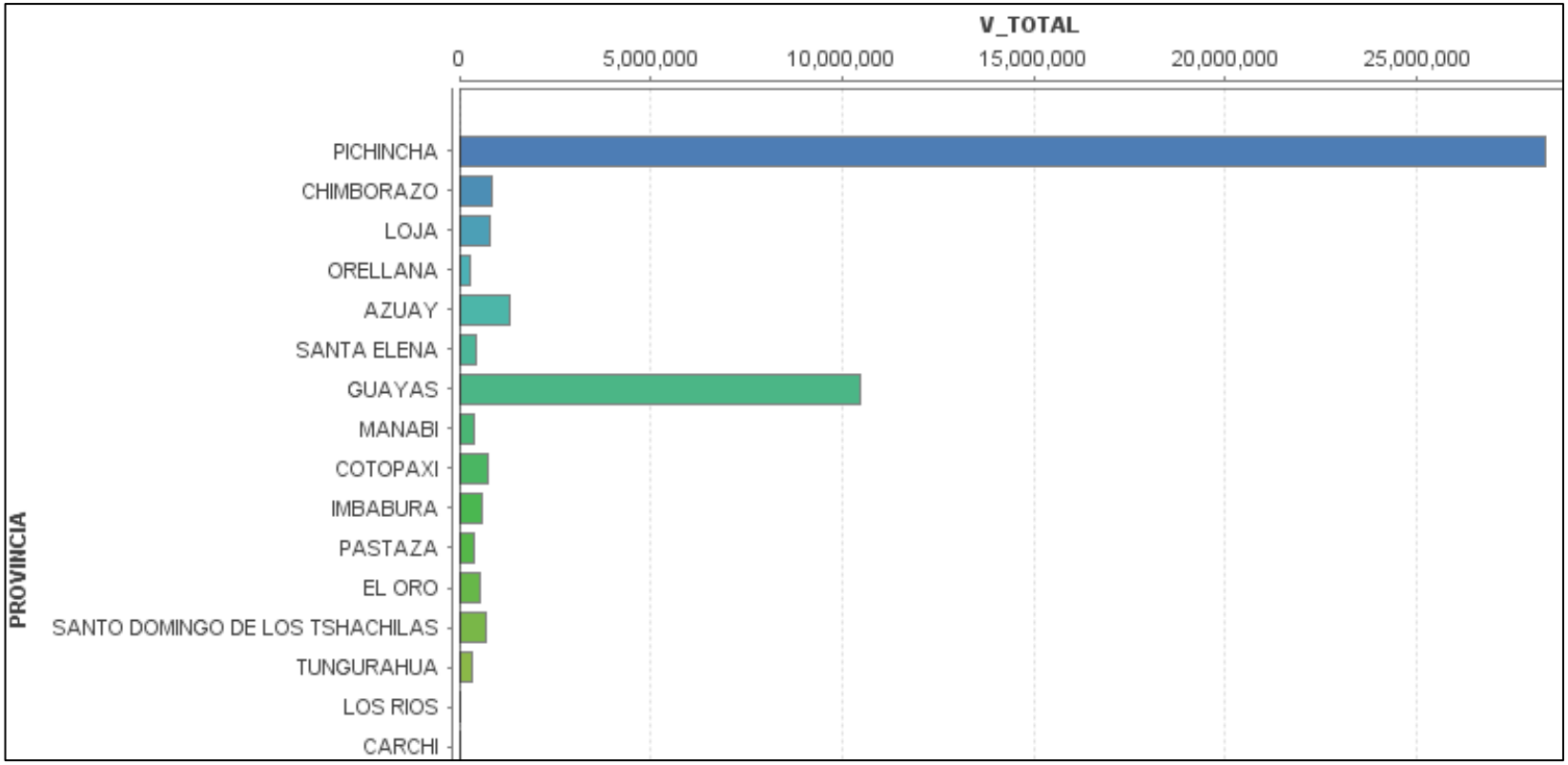


Figura 20 Valor pagado por provincia

Se evidencia que los prestadores de tipo Militar son los que tienen alto índice de facturación a comparación de los de la RPIS, Contratado y Red Complementaria (ver Figura 21)

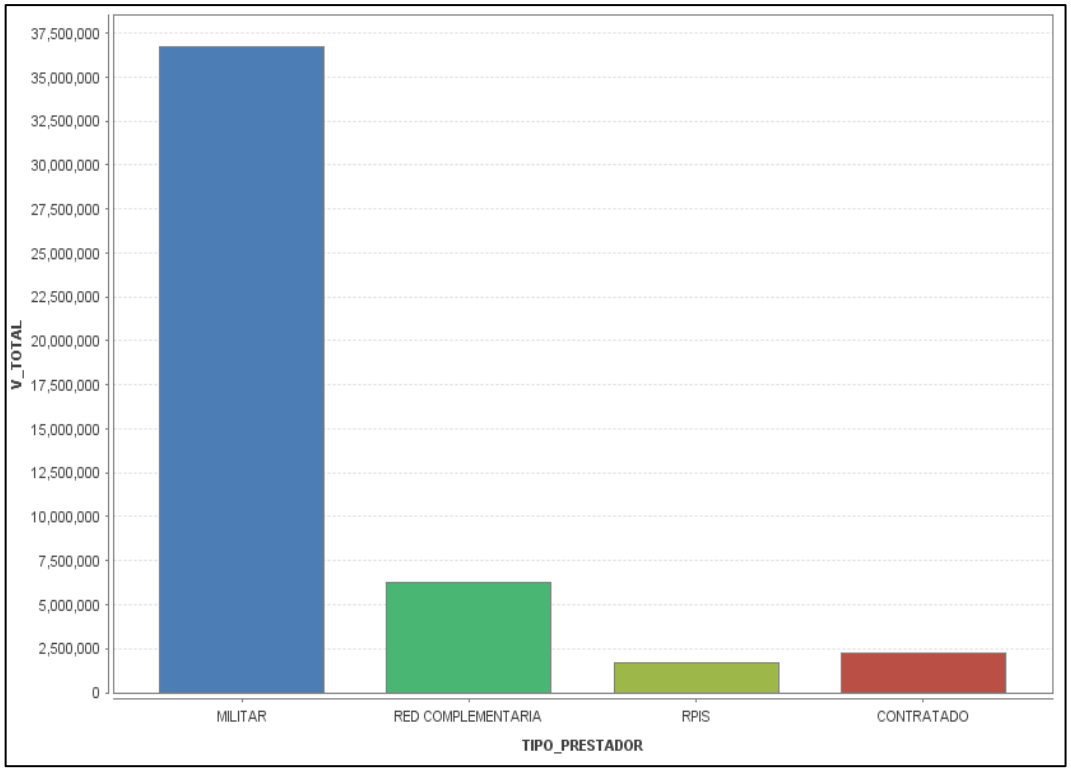


Figura 21 Valor pagado por tipo de prestador

De acuerdo al tipo de lesiones, se visualiza que las Enfermedades del Sistema Osteomuscular son las que representan el mayor pago del ISSFA (ver Figura 22).

CONTINUA

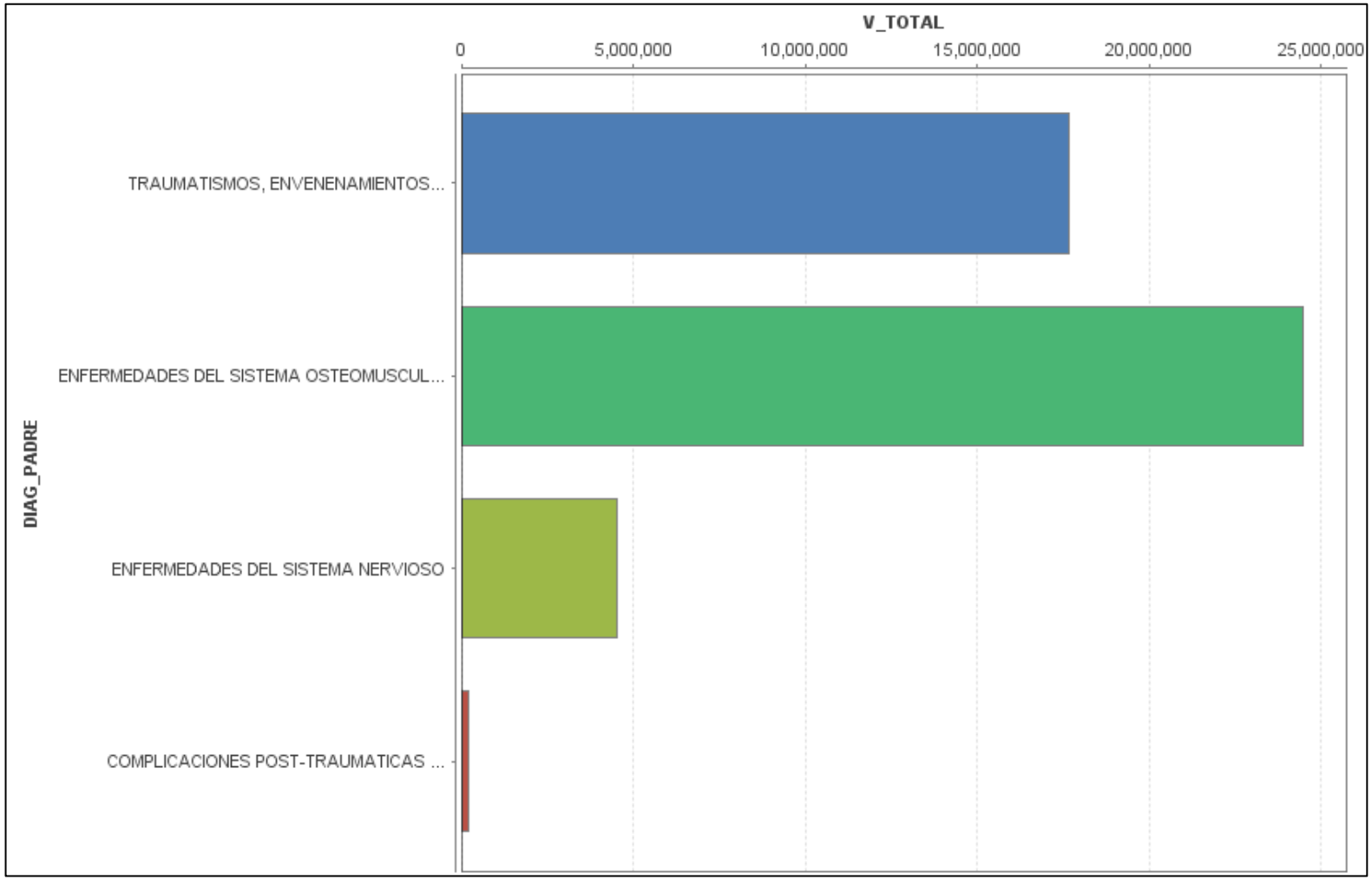


Figura 22 Valor pagado por tipo de patología

En el género del personal militar y sus dependientes, se verifica que el género Masculino es altamente propenso a adquirir este tipo de enfermedades (ver Figura 23)

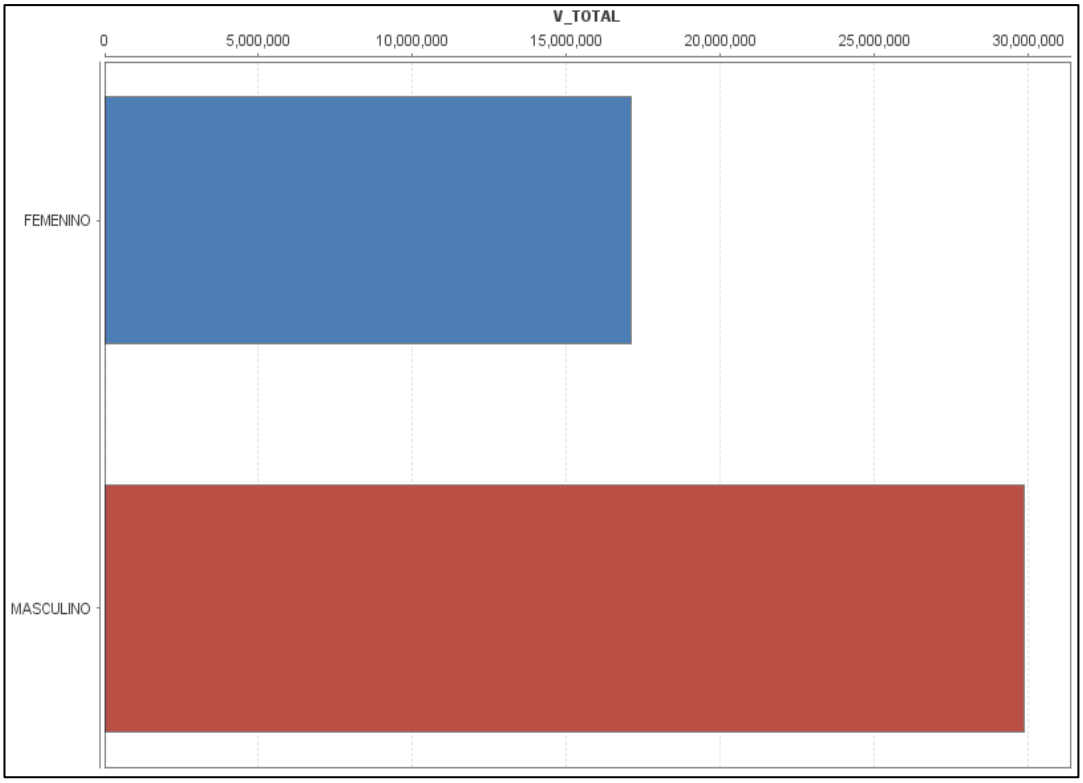


Figura 23 Valor pagado por tipo de militar

En el estado civil de los afiliados existe una mayor concentración de pago en el personal Casado y en menor cantidad el de Unión Libre (ver Figura 24)

CONTINUA

Figura 24 Valor pagado por estado civil

En el campo del valor total, se visualiza un alto grado de dispersión en función del tipo de Servicio de salud (ver Figura 25) , esto se debe a causas como: sucesos extraordinarios que conllevan a pagos extremadamente altos o bajos de acuerdo al tipo de enfermedad, complejidad del tipo de enfermedad, monto del tipo de servicio, entre otros. Según la lógica del negocio, los pagos por servicios no tienen comportamientos similares.

CONTINUA

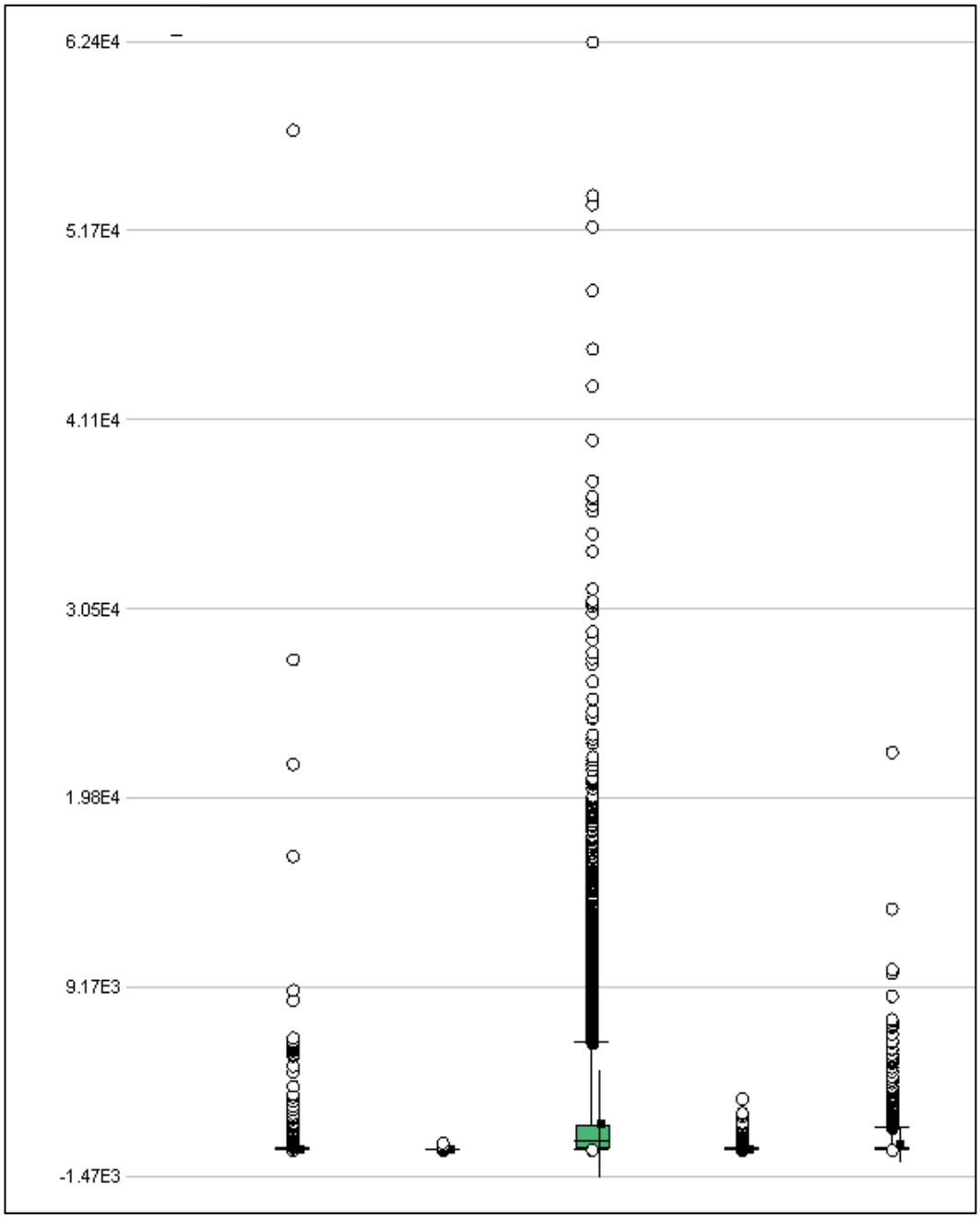


Figura 25 Distribución de datos la clase valor total

3.2.4. Verificación calidad de datos

Con el desarrollo de las actividades anteriores se determina que los datos son de alta calidad por las siguientes razones

- El porcentaje de caracteres especiales o datos almacenados en formatos no estándar de la data es menor al 5%.

- No en todos los casos la ausencia de valores afecta a la completitud de la información, generalmente los valores nulos tienen un significado en el contexto del negocio.
- La presencia de valores nulos que realmente significan valores perdidos se presenta con poca frecuencia (9%) y únicamente en atributos nominales, por lo que esta información puede ser fácilmente completada en base al conocimiento del experto del negocio
- Todos los objetos de base de datos considerados para la obtención de información presentan integridad referencial
- El análisis realizado a los atributos de tipo numérico muestra que menos del 10% de la data contiene valores atípicos

3.3. Fase de Preparación de los Datos

3.3.1. Selección de los datos

En esta sección se considera elaborar criterios de inclusión y exclusión, con el fin de definir un conjunto de datos final que se emplearán para el modelo, todo esto en función del negocio.

3.3.1.1. Criterios de inclusión y exclusión

Los datos se recolectaron únicamente de la base de datos transaccional del ISSFA, en la Tabla 12 se muestran los criterios de exclusión:

Tabla 12
Criterios de Exclusión

Campo	Valores a excluir	Motivo de exclusión
Tipo de servicio	Asistencia odontológica básica Diálisis Reposición de Concentradores de Oxígeno.	Los tipos de servicio no corresponde a trastornos de tipo músculo esqueléticos
Categoría	EX - MILITAR , NINGUNO, NOMINA ISSFA (Servidores públicos que	Las categorías no pertenecen a asegurados del ISSFA

	laboran en la institución)	
Fecha Ingreso	Facturación realizada antes del año 2012 y posterior al año 2016	Incompletitud de datos: El sistema de facturación en línea utilizado por prestadores de salud inicia su utilización en el año 2012, por otro lado en el año actual 2017 aún no se ha realizado por completo la carga de información en la base de datos
Diagnóstico Primario	Nulos	No aportan información para la construcción del modelo
Edad a la fecha de atención	Mayores de 110 años	Datos inconsistentes debido a la fecha de nacimiento ingresadas erróneamente.
Número de solicitud	Nulos	No aportan información para la construcción del modelo
Estado Planilla	F: Facturada E: Edición A: Anulada N: Negada P: Pendiente	No aportan información relevante para la construcción del modelo, porque no registran un valor de pago completo.

Criterios de inclusión

El único criterio de inclusión sobre el conjunto de datos es que el código CIE 10 del diagnóstico primario de la atención pertenezca a la categoría músculo – esquelética. Según los expertos del negocio los diagnósticos que comprenden esta categoría son:

- S00 – T89: Traumatismos
- T90 – T99: Post traumáticas
- G00 – G99: Sistema Nervioso
- M00 – M99: Osteomuscular

3.3.2. Limpieza de los datos

La limpieza de datos se realiza para disminuir o eliminar el ruido, incompletitud e inconsistencia de datos encontrados en los campos seleccionados en el conjunto inicial de datos.

En la sección de Exploración de los datos, se pudo apreciar que existe una gran dispersión del campo Valor Total, por lo que como parte del tratamiento de los datos previo a la construcción del modelo, se realiza la detección de elementos atípicos, que según autores como (Jacob, Ramani, & Nancy, Feature Selection and Classification in Breast Cancer Datasets through Data Mining Algorithm, 2011) representan ruido y desviación del comportamiento de los datos, por lo que deben ser descartados para reducir el costo de procesamiento y esparcimiento de los datos. La detección minuciosa de este tipo de elementos se realizará sobre el campo V_TOTAL puesto que es el atributo Clase.

En la Figura 26, se observa la dispersión del campo V_TOTAL en el servicio Atenciones Médicas por Consulta Externa:

CONTINUA

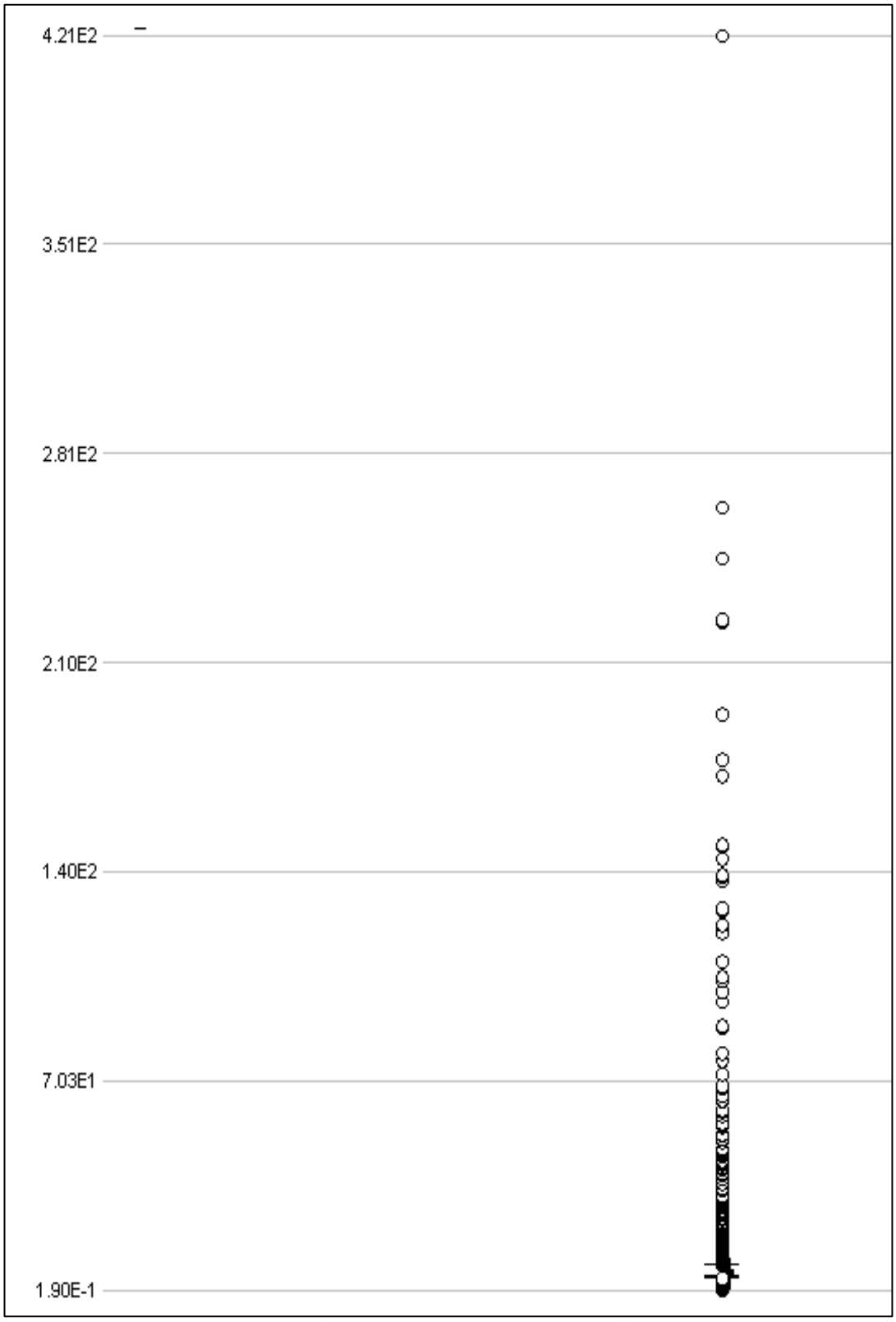


Figura 26 Elementos atípicos de Atenciones médicas por consulta externa

En la Figura 27, se observa la dispersión del campo V_TOTAL en el servicio Emergencia:

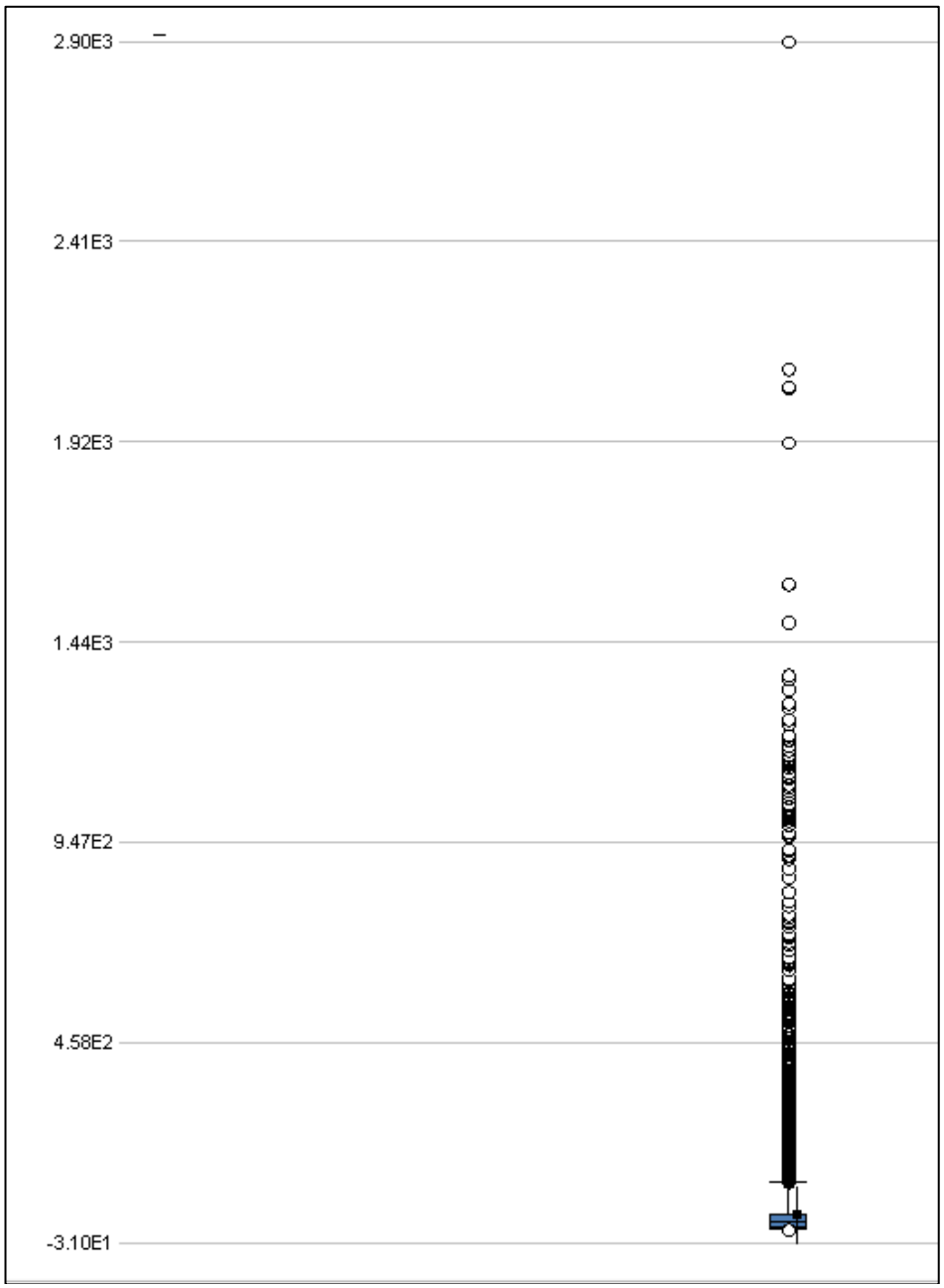


Figura 27 Elementos atípicos en Emergencia

En la Figura 28, se observa la dispersión del campo V_TOTAL en el servicio Exámenes y Procedimientos Consulta Externa:

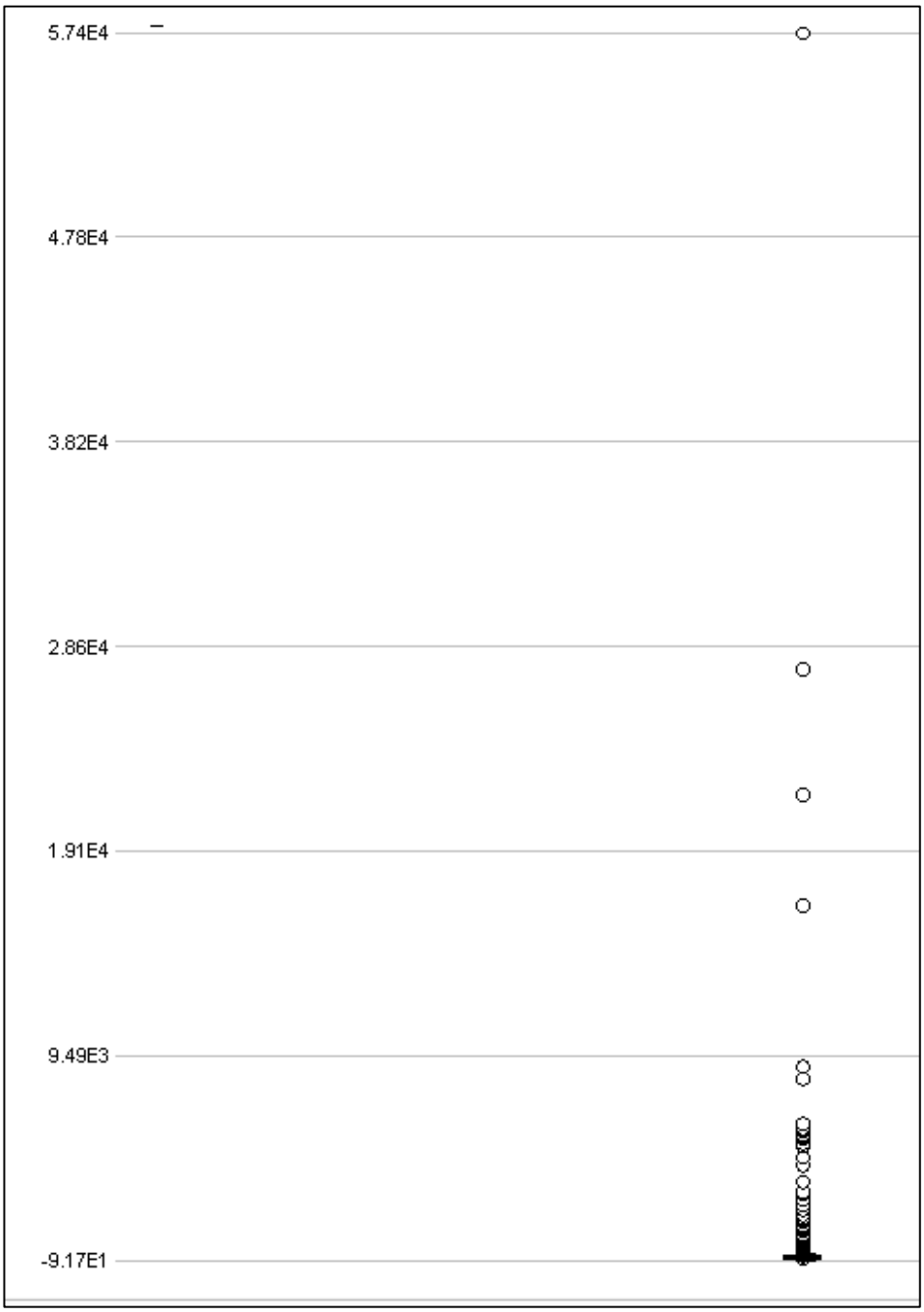


Figura 28 Elementos atípicos de Exámenes y Procedimientos

En la Figura 29, se observa la dispersión del campo V_TOTAL en el servicio Hospitalización:

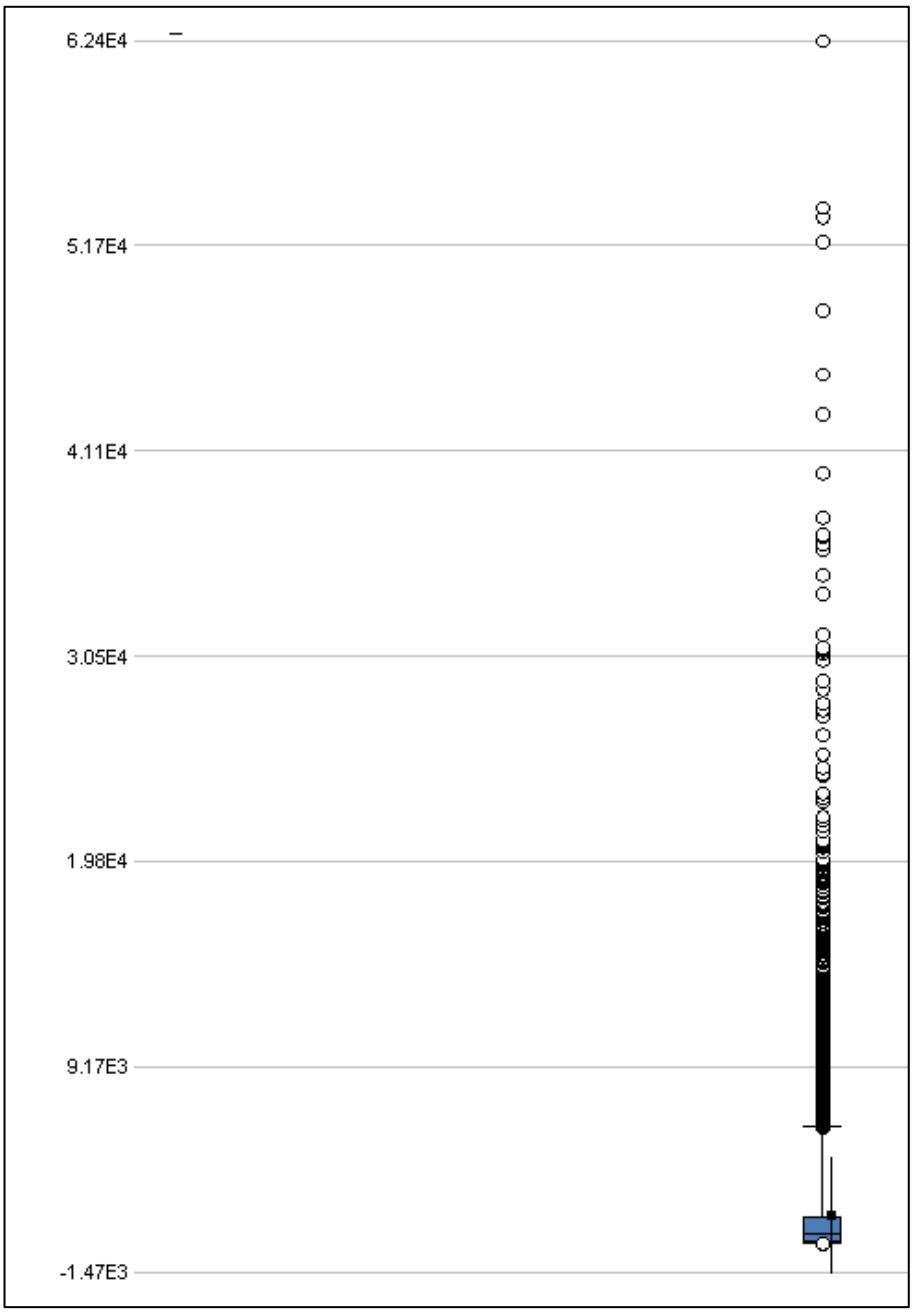


Figura 29 Elementos atípicos en Hospitalización

En la Figura 30, se observa la dispersión del campo V_TOTAL en el servicio Reposición de Gastos Hospitalarios:

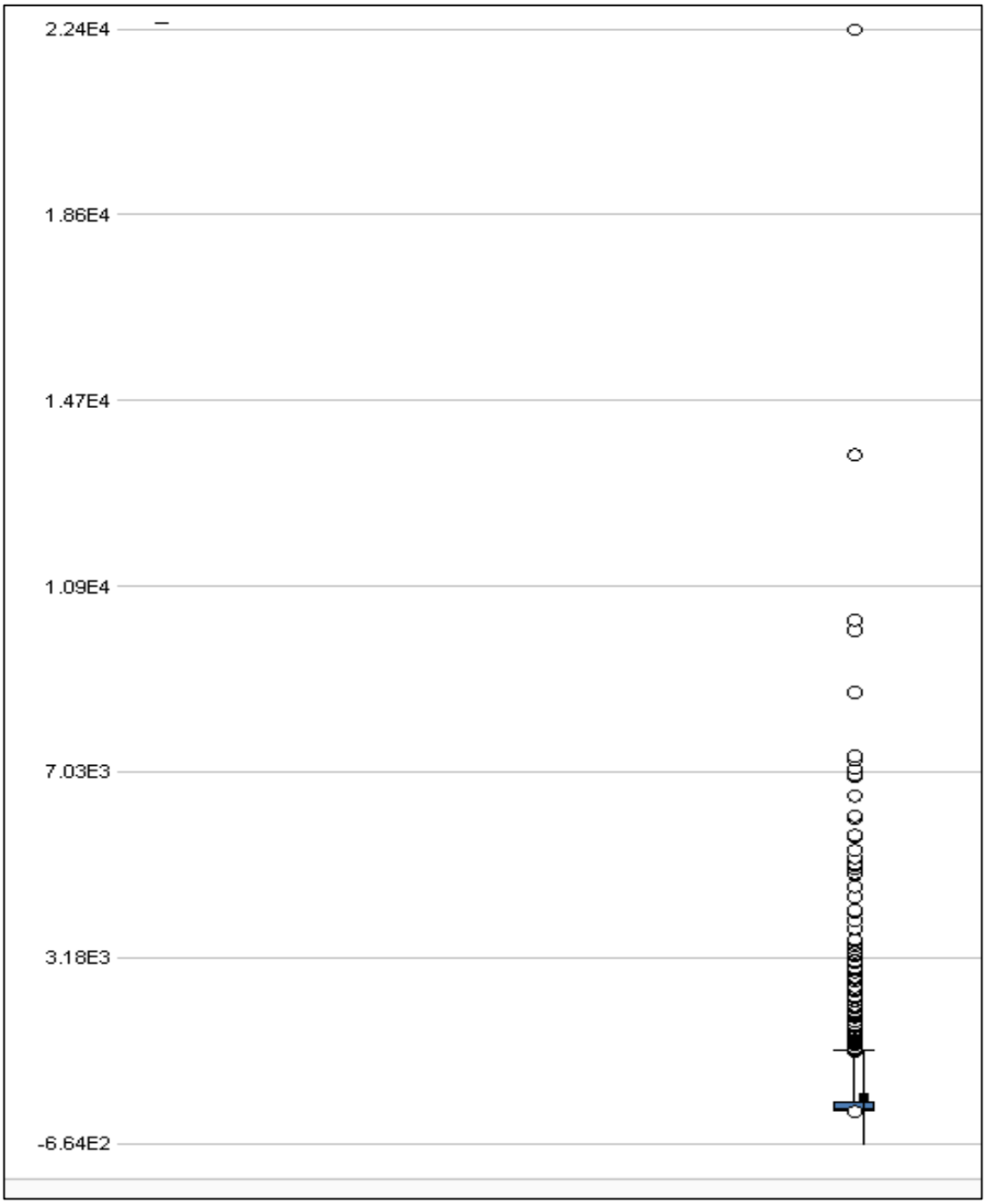


Figura 30 Elementos atípicos en Reposición Gastos Hospitalarios

Como se muestra en la Figura 31, la data inicial tiene una cantidad considerable de elementos atípicos, especialmente en la categoría de Hospitalización y Exámenes y procedimientos, servicios en donde se manejan altas cantidades de dinero.

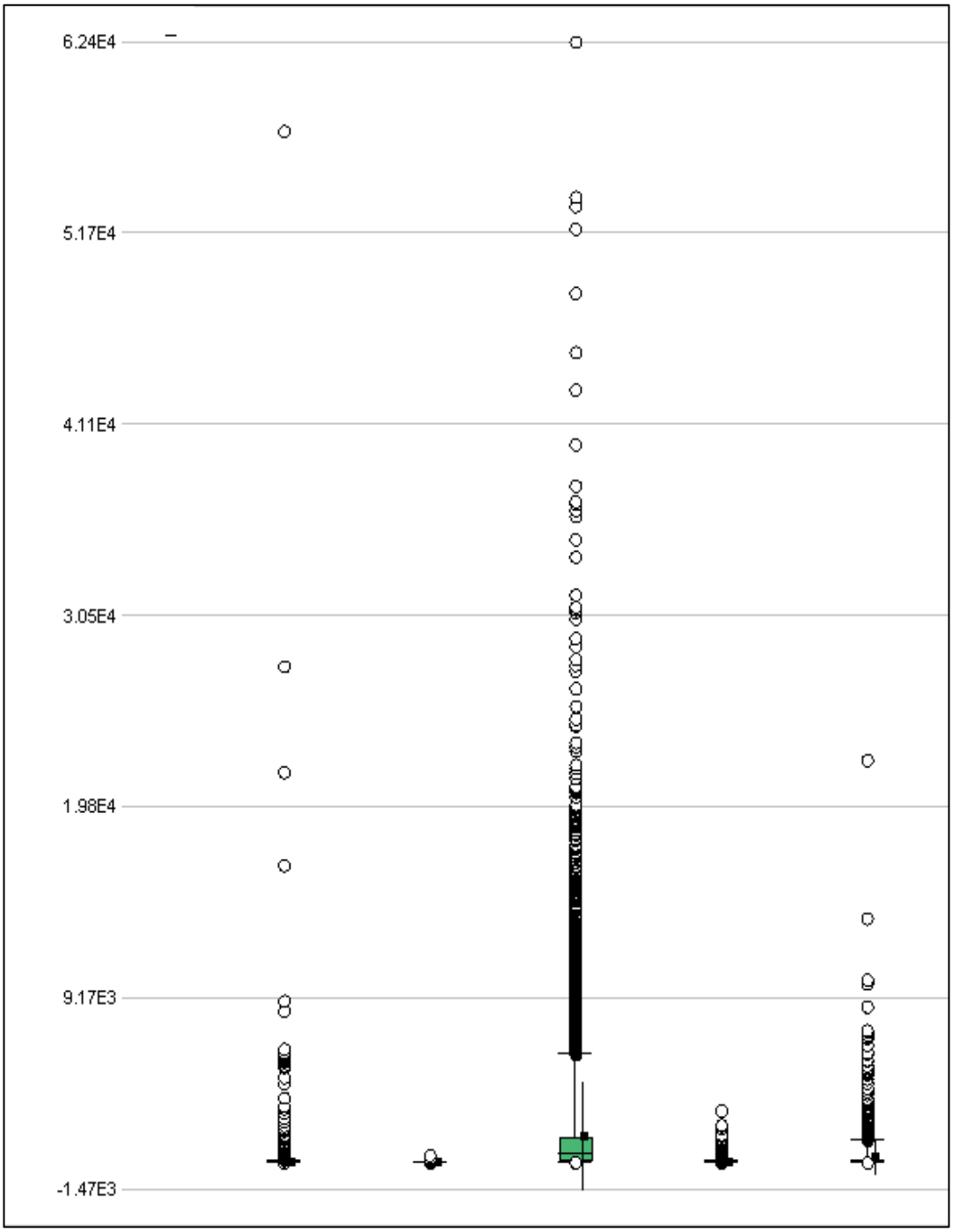


Figura 31 Elementos atípicos por servicio

Por su eficiencia y bajo costo computacional se elige el algoritmo LOF, dentro de los métodos utilizados para el tratamiento de valores atípicos mencionados en el capítulo 2. La herramienta Rapid Miner cuenta con el componente Detect Outlier (LOF), basado en el concepto de densidad, que

identifica las regiones con similar densidad y descarta al resto en base a la distancia configurada, en este caso la euclidiana. En la Figura 32 se muestra el proceso de extracción de outliers del conjunto de datos:

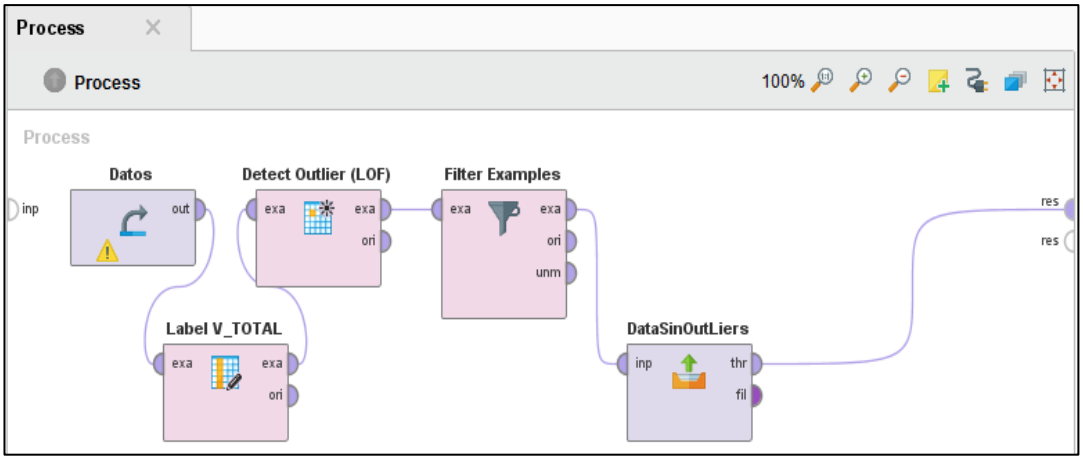


Figura 32 Proceso de extracción de outliers

Fuente: Elaboración Propia

Una vez que se han descartado lo elementos atípicos, se construyen nuevamente los diagramas de caja por servicio, lo que se muestra en la Figura 33.

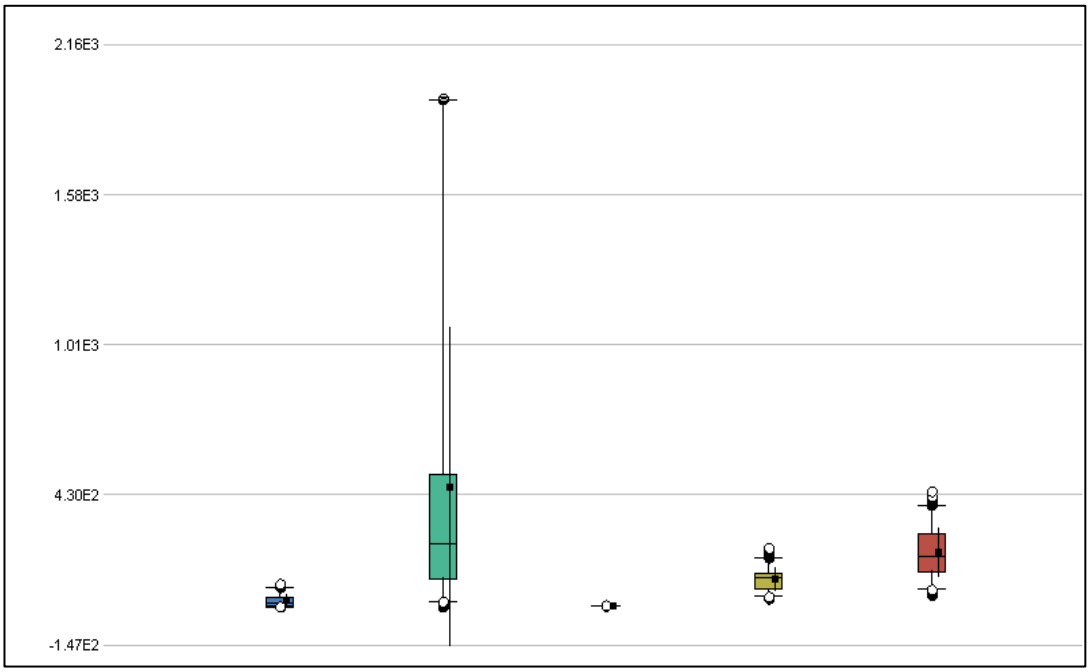


Figura 33 Diagramas de caja de campo V_TOTAL sin elementos atípicos

Fuente: Elaboración Propia

Para la interpretación de los diagramas de caja de la Figura 33, se realiza el cálculo de los cuartiles, rango intercuartil y límites para cada tipo de servicio como se muestra en la Tabla 13.

Tabla 13
Cálculo de Límites por Servicio

Tipo de Servicio	Min	Max	Q1	Q2	Q3	RIC	Limite Superior	Límite Inferior
Hospitalización	0,02	62372,21	170,8	487,995	1427,713	1256,913	3313,08125	-1714,56875
Emergencia	0,01	2902,36	9,1	21,69	40,43	31,33	87,425	-37,895
Atenciones médicas por consulta Externa	0,19	420,79	4,34	5,01	5,01	0,67	6,015	3,335
Exámenes y Procedimientos	0,02	57391,33	11,6	23,96	97,3	87,5	225,85	-116,95
Reposición de Gastos Hospitalarios	0,005	22410	19,75	60	180	160,25	420,375	-220,625

Calidad de datos en elementos atípicos

Con el análisis de elementos atípicos por servicio, en la Tabla 14 se muestra el porcentaje de los ejemplos que se encuentran fuera de rango no supera el 14%, sin embargo sobre el total de la data el porcentaje es de 3.17%, con lo que se concluye que la calidad de datos el alta.

Tabla 14
Porcentaje de elementos atípicos por servicio

Tipo de Servicio	Total Registros	Num. Tipicos	Num. Atipicos	Porcentaje de atípicos
Hospitalización	17158	15074	2084	12,15%
Emergencia	51311	47221	4090	7,97%
Atenciones médicas por consulta Externa	74850	70778	4072	5,44%
Exámenes y Procedimientos	337173	332437	4736	1,40%
Reposición de Gastos Hospitalarios	2337	2028	309	13,22%
Totales	482829	467538	15291	3,17%

3.3.3. Construcción de los datos

Para esta fase se utilizan componentes de la herramienta RapidMiner y scripts de Base de Datos, aplicados en un nuevo esquema llamado “MINERIA_DE_DATOS”

3.3.3.1. Atributos derivados

Los atributos derivados que son necesarios para el modelo son los correspondientes a fecha, provincia. Se toma criterios de reducción de dimensionalidad en los campos de fecha no brinda información significativa, mientras en la Tabla 15 se especifican los campos:

Tabla 15
Transformación de elementos

CAMPO	TRANSFORMACIÓN	VALORES
F_INGRESO	ANIO	TO_CHAR(F_INGRESO, 'YYYY')
	ANIO_MES	TO_CHAR(F_INGRESO, 'YYYY-MM')
	ANIO_TRIMESTRE	TO_CHAR(F_INGRESO, 'YYYY') || '-'||GUPM_REPORTES.F_TRIMESTRE(TO_CHAR(F_INGRESO,'MM'))
	ANIO_SEMESTRE	TO_CHAR(F_INGRESO, 'YYYY') || '-'||GUPM_REPORTES.F_SEMESTRE(TO_CHAR(F_INGRESO,'MM'))
PROVINCIA	REGION	SIERRA COSTA ORIENTE INSULAR

3.3.4. Integración de los datos

Para evitar el impacto en las transacciones de la base de datos, se crea un esquema que contiene el respaldo de las tablas indicadas en la sección anterior. La integración de datos se realiza en la vista “V_PAGOS_SALUD”, construida mediante una sentencia SQL (ver Figura 34), misma que servirá como entrada de datos para le construcción del modelo en RapidMiner.

```
SELECT P.NO_PERSONA, NO_REG,P.NO_SOLICITUD,
NO_DIAG_PRI,
GUPM_DIAGNOSTICO.F_NOMBRE_LARGO(NO_DIAG_PRI) DIAGNOSTICO_PRIMARIO,
NO_PLANILLA,F_INGRESO,F_SALIDA,P.V_TOTAL, P.ESTADO, C_CATEGORIA,
GUPM_CATEGORIA.F_DESCRIPCION(C_CATEGORIA) CATEGORIA_FECHA,
GUPM_REPORTES.F_EDAD_X_FECHA(P.NO_PERSONA,F_INGRESO) EDAD_FECHA,
GUPM_SOLICITUD.F_NO_TIPO(P.NO_SOLICITUD) NO_TIPO,
GUPM_SOLICITUD.F_DESC_TIPO(GUPM_SOLICITUD.F_NO_TIPO(P.NO_SOLICITUD)) TIPO_SERVICIO,
GAFI_CARRERA.F_EMPRESA_PERSONA(P.NO_PERSONA,F_INGRESO) NO_EMPRESA,
GAFI_EMPRESA_GRADO.F_DESC_EMPRESA(GAFI_CARRERA.F_EMPRESA_PERSONA(P.NO_PERSONA,F_INGRESO)) EMPRESA,
GAFI_CARRERA.F_GRADO_PERSONA(P.NO_PERSONA,F_INGRESO) NO_GRADO,
GAFI_EMPRESA_GRADO.F_DESC_GRADO(GAFI_CARRERA.F_EMPRESA_PERSONA(P.NO_PERSONA,F_INGRESO),
      GAFI_CARRERA.F_GRADO_PERSONA(P.NO_PERSONA,F_INGRESO)) GRADO,
GUPM_SOLICITUD.F_NO_HOSPITAL(P.NO_SOLICITUD) NO_PRESTADOR,
      GPRE_PROVEEDOR.F_NOMBRE(GUPM_SOLICITUD.F_NO_HOSPITAL(P.NO_SOLICITUD)) PRESTADOR,
GUPM_HOSPITAL.F_NO_LOC(GUPM_SOLICITUD.F_NO_HOSPITAL(P.NO_SOLICITUD)) NO_LOC,
S.F_DESPACHO, GUPM_HOSPITAL.F_ADMINISTRADOR(S.NO_HOSPITAL) ADMINISTRADOR
FROM T_UPM_PLANILLAS_X_PERSONA P, T_UPM_SOLICITUD S
WHERE F_DESPACHO  BETWEEN '01/01/2012' AND '30/06/2017'
AND P.ESTADO = 'C' OR (P.ESTADO = 'P' AND  S.NO_SOLICITUD IN (SELECT NO_SOLICITUD
FROM T_UPM_SOLICITUD WHERE ESTADO = 'D' AND F_DESPACHO  BETWEEN '01/01/2012' AND '30/06/2017'))
AND F_CATEGORIA(P.NO_PERSONA, F_INGRESO) NOT IN ('NINGUNO','EX-MILITAR','NOMINA ISSFA')
AND (NO_DIAG_PRI LIKE 'M%' OR NO_DIAG_PRI LIKE 'S%' OR NO_DIAG_PRI LIKE 'T%'OR NO_DIAG_PRI LIKE 'G%')
AND V_TOTAL!=0
AND P.NO_SOLICITUD = S.NO_SOLICITUD;
```

Figura 34 Sentencia SQL para integración de datos

3.4. Fase de Generación del Modelado

3.4.1.1. Selección de la técnica de modelamiento

En esta sección se construyen cuatro modelos, de acuerdo a las técnicas definidas en la sección 3, estas técnicas son:

- Árboles de decisión
- Regresión Lineal
- Máquina de vector de soporte SVM[17]
- Redes Neuronales

3.4.1.2. Supuestos del modelamiento

Los supuestos considerados para la construcción del modelo son:

- Considerando que el conjunto inicial de datos cuenta con un número extenso de atributos, se presume que no todos tienen relevancia para la construcción del modelo.
- Existen valores atípicos en el atributo clase que generen ruido en el modelo.
- No se presenta estacionalidad en el atributo clase valor pagado, debido al alto grado de variabilidad que presentan los diferentes tipos de servicio de salud y diagnósticos.

3.4.2. Generación del diseño de pruebas

En esta sección se define el proceso utilizado para la validación y pruebas del modelo, a continuación se detalla la entrada, procesos y salidas esperadas, que serán ejecutadas para todas las técnicas seleccionadas.

Entrada:

Conjunto de datos de planillaje correspondiente a los años 2012 al 2016.

[17] Algoritmo Support Vector Machine

Proceso:

Se utilizará la técnica de validación cruzada o cross-validation, al basarse en la “Ley de Paretto”, se tomará el 80% del conjunto de datos para la fase de entrenamiento y el 20% restante para la fase de pruebas a fin de reflejar la precisión del modelo.

Al tener datos de 5 años completos, se utilizará el componente de RapidMiner Cross Validatión o validación cruzada (ver Figura 35), que divide todo el conjunto de datos en N partes iguales, entrenar N-1 partes y validar la parte restante. La precisión del modelo es la media aritmética del error resultante en cada partición. (RapidMiner, 2017)

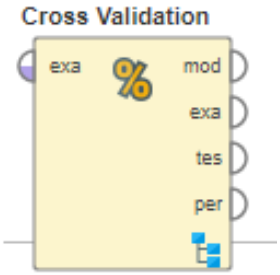


Figura 35 Componente de RapidMiner Cross Validation
Fuente: (RapidMiner, 2017)

Mediante artificios de las autoras y componentes de RapidMiner, por cada tipo de servicio de salud, se obtienen particiones con el mismo número de registros, cada partición representa a un año, de esta forma se entrenan datos de 4 años y se validan los de uno.

A continuación se describe el proceso de construcción del diseño de pruebas en RapidMiner:

- Con el componente Filter Examples se seleccionan los datos de un tipo de servicio y posteriormente de cada año.
- Se construyen cinco particiones con igual número de registros tomados de forma aleatoria utilizando el componente Filter Examples Range.

- Se combinan y ordenan las particiones cronológicamente, utilizando los componentes: Union y Sort respectivamente.

En la Figura 36, se observa el proceso descrito:

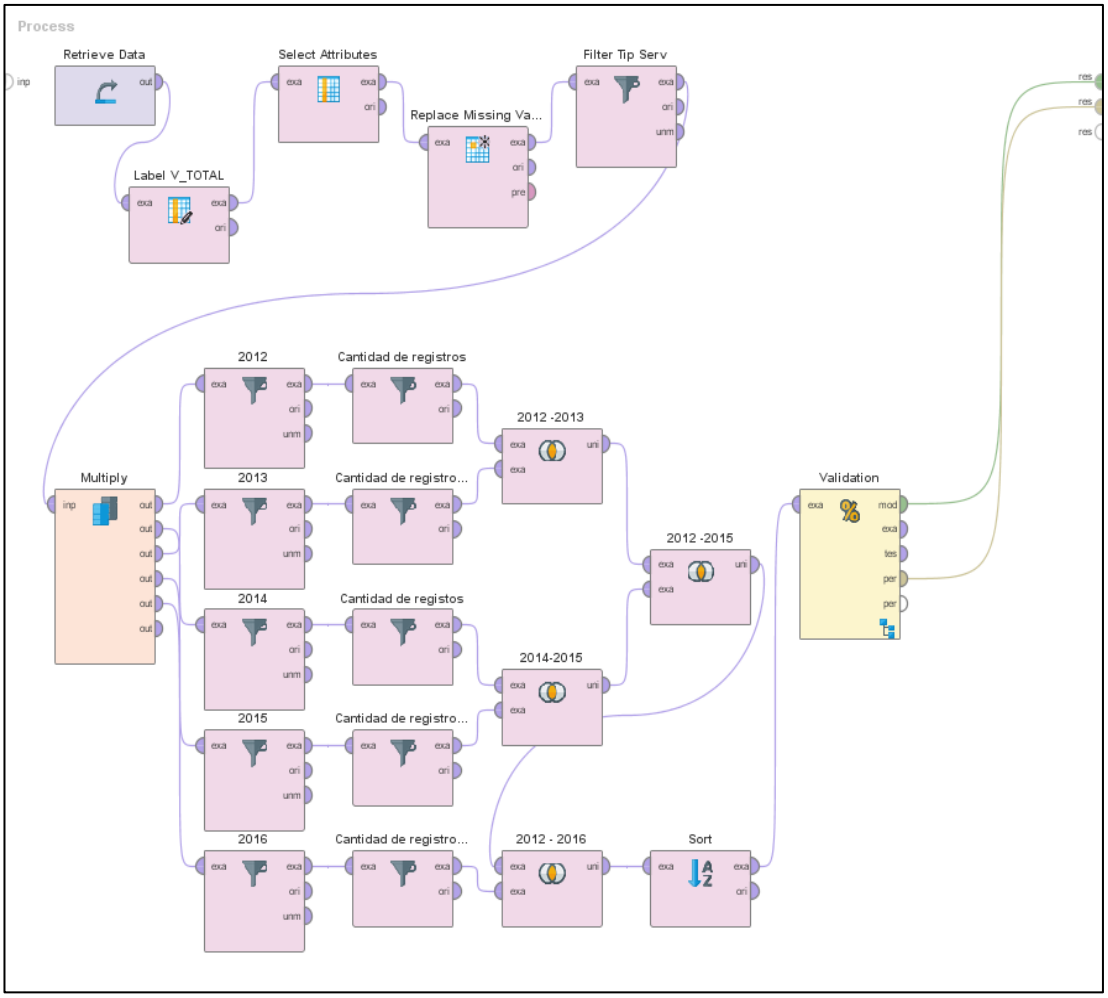


Figura 36 Proceso de validación

- Finalmente, se configura los parámetros del componente Cross Validation:
  - Número de particiones igual al número años del conjunto de datos.
  - Tipo de muestreo lineal, para conservar el orden cronológico de los datos (ver Figura 37)

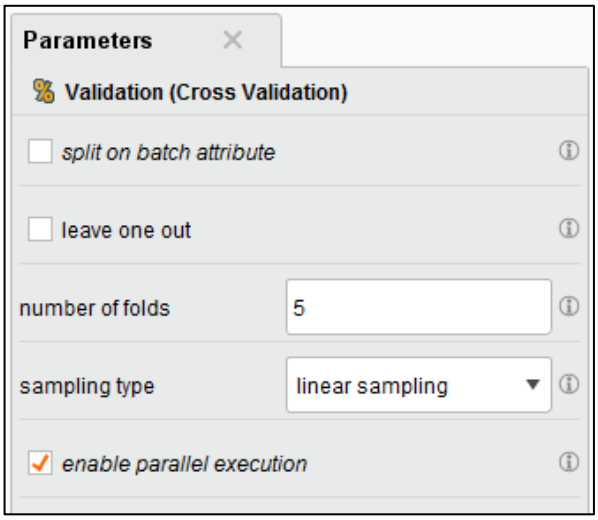


Figura 37 Parametrización de Validación Cruzada

Salida

- Raíz del error cuadrático medio o RMSE[18]
- Suma del valor real y su predicción anual, utilizando el operador Agregate.
- Tiempo de ejecución del modelo en segundos.

3.4.3. Construcción del modelo

Para satisfacer a los objetivos planteados en el capítulo 3, se construyen diferentes modelos, cuya definición del atributo clase es diferente, para algunos modelos corresponde al campo Diag_Padre o diagnóstico padre, y en otros el campo V_Total o valor pagado.

La construcción inicia con la selección de atributos de mayor relevancia con respecto a la clase definida.

3.4.3.1. Selección de atributos

Debido a su amplia efectividad [36] y bajo consumo de memoria (RapidMiner, 2017), el método de selección de características Sequential Floating Search con propagación hacia adelante es utilizado para identificar la relevancia de atributos.

[18] Root Mean Squared Error

Esta técnica inicia con un conjunto vacío de atributos, su selección se realiza mediante un proceso de validación cruzada con el algoritmo del Vecino más cercano en la fase de entrenamiento. Este proceso se repite N veces, obteniendo en cada iteración un subconjunto de atributos que provocan un incremento del rendimiento con respecto a la clase.

En la Figura 38 se observa el proceso de selección de características en RapidMiner con validación cruzada de 10 particiones sobre un conjunto inicial de 50 atributos.

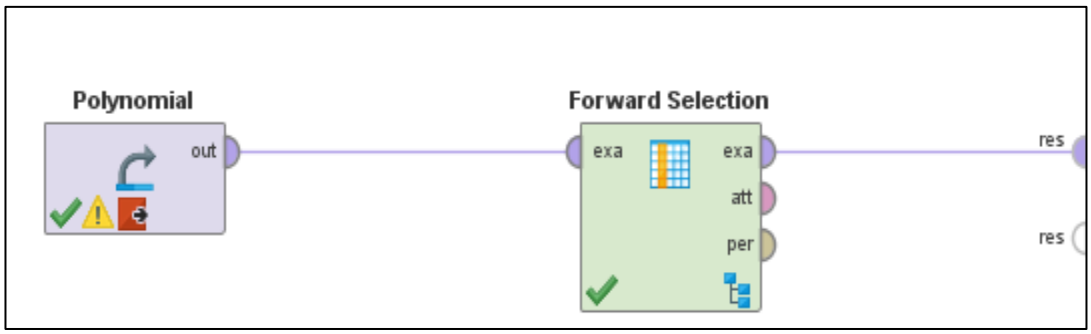


Figura 38 proceso de selección de características

Al definir como clase al campo Diag_Padre de tipo Polinomial, resultado es un subconjunto de ocho atributos compuestos por no_empresa, empresa, grupo_etario, tipo_militar, sexo, ts_meses, grado y estado_civil.

Por otro lado, al definir como atributo clase al campo V_TOTAL, el subconjunto resultante se compone de doce atributos de tipo numérico, entre lo que se encuentran no_persona, no_reg, c_categoria, edad_fecha,no_tipo,no_empresa, no_grado, no_prestador, nivel_prestador, anio, ts_meses y dias_consulta.

Una vez definidos los atributos relevantes con respecto a su clase, se da inicio a la fase de construcción de los diferentes modelos: árbol de decisión, regresión lineal, SVM y redes neuronales.

3.4.3.2. Árbol de decisión

Con el Árbol de decisión, se pretende conocer la combinación de factores que influyen en la adquisición de una enfermedad músculo - esquelética de un afiliado. Para lo cual, se define como el atributo clase al campo diagnóstico padre.

Este componente genera una colección de nodos para crear decisiones sobre los valores de una clase, cada nodo representa una regla de clasificación para los atributos. La ventaja de este componente es que permite la utilización un conjunto de datos compuesto por atributos numéricos y nominales, a diferencia de sus variaciones Chaid y RandomForest que solo tiene conjuntos de datos nominales. (RapidMiner, 2017)

La Figura 39, muestra el modelo utilizando árbol de decisión:

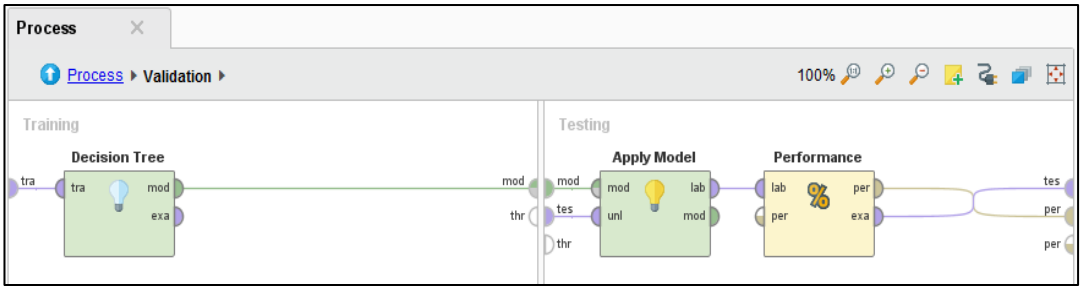


Figura 39 Árbol de decisión

Para un mejor entendimiento del árbol y a razón de que el diagnóstico padre es extenso, se realiza el mapeo de los nombres. (ver Figura 40):

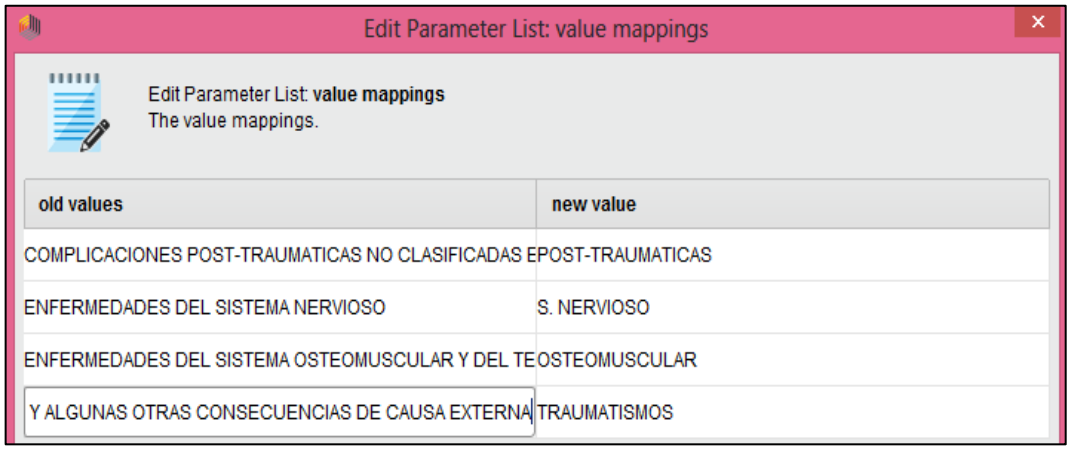


Figura 40 Conversión de nombres de diagnóstico

Con finalidad de mejorar la visualización, el árbol de decisión se muestra en secciones Figura 41, Figura 42, Figura 43 de acuerdo a atributo de mayor peso que es el tiempo de servicio (TS_MESES), seguido del Grupo Etario. En el ANEXO B se encuentra el árbol completo.

CONTINUA

La Figura 41, muestra el árbol de decisión correspondiente al tiempo de servicio menor e igual a 477,5 meses y grupos etario de entre 0-9 años y 10-19 años:

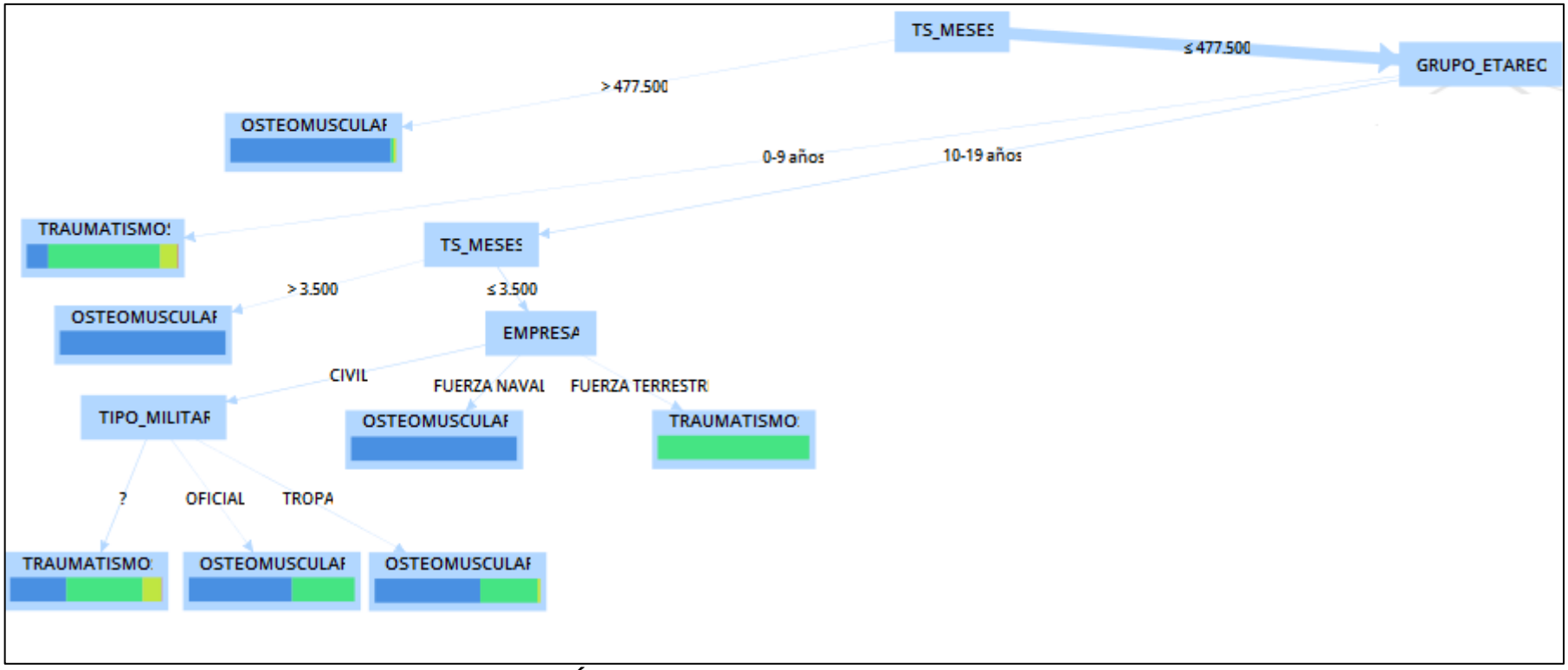


Figura 41 Árbol de decisión - ramificación izquierda

La Figura 42, muestra el árbol de decisión correspondiente al tiempo de servicio menor e igual a 477,5 meses y grupo etario de 20 a 39 años.

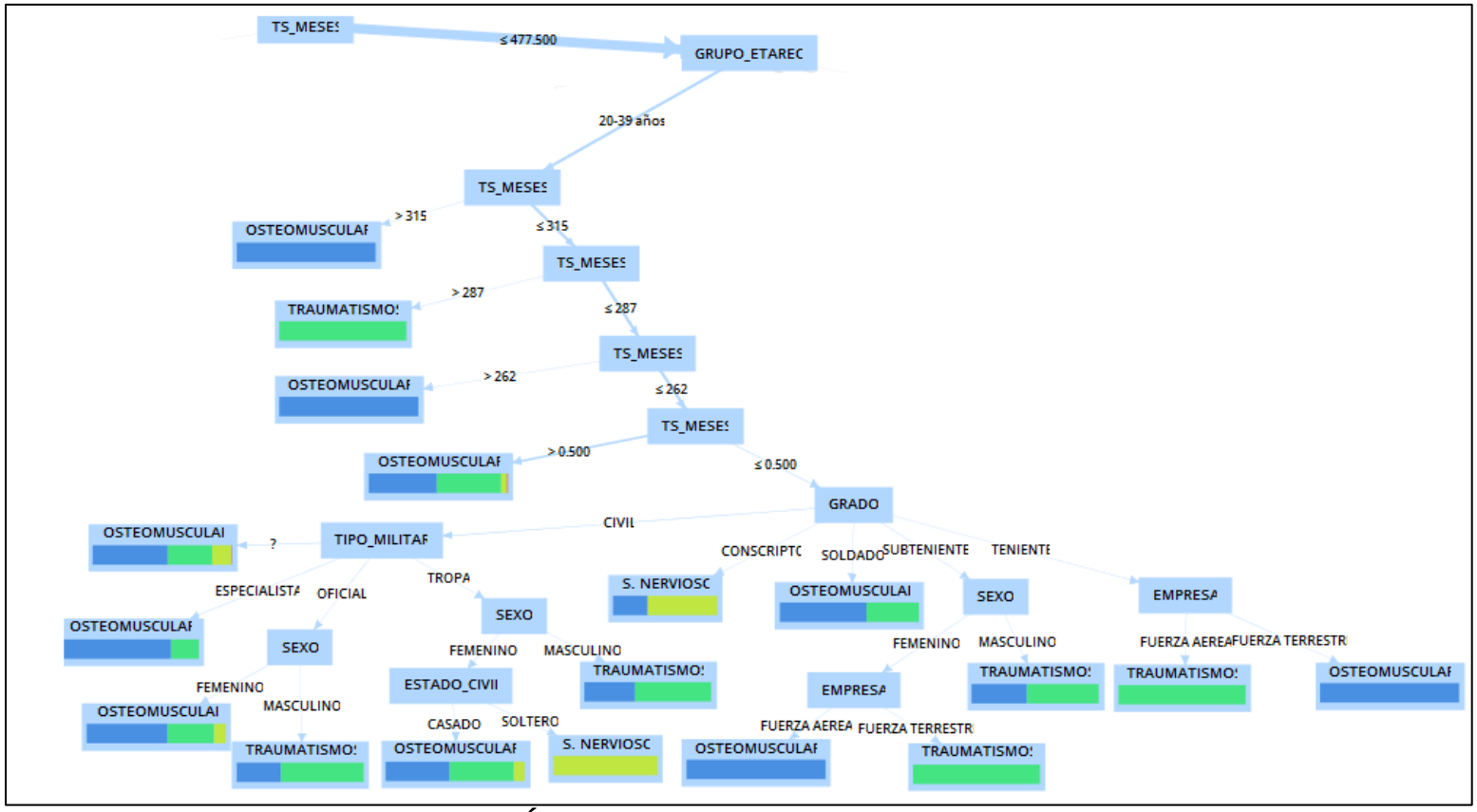


Figura 42 Árbol de decisión - ramificación central

La Figura 43, muestra el árbol de decisión correspondiente al tiempo de servicio menor e igual a 477,5 meses y grupo etario de 40-64 años y 85 años o más:

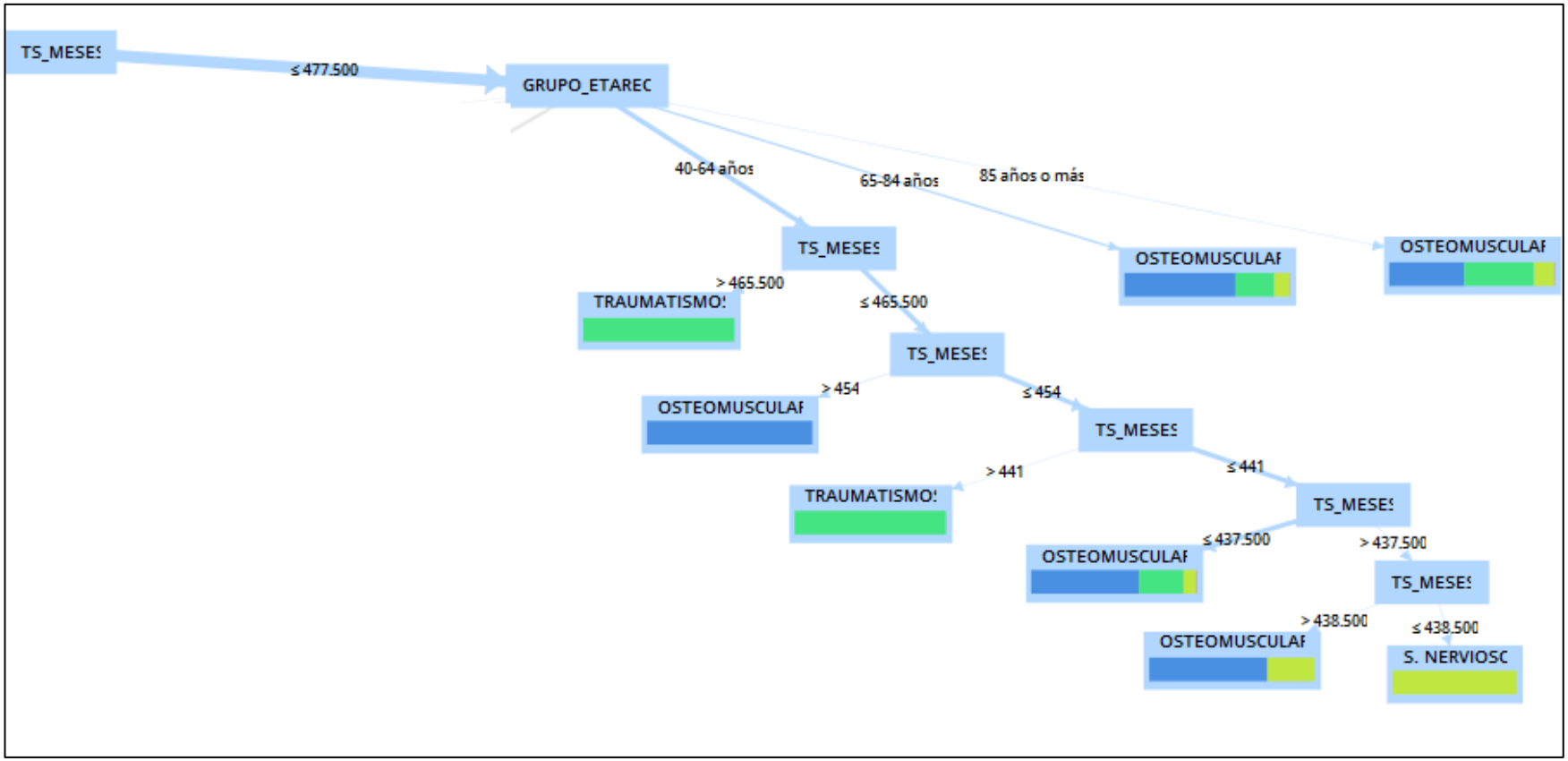


Figura 43 Árbol de decisión - ramificación derecha

En el ANEXO C, se encuentra el conjunto completo de reglas generadas por el árbol de decisión.

Los siguientes modelos se orientan a la predicción del valor total a pagar por servicio de salud, para lo cual, se define como el atributo clase al campo v_total.

3.4.3.3. Regresión Lineal

La Figura 44 muestra el modelo construido en RapidMiner, se establece un valor de tolerancia de 0,05. Este límite indica la presencia de una variable como combinación lineal de las restantes.

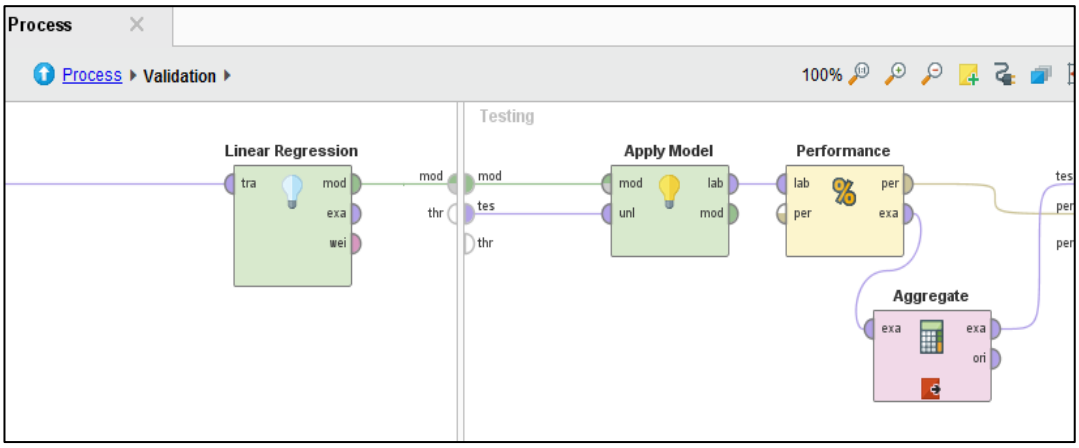


Figura 44 Construcción del modelo con regresión lineal
Fuente: Elaboración Propia

El modelo se ejecuta por servicio, resultando la matriz de la Figura 45 que contiene factores estadísticos como p valor y tolerancia que advierten que variables deben conservarse en la ecuación. En el ANEXO D,E,F y G se encuentran todas las matrices por los servicios restantes.

CONTINUA

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value
NO_PERSONA	0.000	0.000	0.008	0.982	1.959	0.050
C_CATEGORIA	-0.344	0.108	-0.019	0.991	-3.194	0.001
EDAD_FECHA	-0.040	0.014	-0.011	0.957	-2.906	0.004
NO_EMPRESA	-1.772	0.199	-0.029	0.977	-8.908	0
NO_GRADO	-0.242	0.044	-0.022	0.994	-5.547	0.000
NO_PRESTADOR	0.000	0.000	0.020	0.935	6.851	0.000
NIVEL_PRESTAD...	-19.976	0.213	-0.285	0.915	-93.704	0
ANIO	0.971	0.136	0.021	0.994	7.140	0.000
TS_MESES	0.011	0.003	0.017	0.998	3.299	0.001
DIAS_CONSULTA	1.281	0.081	0.045	0.999	15.756	0
(Intercept)	-1872.677	273.382	?	?	-6.850	0.000

Figura 45 Matriz de regresión lineal en exámenes y procedimientos

La Tabla 16, muestra las ecuaciones resultantes por servicio. Para conservar o descartar una variable en el modelo se ha considerado que p valor debe ser menor a 0,05 y la tolerancia debe ser aproximada a 1.

Tabla 16
Ecuaciones del método de regresión lineal

Tipo de Servicio	Ecuación Resultante
Exámenes y Procedimientos	+ 0.021NO_REG- 3.222C_CATEGORIA- 24.699NIVEL_PRESTADOR+ 1.106DIAS_CONSULTA- 1121.420
Hospitalización	+ 2.795EDAD_FECHA+ 39.607NO_EMPRESA+ 8.226NO_GRADO+ 0.002NO_PRESTADOR+ 460.721NIVEL_PRESTADOR+ 61.599ANIO+ 19.351DIAS_CONSULTA- 125963.087
Emergencia	- 0.010NO_REG- 0.053EDAD_FECHA+ 0.976NO_EMPRESA+ 7.687NIVEL_PRESTADOR- 11128.798
Reposición de gastos hospitalarios	- 4.693NO_GRADO- 0.322DIAS_CONSULTA+ 158.828
Atenciones Médicas por Consulta Externa	- 0.002C_CATEGORIA- 0.004NO_EMPRESA + 0.005ANIO- 5.912

Como se puede ver en la Tabla 16, las variables como número de prestador o el tiempo de servicio en meses de un afiliado no son influyentes para el modelo debido a que no se consideran en la ecuación, mientras que el número de días de consulta y el nivel del prestador aportan información valiosa.

La ecuación con mayor número de variables es la que corresponde al servicio de hospitalización, una de las causas es que en esta categoría se maneja altos montos de dinero dependiendo del nivel, lo que explica la alta influencia de la variable NIVEL_PRESTADOR sobre la ecuación resultante.

3.4.3.4. SVM

Para la construcción del modelo con el predictor SVM se establece la función núcleo “punto” definido como k (x,y) = x*y. La Figura 46 muestra el proceso de construcción del modelo.

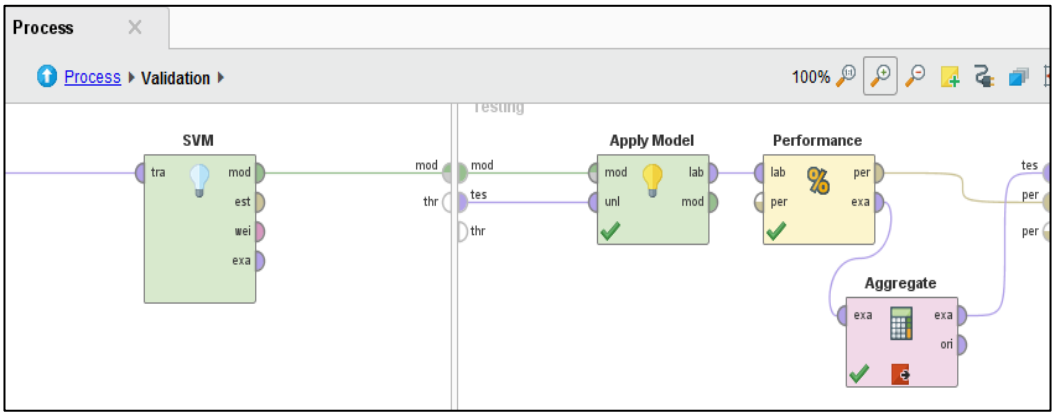


Figura 46 Construcción del modelo con SVM

El proceso se ejecuta por servicio, resultando el vector núcleo de la Figura 47 en donde se estipulan el valor de los pesos de las variables involucradas y sus sesgos, los cuales advierten que variables deben conservarse o descartarse en el modelo.

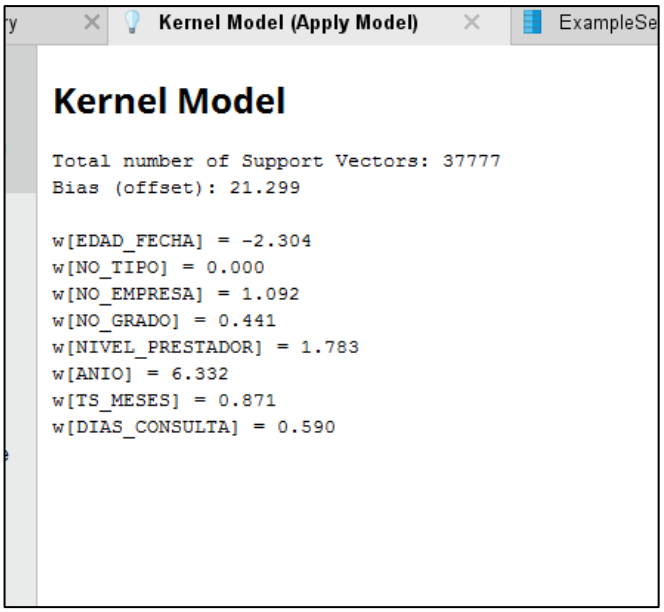


Figura 47 Vector núcleo resultante en exámenes y procedimientos

En la Figura 48 se visualiza la forma de separación de datos en el hiperplano.

CONTINUA

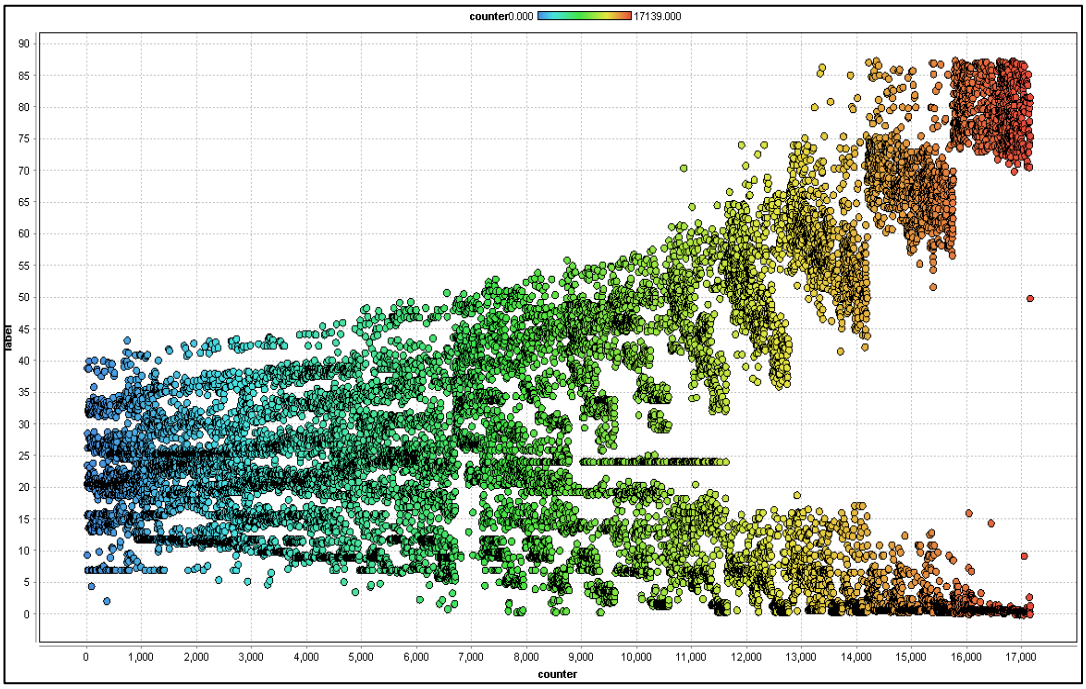


Figura 48 Clasificación de datos en el hiperplano

La Tabla 17, muestra los vectores resultantes por servicio. Para conservar o descartar una variable en el modelo se ha considerado que el valor de los pesos sea distinto de 0.

Tabla 17
Vectores resultantes de SVM

Servicio	Vector de Soporte	Sesgo
Exámenes y Procedimientos	w[EDAD_FECHA] = 89.03 w[NO_EMPRESA] = 0.051 w[NO_GRADO] = 6.32 w[NIVEL_PRESTADOR] = 100.276 w[TS_MESES] = 5.36 w[DIAS_CONSULTA] = 87.43	Total number of Support Vectors: 498 Bias (offset): 97.042
Hospitalización	w[EDAD_FECHA] = 11.121 w[NO_EMPRESA] = 7.044 w[NO_GRADO] = 8.397 w[NIVEL_PRESTADOR] = 130.216 w[ANIO] = 31.696 w[TS_MESES] = 6.553 w[DIAS_CONSULTA] = 140.662	Total number of Support Vectors: 8400 Bias (offset): 457.966
Emergencia	w[EDAD_FECHA] = 0.424 w[NO_EMPRESA] = 0.938	Total number of Support

	w[NO_GRADO] = -1.190 w[NIVEL_PRESTADOR] = 3.435 w[ANIO] = 8.576 w[TS_MESES] = -0.968 w[DIAS_CONSULTA] = 0.740	Vectors: 17140 Bias (offset): 24.932
Reposición de gastos hospitalarios	w[EDAD_FECHA] = -2.627 w[NO_EMPRESA] = 0.479 w[NO_GRADO] = -20.081 w[ANIO] = -3.660 w[TS_MESES] = -1.443 w[DIAS_CONSULTA] = -7.669	Total number of Support Vectors: 1622 Bias (offset): 49.670
Atenciones Médicas por Consulta Externa	w[EDAD_FECHA] = -0.002 w[NO_GRADO] = 0.001 w[NIVEL_PRESTADOR] = 0.311 w[ANIO] = 0.001 w[TS_MESES] = -0.002	Total number of Support Vectors: 56622 Bias (offset): 4.800

Como se muestra en la Tabla 17, las variables como no_persona y no_prestador fueron descartadas por no tener relevancia en el modelo, a diferencia del nivel, grado y días de consulta que tienen pesos considerables en los vectores resultantes.

3.4.3.5. Redes Neuronales

En la Figura 49 se observa el modelo construido en RapidMiner:

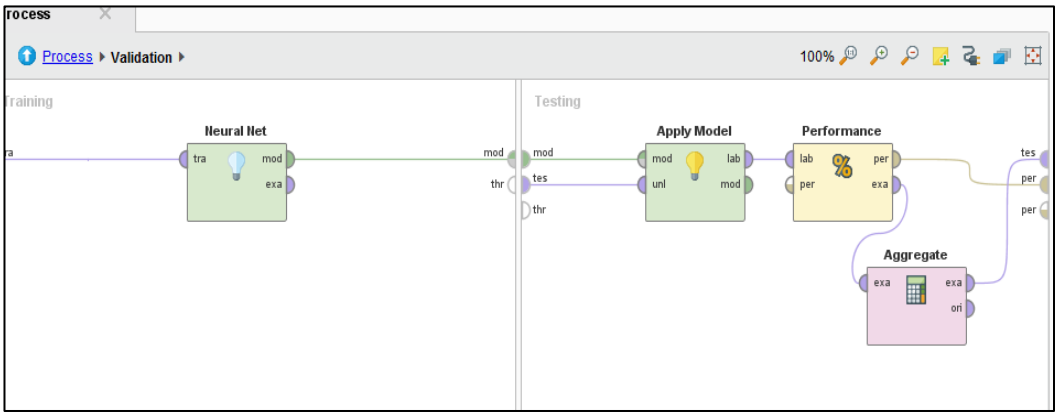


Figura 49 Modelo construido con redes neuronales

El resultado del modelo, es una red neuronal de 3 capas (ver Figura 50):

- La primera capa representa a los atributos de entrada seleccionados.
- La segunda contiene los pesos de los atributos para la capa de salida.
- La tercera capa muestra la neurona resultante con los valores esperados.

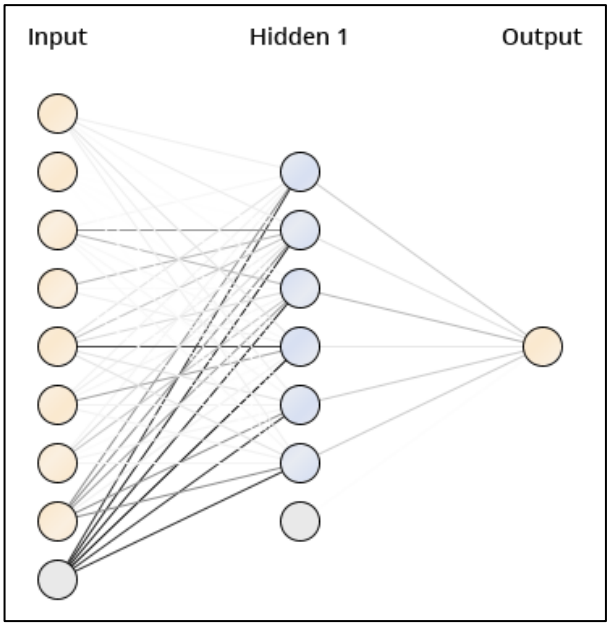


Figura 50 Capas de la red neuronal

Considerando que según estudios como (Huang, McGregor, & James, 2014) las redes neuronales son modelos complejos de interpretar, se tomará como indicadores a la tasa de error RMSE y la comparación del valor real con su predicción, obviando cualquier otra interpretación.

En la siguiente fase, se realiza la evaluación de este y los anteriores modelos construidos.

En el ANEXO J, se encuentran los archivos ejecutables de todos los modelos construidos.

3.5. Fase de Evaluación del modelo

Una vez construidos los modelos, se procede con la evaluación de resultados en función del cumplimiento de los objetivos del negocio y de minería de datos. Los modelos seleccionados serán los que cumplan con los criterios establecidos en la sección del Proceso de revisión.

3.5.1. Evaluación de resultados

En la Tabla 18, se contrasta el cumplimiento de los objetivos del negocio con las técnicas de minería de datos utilizadas.

Tabla 18
Validación de objetivos del negocio

Objetivos del negocio	Árboles de decisión	Regresión lineal	Support vector machine	Neural net
Determinar la combinación de factores que influye en la adquisición de una enfermedad músculo - esquelética en un afiliado	SI	NO	NO	NO
Identificar el pago de enfermedades músculo esquelético por tipo de servicio y el porcentaje de consumo a nivel del presupuesto.	NO	SI	SI	SI
Utilizar los resultados obtenidos del modelo como un insumo para la elaboración de planes preventivos.	SI	NO	NO	NO

Por otro lado, la Tabla 19, contrasta el cumplimiento de los objetivos de minería de datos con las técnicas utilizadas.

CONTINUA

Tabla 19
Validación de objetivos de minería de datos

Objetivos de minería de datos	Árboles de decisión	Regresión lineal	Support vector machine	Neural net
Clasificar la enfermedad del tipo Musculo esquelética a la que es propenso un afiliado de acuerdo a sus características como género, fuerza, categoría, grado, grupo etario.	SI	NO	NO	NO
Predecir el pago de enfermedades músculo esquelético facturado por servicios de salud.	NO	SI	SI	SI
Identificar las características de los afiliados en los que se paga mayor cantidad de dinero.	NO	SI	SI	SI

Se evidencia que todas las técnicas cumplen de forma parcial o total con los objetivos de minería de datos y del negocio.

3.5.2. Modelos aprobados

En esta etapa, las autoras seleccionan los modelos que satisfagan los siguientes criterios para la predicción del valor total:

a) Tasa menor del error RMSE [19] definida como una medida de desempeño para evaluar predicciones de valores numéricos (Barnston, 1992), su fórmula es:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}i - yi)^2}{n}}$$

[19] Raíz del error cuadrático medio o Root Mean Squared

Este error se calcula mediante la raíz del cuadrado de diferencia del valor real con el valor obtenido dividido para el número de muestras.

b) Costo computacional tomando como referente el tiempo de ejecución no mayor a 30 minutos.
c) Aproximación del valor real con su predicción.

Criterios para la evaluación del Árbol de Decisión:

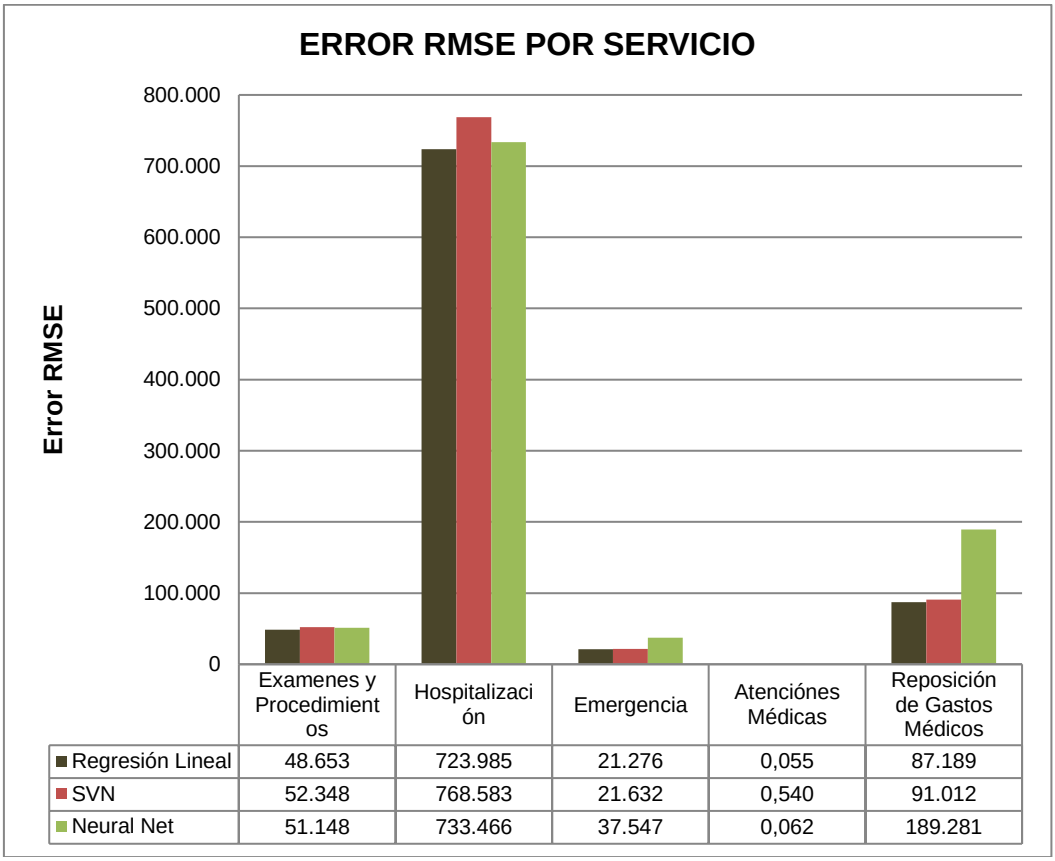
d) Porcentaje de precisión: Porcentaje de la clasificación correcta de los ejemplos en el grupo correspondiente. (RapidMiner Documentation, 2017)
e) Indicador Kappa: medida estadística que indica la proporción de la concordancia observada para elementos cuantitativos. (RapidMiner Community, 2017)

Considerando que el comportamiento de los datos varía por cada tipo de servicio la evaluación se realizará de forma individual.

TASA ERROR RMSE

En la Figura 51, se muestra un gráfico comparativo de la tasa de error RMSE presentada en cada modelo; en este criterio el modelo más acertado es Regresión Lineal.

CONTINUA



	Examenes y Procedimientos	Hospitalización	Emergencia	Atenciónes Médicas	Reposición de Gastos Médicos
Regresión Lineal	48.653	723.985	21.276	0,055	87.189
SVN	52.348	768.583	21.632	0,540	91.012
Neural Net	51.148	733.466	37.547	0,062	189.281

Figura 51 Error RMSE por servicio

Hospitalización es el servicio que tiene mayor tasa de error RMSE debido a la variabilidad de sus tarifas, al contrario de Atenciones médicas por consulta externa.

Con fines de investigación, se ejecutaron los mismos modelos utilizando el conjunto de datos previo al tratamiento de valores atípicos. En la Figura 52, se verifica el porcentaje de error RMSE incrementa en un 40% en relación a lo obtenido con valores limpios.

CONTINUA

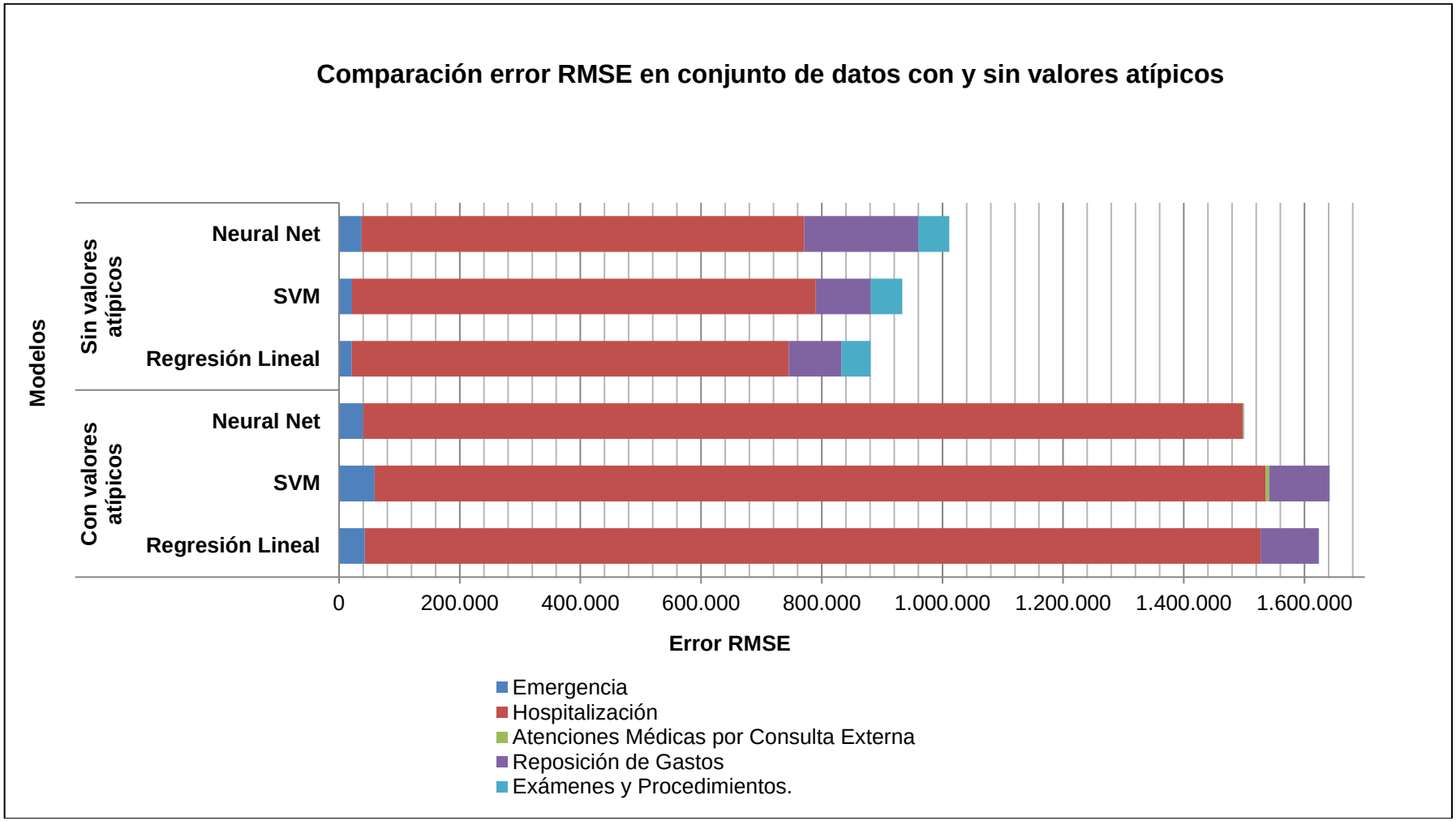
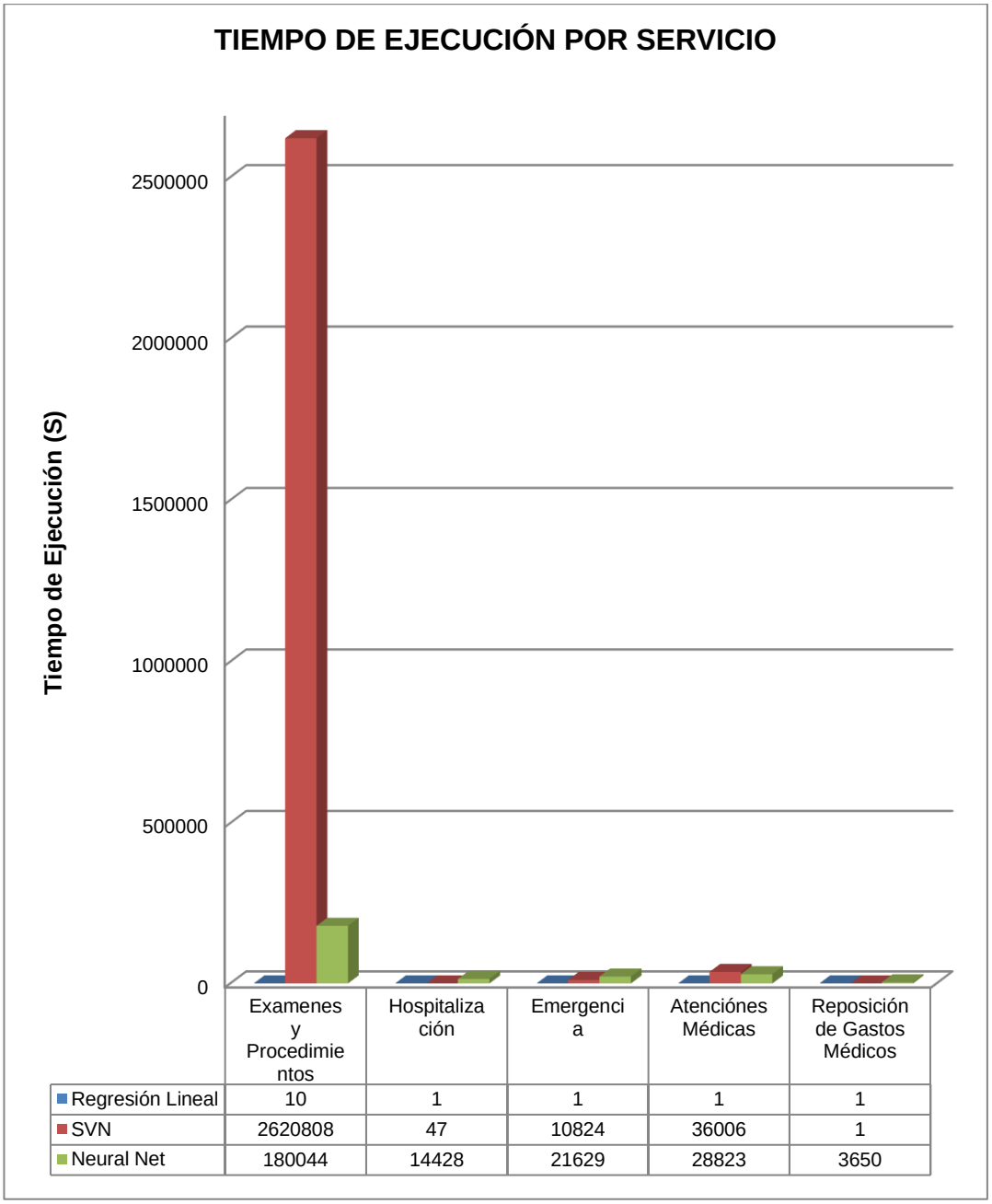


Figura 52 Comparación error RMSE con y sin valores atípicos

COSTO COMPUTACIONAL

Para evaluar el costo computacional se toma como referencia el tiempo de ejecución en segundos de cada modelo. En la Figura 53, se muestra la comparativa de este criterio, verificándose que Regresión Lineal tiene menor tiempo de ejecución.



	Examenes y Procedimientos	Hospitalización	Emergencia	Atenciónes Médicas	Reposición de Gastos Médicos
Regresión Lineal	10	1	1	1	1
SVN	2620808	47	10824	36006	1
Neural Net	180044	14428	21629	28823	3650

Figura 53 Tiempo de Ejecución por Servicio

En comparación de SVN que tiene mayor tiempo de ejecución, regresión lineal es 2600808 veces más rápido, siendo la técnica que satisface el segundo criterio de evaluación. Este es un indicio que el tiempo de ejecución de directamente proporcional al número de registros.

PREDICCIÓN DE VALORES

En este punto, la evaluación se realiza por tipo de servicio. A continuación se describen los resultados obtenidos.

En el servicio Exámenes y Procedimientos se puede apreciar (ver Figura 54), que Regresión lineal es la técnica cuya predicción se aproxima más al valor real en todos los años, a diferencia de Redes Neuronales[20]. Este patrón se repite en el servicio de Emergencia (ver Figura 55), en donde se presentan valores negativos que pueden ser indicios del sobreajuste del modelo en redes neuronales.

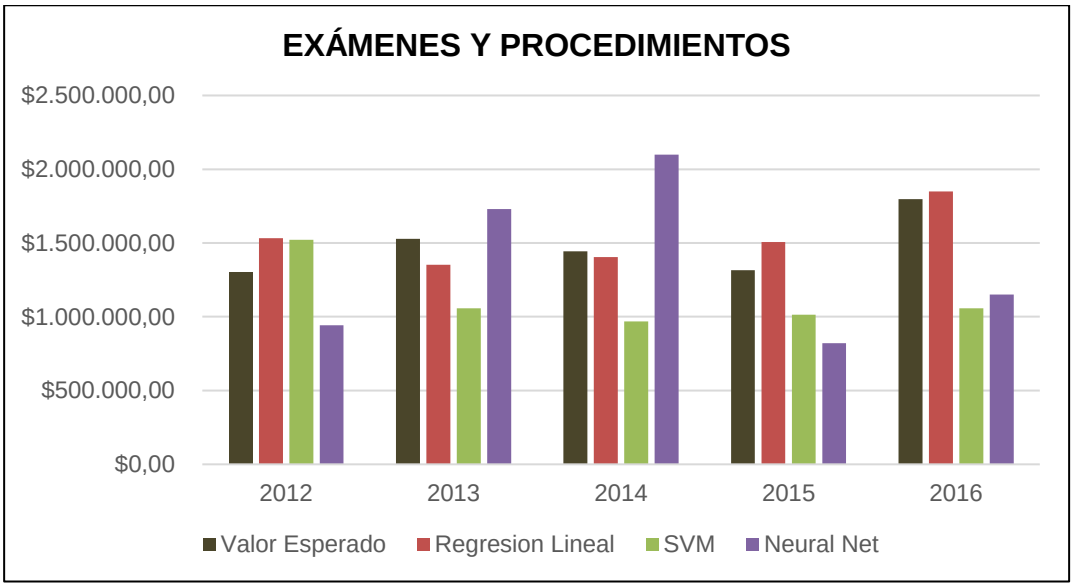


Figura 54 Valores predichos en Exámenes y Procedimientos

CONTINUA

[20] Neural Net

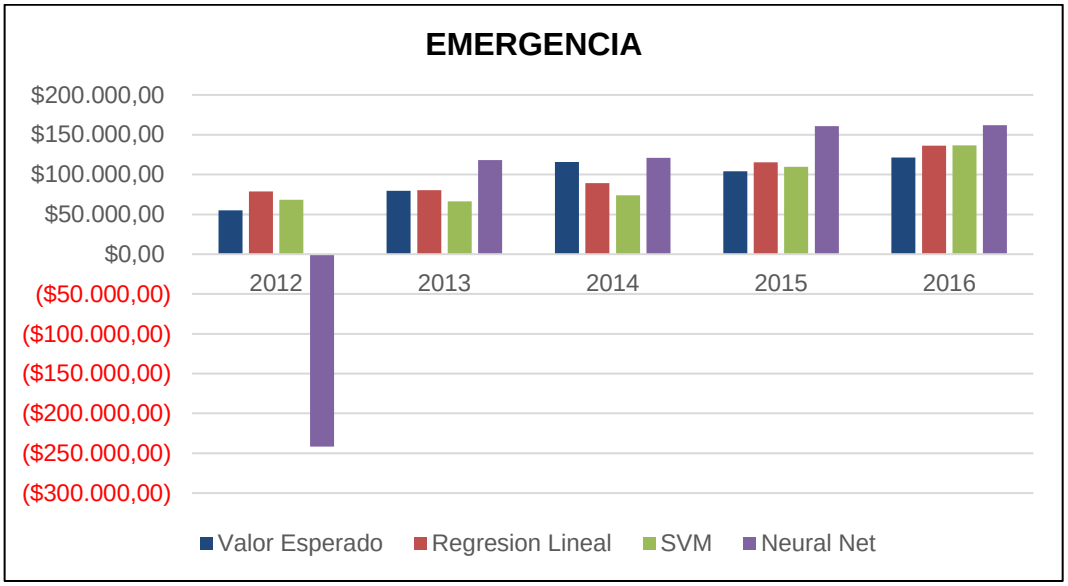


Figura 55 Valores predichos en Emergencia

Este comportamiento se repite para el servicio de Reposición de Gastos Hospitalarios, en donde se puede comprobar (ver Figura 56), que el modelo Neural Net genera una predicción negativa del valor, lo que no es consistente con el negocio, razón por la que es descartado.

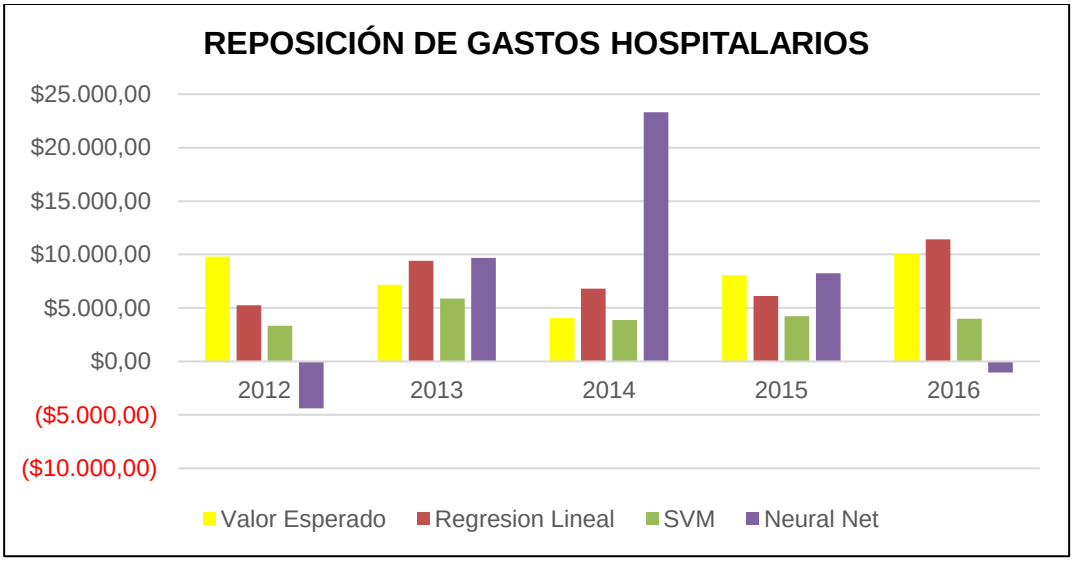


Figura 56 Valores predichos en Reposición de Gastos Hospitalarios

Para el servicio de Hospitalización, se puede apreciar (ver Figura 57), que Regresión lineal es la técnica cuya predicción se aproxima más al valor real en todos los años, a diferencia de SVM.

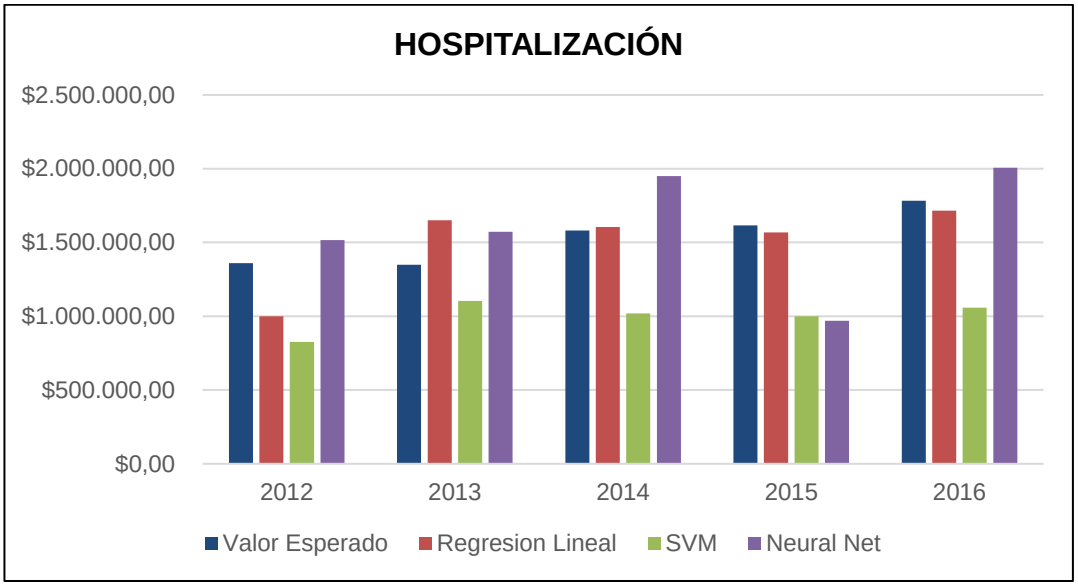


Figura 57 Valores predichos en Hospitalización

En cuanto a Atenciones Médicas por Consulta Externa, se evidencia que en los últimos tres años (ver Figura 58) todas las técnicas predicen valores aproximados a la realidad, sin embargo Regresión Lineal muestra menor diferencia.

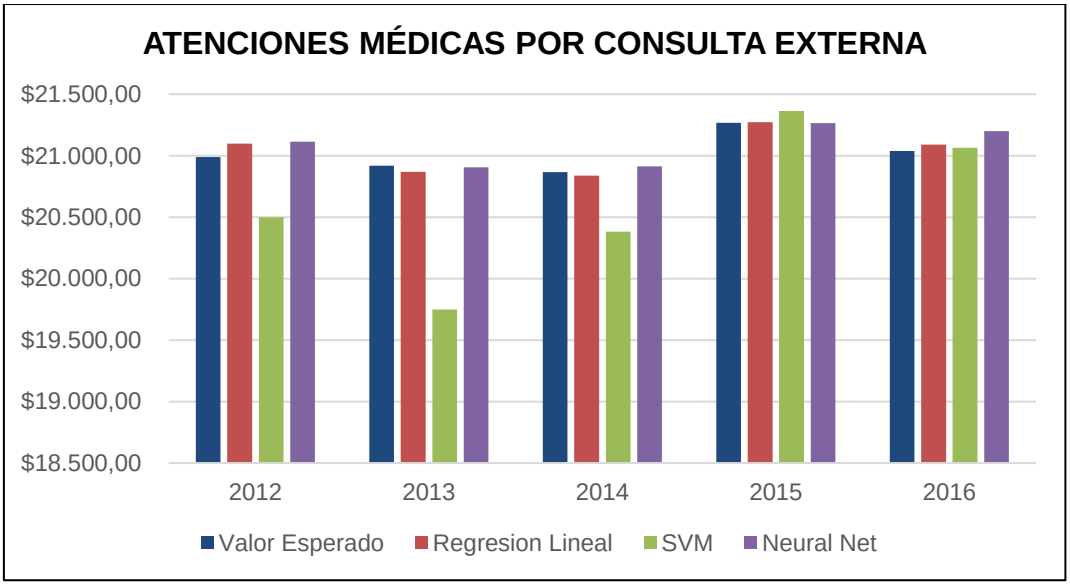


Figura 58 Valores predichos en Atenciones Médicas por consulta externa

CONTINUA

PORCENTAJE DE PRESICIÓN

Para la evaluación del Árbol de Decisión, se toma como indicador al porcentaje de precisión, definido por el número relativo de clasificaciones correctas de ejemplos. Una vez que se ejecuta el modelo con diferentes parametrizaciones, se obtiene el árbol con el mayor porcentaje de precisión (60.16%). La Figura 59 muestra el porcentaje de precisión por clase.

accuracy: 60.16% +/- 0.07% (mikro: 60.16%)

	true OSTEOMUSCULAR	true TRAUMATISMOS	true S. NERVIOSO	true POST-TRAUMATICAS	class precision
pred. OSTEOMUSCULAR	249022	131204	31586	1335	60.27%
pred. TRAUMATISMOS	15429	32225	6526	195	59.26%
pred. S. NERVIOSO	3	7	6	0	37.50%
pred. POST-TRAUMATICAS	0	0	0	0	0.00%
class recall	94.16%	19.72%	0.02%	0.00%	

Figura 59 Matriz de evaluación de precisión

INDICADOR KAPPA

Se interpreta el valor Kappa de la siguiente manera (RapidMiner Community, 2017):

- Kappa = 1 indica que todas las predicciones son correctas
- Kappa entre 0 y 1 indica que la mayoría de predicciones son correctas
- Kappa menor a 0 indica que todas las predicciones son incorrectas.

Como se aprecia en la Figura 60, el valor kappa es 0,001, por lo tanto cumple con el criterio de evaluación.

kappa: 0.001 +/- 0.000 (mikro: 0.001)

	true OSTEOMUSCUL...	true TRAUMATISMOS	true S. NERVIOSO	true POST-TRAUMATI...	class precision
pred. OSTEOMUSCU...	264449	163406	38059	1530	56.57%
pred. TRAUMATISMOS	3	29	0	0	90.62%
pred. S. NERVIOSO	2	1	59	0	95.16%
pred. POST-TRAUMAT...	0	0	0	0	0.00%
class recall	100.00%	0.02%	0.15%	0.00%	

Figura 60 Matriz de evaluación de indicador kappa

Por lo antes mencionado, se seleccionan al árbol de decisión y regresión lineal como los modelos óptimos.

3.6. Fase de despliegue

Esta fase tiene por objetivo, describir el proceso que permitirá al Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, incorporar y dar mantenimiento a los modelos construidos, así como utilizar los resultados de esta investigación en su proceso de toma de decisiones.

3.6.1. Plan de Despliegue

Los requerimientos para llevar a cabo el despliegue de los modelos son:

- Acceso al esquema “UPM” de base de datos del ISSFA.
- Acceso al esquema “MINERIA_DE_DATOS” de base de datos del ISSFA.
- Instalación del software RapidMiner Studio versión 7.0 o superior.
- Capacidad de memoria RAM de 8GB o superior.

Proceso

- Actualizar las tablas en el esquema de respaldo “MINERIA_DE_DATOS”
- Acceder a la información de la vista “V_PAGOS_SALUD”
- Ejecutar los modelos con la información de la vista.
- Realizar el análisis de los resultados obtenidos y llevar un registro de la tasa de error.

3.6.2. Plan de Mantenimiento y Monitoreo

Para el mantenimiento del modelo se debe considerar que el pago por servicios de salud se realiza mensualmente, razón por la cual, se debe actualizar la información del esquema “MINERIA_DE_DATOS” con esta misma periodicidad y monitorear los resultados obtenidos, verificando que la tasa de error sea mínima.

3.6.3. Reporte final

La elaboración del reporte final presenta los resultados obtenidos en la ejecución de los modelos y se muestran en el Capítulo 4.

CAPITULO 4: ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

En esta sección se realiza el análisis e interpretación de los resultados arrojados por los modelos. Cabe mencionar que todas suposiciones expresadas en este capítulo se fundamentan en el criterio de los expertos del negocio. A continuación se describen los hallazgos encontrados con respecto al tipo de enfermedad y pago por servicio de salud.

4.1. Análisis por tipo de enfermedad

El análisis por enfermedad satisface a los siguientes objetivos:

Objetivos del negocio:

- Determinar la combinación de factores que influyen en la adquisición de una enfermedad músculo -esquelética en un afiliado.
- Utilizar los resultados obtenidos del modelo como un insumo para la elaboración de planes preventivos.

Objetivos de minería de datos:

- Clasificar la enfermedad del tipo Músculo esquelética a la que es propenso un afiliado de acuerdo a sus características como género, fuerza, categoría, grado, grupo etario.
- Identificar las características de los afiliados en los que se paga mayor cantidad de dinero.

En cuanto a enfermedades musculo esqueléticas, de forma general se observa que en el periodo 2012 - 2016, el ISSFA paga mayores montos por Osteomusculares (ver Figura 61) Esta categoría consume 52% del pago, seguida de Traumatismos con 38%, enfermedades del Sistemas Nervioso con 10% y en menor proporción se encuentran las enfermedades Post traumáticas que no superan el 1%.

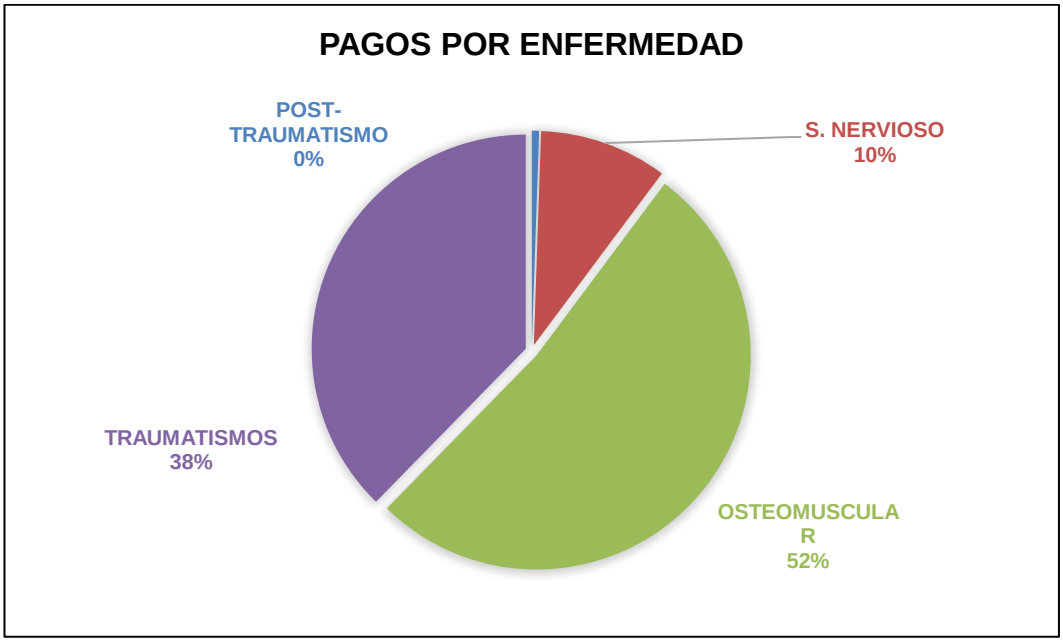
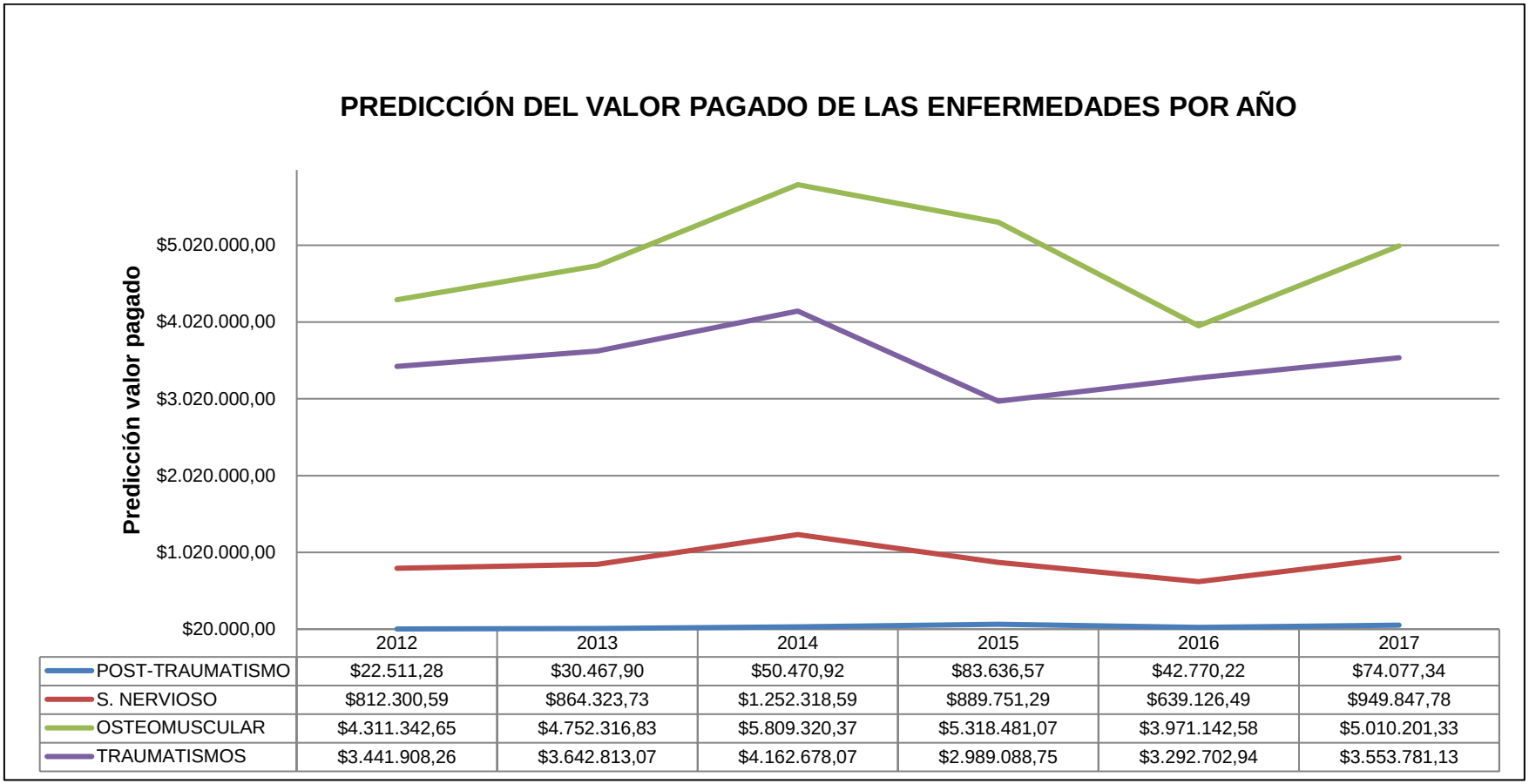


Figura 61 Pagos por Enfermedad musculo esquelética

Al estudiar el comportamiento anual de pagos (ver Figura 62), para el año 2017 se predice una variación incremental en todos los servicios de salud. Las enfermedades Osteomusculares siguen siendo las más costosas para el ISSFA superando los $5000000, a diferencia de las Post traumas que se encuentra por debajo del 1’000.000.

CONTINUA



	2012	2013	2014	2015	2016	2017
POST-TRAUMATISMO	$22.511,28	$30.467,90	$50.470,92	$83.636,57	$42.770,22	$74.077,34
S. NERVIOSO	$812.300,59	$864.323,73	$1.252.318,59	$889.751,29	$639.126,49	$949.847,78
OSTEOMUSCULAR	$4.311.342,65	$4.752.316,83	$5.809.320,37	$5.318.481,07	$3.971.142,58	$5.010.201,33
TRAUMATISMOS	$3.441.908,26	$3.642.813,07	$4.162.678,07	$2.989.088,75	$3.292.702,94	$3.553.781,13

Figura 62 Predicción del valor pagado en las enfermedades por año

A continuación se describe la combinación de características que influyen en el desarrollo de una enfermedad músculo – esquelética.

4.1.1. Enfermedad Osteomuscular

Según (Paredes Chicaiza, 2017), este tipo de padecimientos se producen por trabajar en condiciones inadecuadas, sus causa son directamente relacionadas con actividades como: actividades físicas que demandan fuerza excesiva, trabajos repetitivos, estar de pie o sentado en tiempos prolongados y cargar objetos pesados. Son las principales causas de ausentismo laboral, afectando claramente al desempeño físico del personal militar. Por lo tanto, deberían ser consideradas en programas de prevención de la salud ocupacional.

Los afiliados que son propensos a adquirir esta enfermedad cumplen con las siguientes características:

- Tiempo de servicio en la milicia mayor a 39 años.
- Subtenientes, tenientes, soldados y cabos segundos entre 18 y 19 años de edad con tiempo de servicio mayor a 3 meses y medio.
- Activos de la fuerza naval entre 18 y 19 años de edad.
- Afiliados de 20 a 39 años de edad con tiempo de servicio entre 21 a 23 años o con más de 26 años de servicio.
- Soldados de 20 a 39 años de edad con tiempo de servicio menor a 15 días.
- Mujeres subtenientes de la fuerza aérea entre 20 a 39 años de edad con tiempo de servicio menor a 15 días.
- Tenientes de la fuerza terrestre entre 20 a 39 años de edad.
- Afiliados entre 20 a 39 años de edad con tiempo de servicio menor a 15 días y que pertenecen a:
    - Especialistas
    - Mujeres oficiales
    - Mujeres casadas de tropa

- Afiliados y pensionistas entre 40 a 64 años de edad y tiempo de servicio de 36 a 38 años.

En la **Figura 63**, se muestra la predicción de pago en lesiones osteomusculares, la más costosa es dorsalgia[21] con el consumo del 22% del pago, seguida de trastornos en la rodilla, discos invertebrales, coxartrosis[22], gonartrosis[23] y finalmente lesiones del hombro.

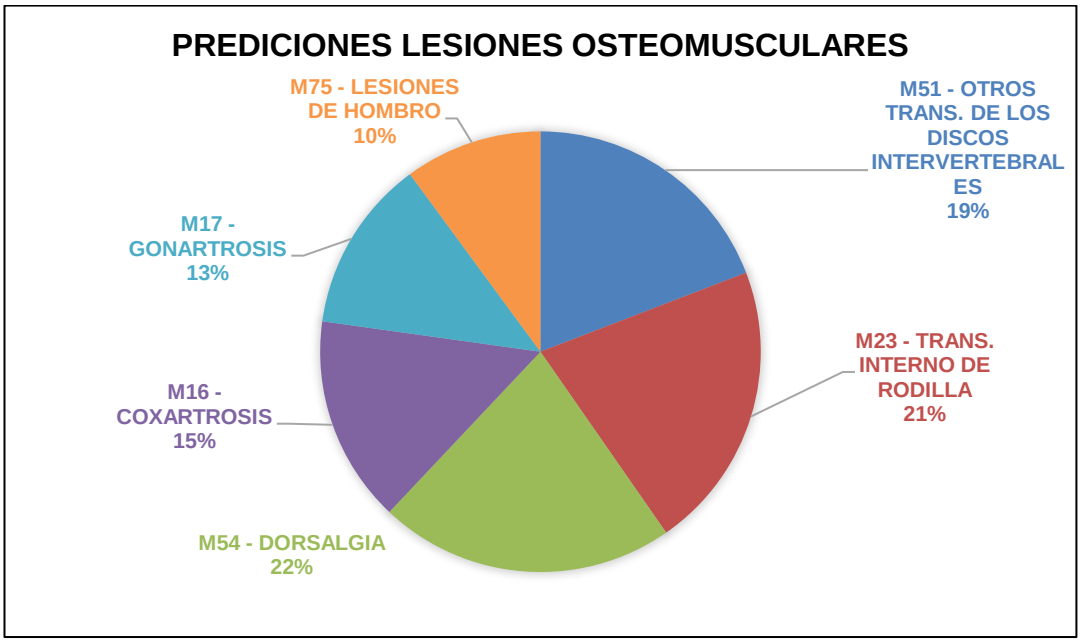


Figura 63 Predicciones lesiones osteomusculares

Según criterios de expertos del negocio, en el caso de pasivos, este tipo de lesiones se produce como resultado de la acumulación de desgaste físico durante la carrera militar, por el contrario para el personal activo joven son indicios de una inadecuada preparación previa a los ejercicios físicos, malos hábitos alimenticios e incluso podrían indicar que el proceso de reclutamiento tiene ciertas falencias.

[21] Dolor a nivel de la parte posterior de la espalda
[22] Dolor a nivel de la cadera
[23] Dolor crítico de la rodilla que puede implicar cirugía o incluso prótesis

4.1.2. Traumatismos

Se define a traumatismos como lesiones o daños de los tejidos orgánicos o de los huesos producidos por algún tipo de violencia externa, como un golpe, una torcedura u otra circunstancia, implica un daño físico que pueden derivar en complicaciones secundarias que ponen en riesgo la vida. (Wikipedia, 2017)

La combinación de características del personal propenso a adquirirlos son:

- Dependientes menores a 19 años de edad
- Personal activo de la Fuerza terrestre de 18 a 19 años de edad con tiempo servicio menor a 3.5 meses.
- Subtenientes y tenientes o soldados y cabos segundos, hombres de 20 a 39 años de edad.
- Afiliados con tiempo de servicio mayor a 38 años y dentro del grupo Etérea de 40 a 64 años

A mayor detalle, en la Figura 64, se presenta la predicción del pago por tipo de traumatismo, las Fracturas de pierna y tobillo son las más costosas y consumen un 22% del valor total y en menor proporción se encuentran las fracturas de hombro y brazo.

CONTINUA

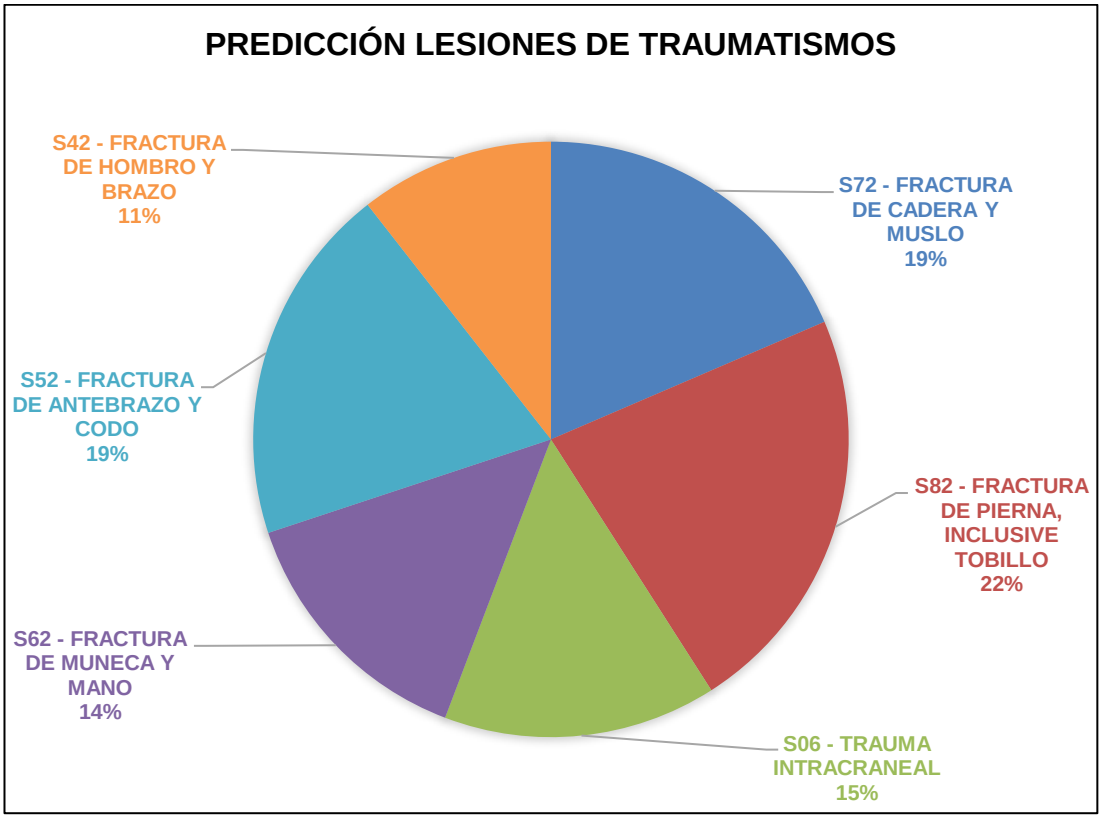


Figura 64 Predicción lesiones de Traumatismos

Según lo antes descrito, estos padecimientos se presentan con más frecuencia al inicio o fin de la carrera militar, esto podría ser causa del entrenamiento al que son sometidos en etapas iniciales de su profesión y las secuelas que estas provocan.

4.1.3. Enfermedad del sistema nervioso

Las enfermedades del Sistema nervioso están relacionadas con la coordinación de las funciones básicas del cuerpo, Los órganos principales del sistema nervioso, aparte del cerebro y la médula, son: Los ojos, los oídos, los órganos del gusto, los órganos del olfato y los receptores sensoriales de la piel.

Las características de afiliados más propensos a adquirir enfermedades del sistema nervioso son:

- Conscriptos entre 20 y 39 años de edad.

- Mujeres solteras de tropa de 20 a 39 años de edad y tiempo de servicio menor a 15 días.
- Afiliados de 40 a 64 años de edad y con menos de 36 años de tiempo de servicio en las Fuerzas Armadas.

Para mayor detalle en la Figura 65, se observa que el valor de predicción de pago para el año 2017, la lesión más costosa corresponde a hernias discales o compresiones de raíces y plexos nerviosos con un 40% del pago, lo que se relaciona con cargas de peso excesivo, dando indicios de que para la ejecución de entrenamientos del personal militar no se considera su tamaño y contextura.

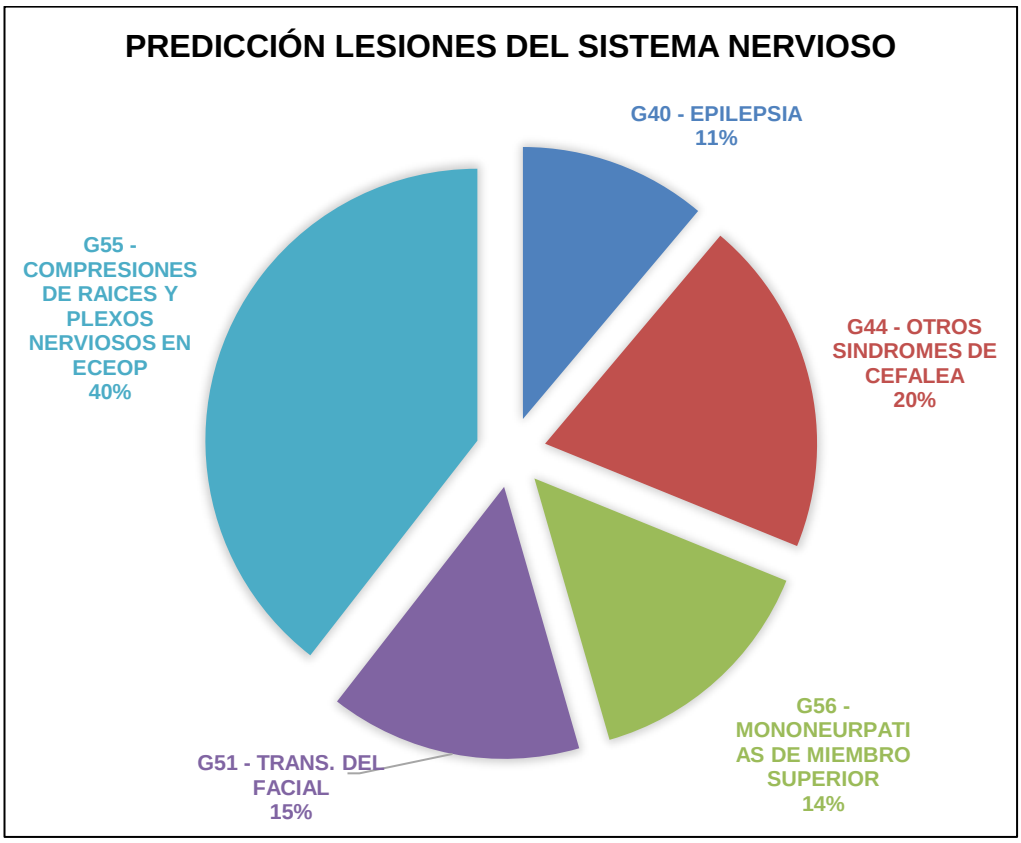


Figura 65 Predicción lesiones del sistema nervioso

CONTINUA

4.1.4. Enfermedades post-traumáticas

Son enfermedades que se producen como consecuencia de golpes, accidentes o enfermedades preexistentes. Estas enfermedades se hallan en menor frecuencia en afiliados y pensionistas mayores a 65 años.

La Figura 66, muestra a mayor detalle el valor de predicción de pago para año 2017, las secuelas del trauma de miembro inferior, son las más costosas para el ISSFA.

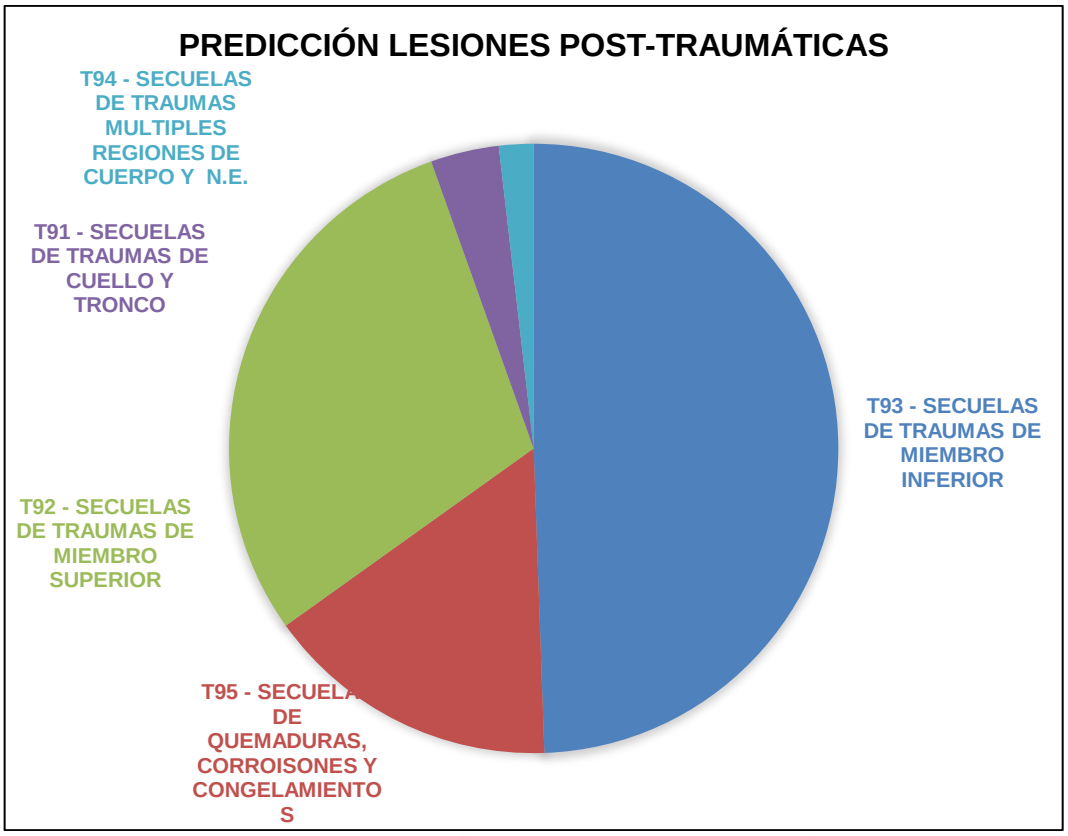


Figura 66 Predicción lesiones post-traumáticas

Las secuelas de traumas de miembro inferior podrían ser consecuencias de operaciones en las que no existieron terapias de rehabilitación adecuadas, ocasionando que la persona no se recupere al 100%.

En el **ANEXO J,** se encuentra una tabla específica con la combinación de características de los afiliados propensos a adquirir una enfermedad post-traumática.

4.2. Análisis por tipo de servicio

El análisis por tipo de servicio satisface a los siguientes objetivos:

Objetivos del negocio:

- Identificar el pago de enfermedades músculo esquelético por tipo de servicio y el porcentaje de consumo a nivel del presupuesto.

Objetivos de minería de datos:

- Predecir el pago de enfermedades músculo esqueléticas facturadas por servicios de salud.

El patrón más evidente es la disminución del pago en el año 2015 con respecto al 2014 en todos los servicios, lo que se explica por las nuevas definiciones de precios estipuladas en el Tarifario de Prestaciones del Sistema Nacional aplicado por el ISSFA en el mes de abril del 2015 por normativas del Ministerio de Salud.

A continuación se muestra el análisis realizado por tipo de servicio:

4.2.1. Exámenes y Procedimientos.

La Figura 67, muestra el pago realizado por el ISSFA en Exámenes y Procedimientos desde el año 2012 al 2016, incluyendo la predicción de pago para el 2017. Este servicio genera un pago mayor en el año 2016, mientras que la predicción para el año 2017 baja en un 4%. Es importante mencionar, que al mes de julio del 2017 se ha pagado un 63% del valor predicho. Según criterio del negocio, una de las causas en la disminución del pago es la implementación de un nuevo Tarifario Nacional en Febrero 2017 en prestadores de salud Militar, lo que implica la disminución de precios en estos servicios

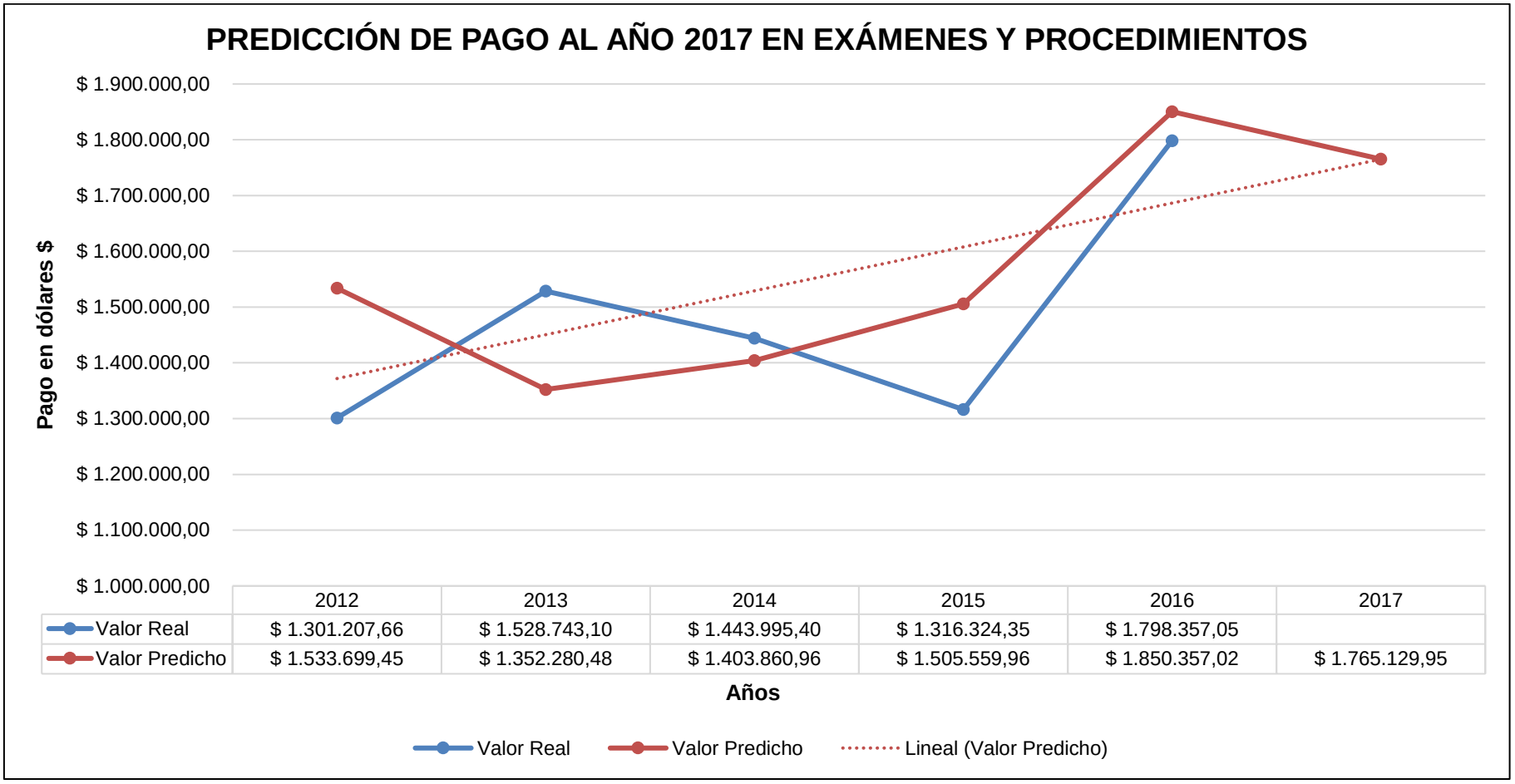


Figura 67 Predicción de pago al año 2017 en Exámenes y procedimientos

En relación al presupuesto de Exámenes y Procedimientos, se observa (ver Figura 68) que no ha sufrido un impacto mayor al 17% en los últimos 3 años y su predicción al año 2017 será de 16%. En otras palabras, habrá un impacto mínimo en el presupuesto de este servicio.

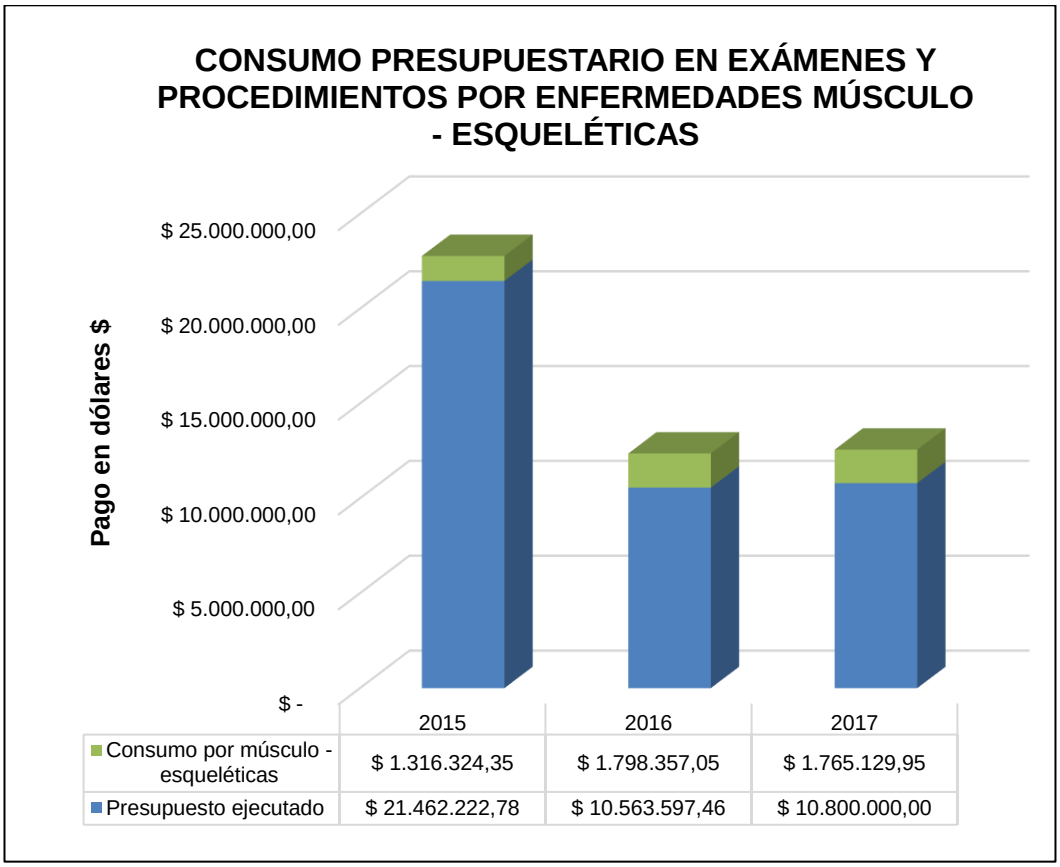


Figura 68 Consumo presupuestario en Exámenes y procedimientos

Según criterios de expertos del negocio, lo antes mencionado podría indicar que:

- Las personas no tienen una cultura de prevención de lesiones, es decir el ingreso por emergencia es el disparador para la realización de exámenes y procedimientos.
- En cierta forma, se están aplicando los controles estipulados en capítulo VII de la Norma Técnica Sustitutiva de Relacionamiento emitida por el Ministerio de Salud en donde se establece la

responsabilidad de pago debe ser compartida por las aseguradoras privadas dependiendo de su porcentaje de cobertura. (Ministerio de Salud Pública del Ecuador, 2017)

A mayor nivel de detalle, con respecto a la población de activos y pasivos, se evidencia que en el año 2016 el valor pagado en personal activo triplica al valor del pago en pasivo, mientras que para el año 2017 el incremento será únicamente del 27% (ver Figura 69). Este patrón podría deberse a la cantidad de exámenes que el personal militar se realiza previo a rendir las pruebas físicas periódicas propias de su evaluación de carrera.

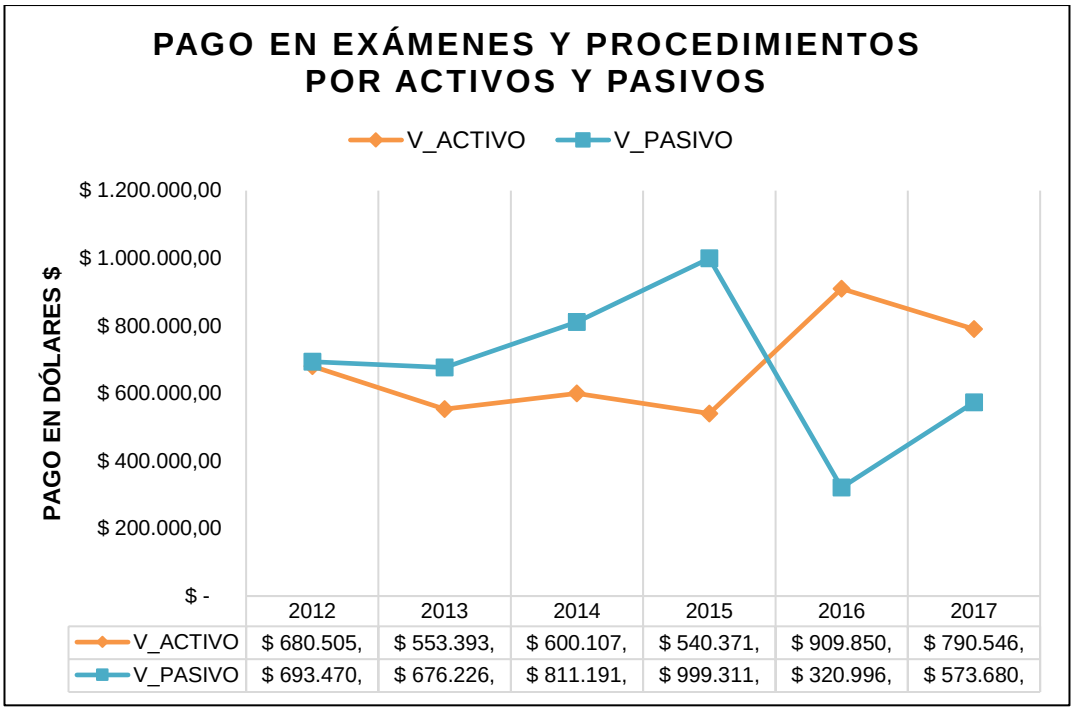


Figura 69 Pago en Exámenes y procedimientos por activos y pasivos

En cuanto a la fuerza, se predice que el pago destinado a los miembros de la fuerza terrestre incrementará en un 10% con respecto al año anterior, en la fuerza naval baja en 4% y en la fuerza aérea se mantiene el valor. (ver Figura 70 y Figura 71).

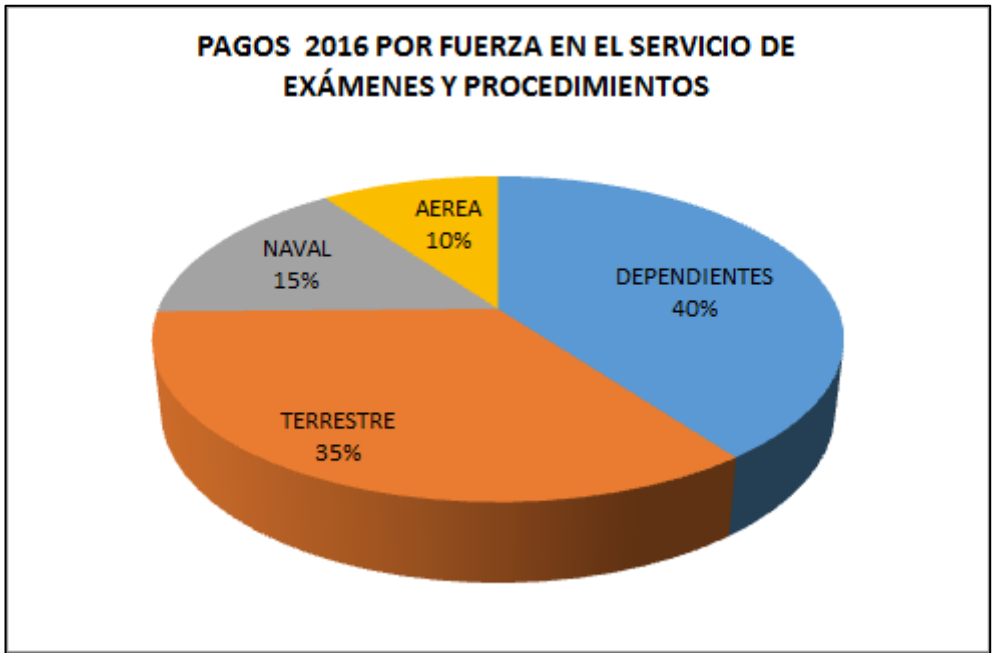


Figura 70 Pagos 2016 por fuerza en Exámenes y procedimientos

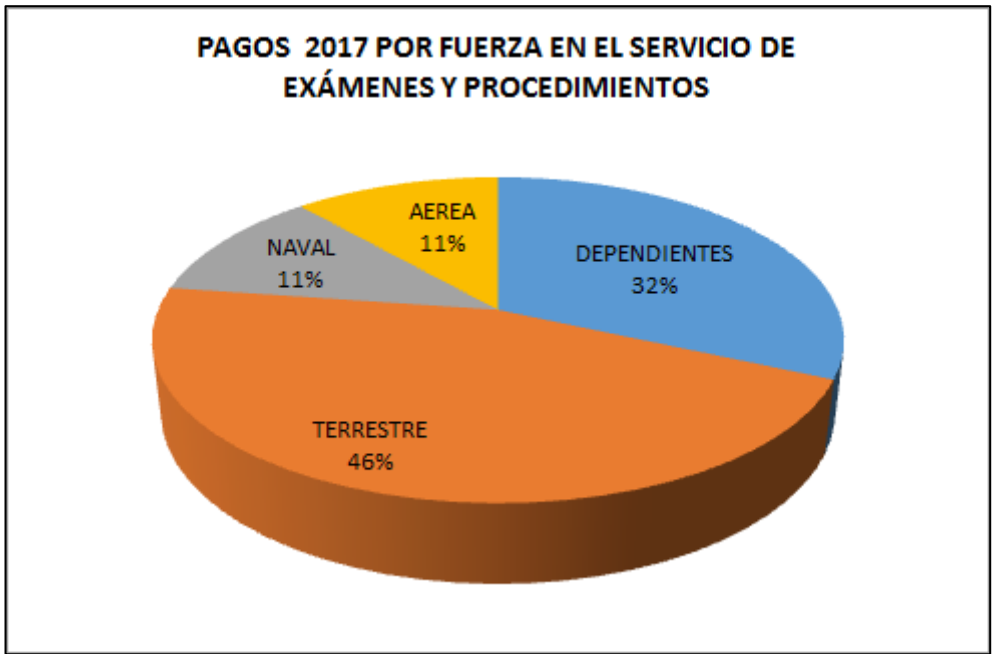


Figura 71 Pagos 2017 por fuerza en el servicio de Exámenes y procedimientos

Este hecho se puede atribuir a que existe una mayor cantidad de afiliados que pertenecen a la fuerza terrestre.

4.2.2. Hospitalización

Este servicio está orientado a proporcionar cuidados básicos y especializados. Desarrolla actividades médicas y de enfermería encaminadas al tratamiento y recuperación de la salud del paciente, ofreciendo la prestación de los servicios con recurso humano calificado, comprometido y humanitario. (Clínica Sebastián de Benálcazar, 2017)

El servicio de Hospitalización tiene una tendencia incremental desde el año 2013 (ver Figura 72). Al mes de julio se ha consumido el 52% del presupuesto, se predice que a final de 2017 existirá un incremento de 11% respecto al año anterior siendo a la vez el valor pagado más alto en los últimos cinco años.

CONTINUA

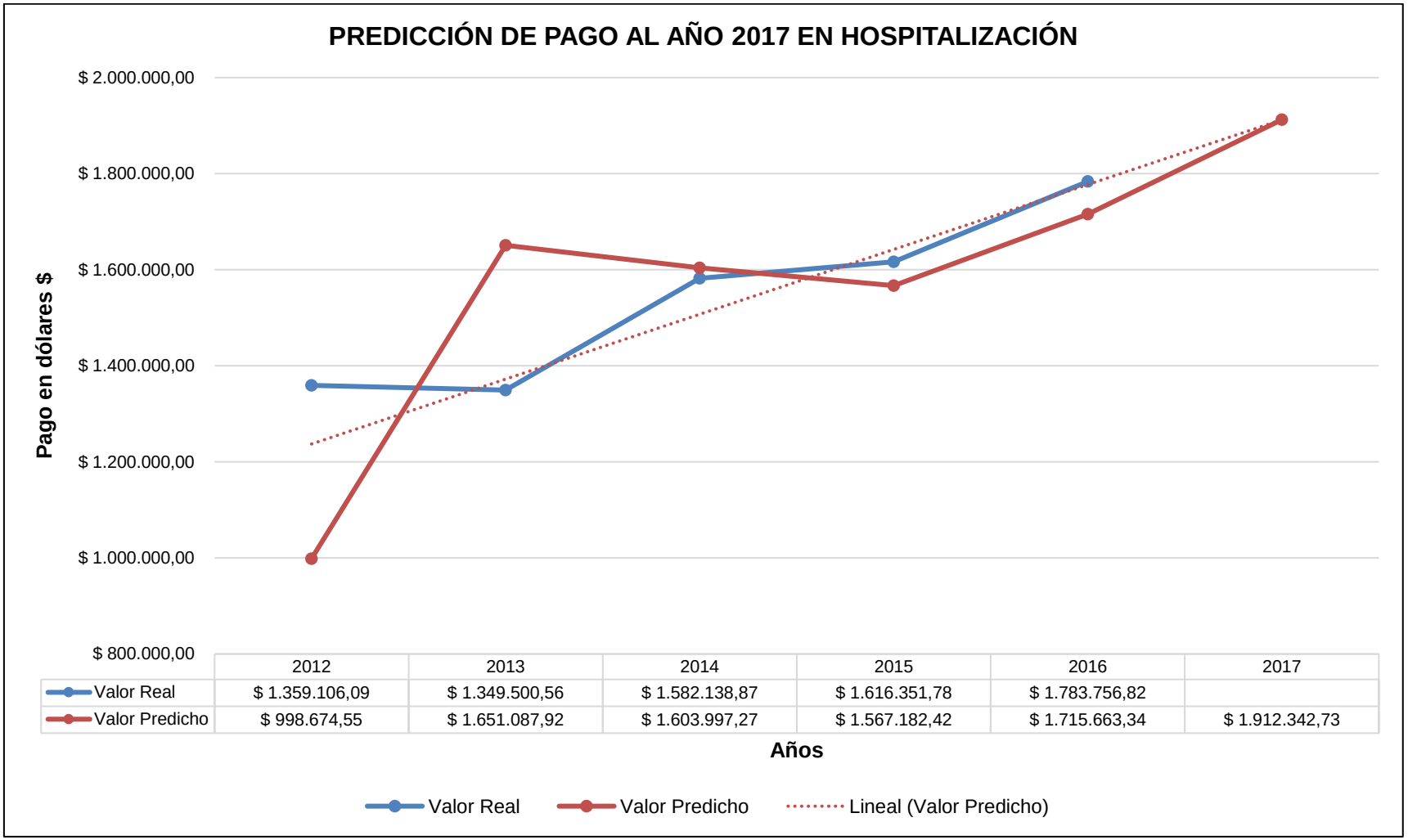


Figura 72 Predicción de pago al año 2017 en Hospitalización

Para más información, la Tabla 20, muestra el porcentaje de consumo en el presupuesto de los últimos 3 años, se evidencia un patrón incremental, de tal forma que en el año 2017 se consumirá un 12,83% del presupuesto, el doble del año 2015

Tabla 20

Consumo del presupuesto en enfermedades músculo esqueléticas

Año	Valor por músculo esqueléticas	Porcentaje pago por músculo esqueléticas
2015	$ 1.616.351,78	5,63
2016	$ 1.783.756,82	7,30
2017	$ 1.912.342,75	12,83

El incremento del consumo presupuestal de este servicio podría deberse a que una alta facturación por prestadores del ISSFA:

En cuanto al pago por la categoría del afiliado, en la Figura 73 se puede apreciar que en los dos últimos años el pago en hospitalizaciones ha subido en 14% para los activos y ha disminuido en un 5% para la población en servicio pasivo. En otras palabras para finales del 2017 el ISSFA habrá destinado a activos, el doble de recursos económicos de los pasivos, lo que se podría explicar debido a la cantidad de operaciones a las que se someten los afiliados con la finalidad de mantenerse en buena condición para continuar ejecutando actividades de alto riesgo, propias de su profesión.

CONTINUA

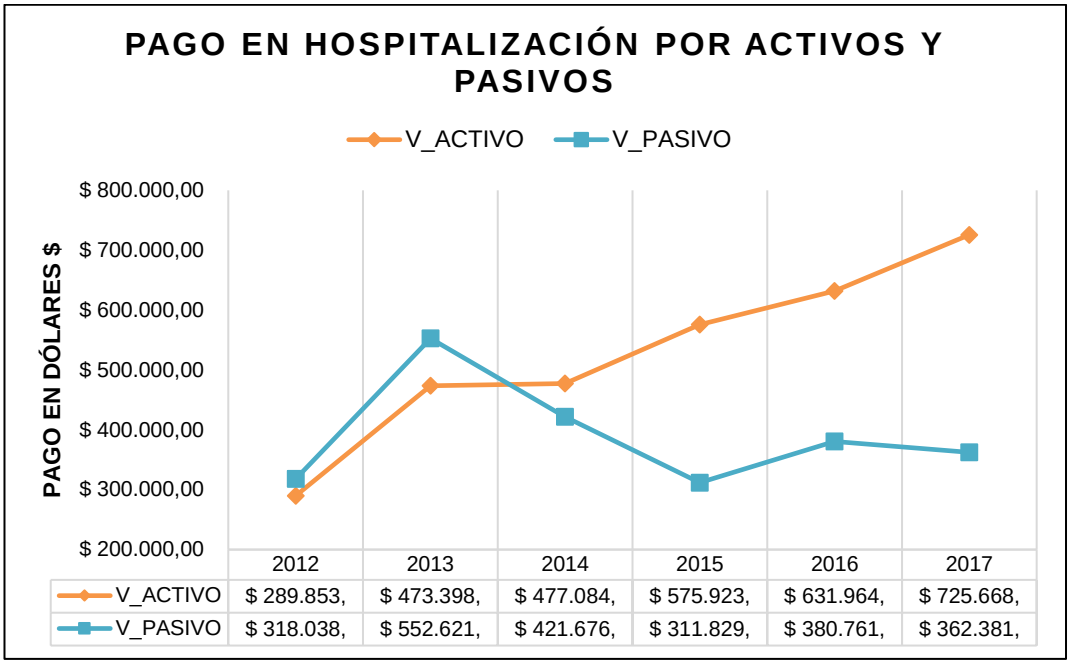


Figura 73 Pago en Hospitalización por activos y pasivos

En referencia al pago por fuerza el modelo predice que la proporción del pago se mantendrá (ver Figura 74 y Figura 75).

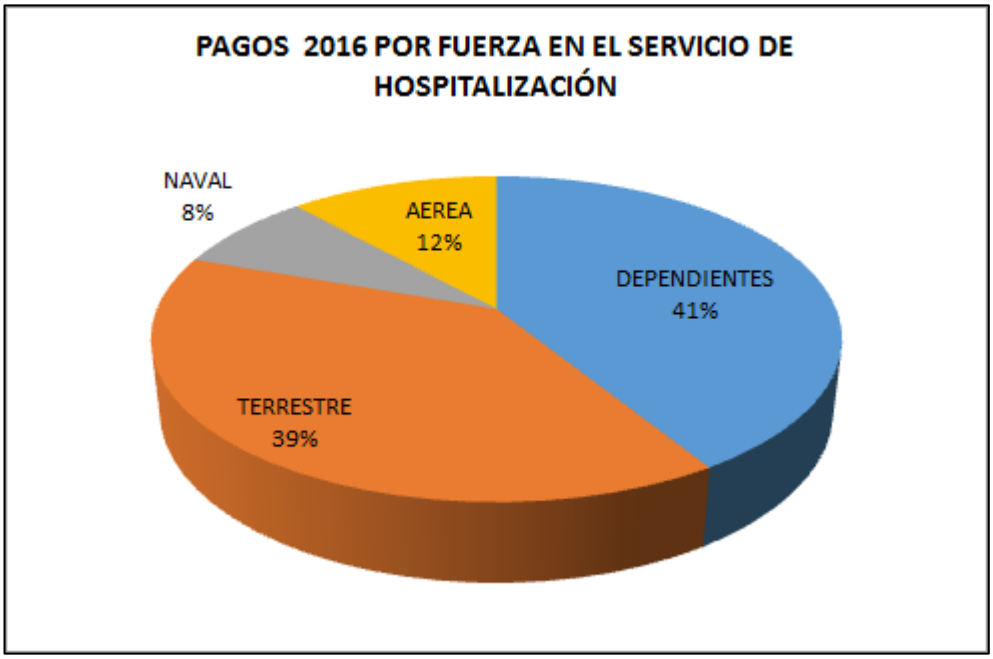


Figura 74 Pagos 2016 por fuerza en el servicio de Hospitalización

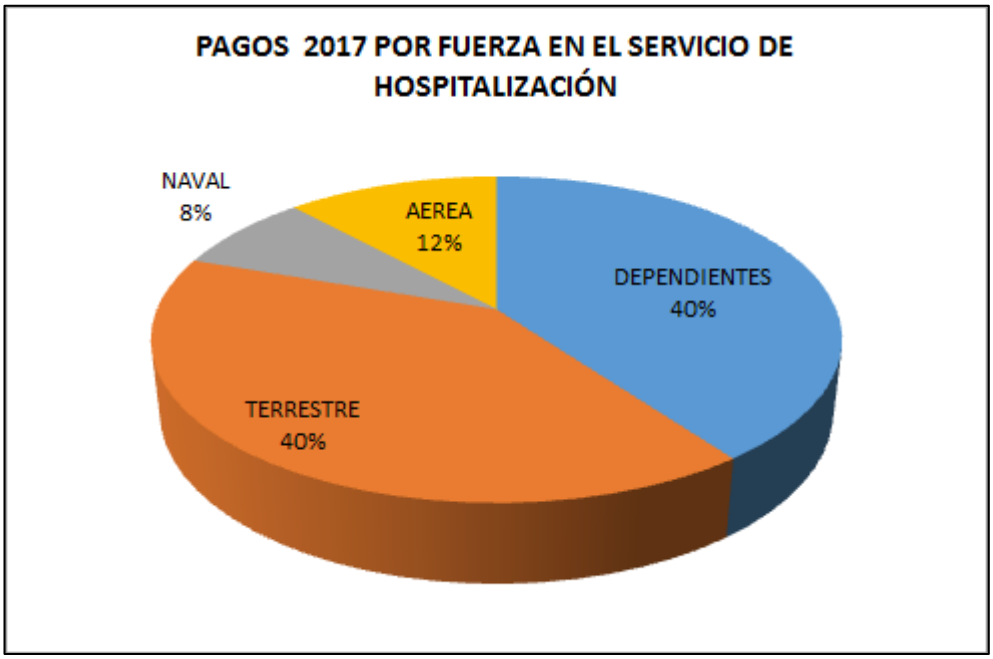


Figura 75 Pagos 2017 por Fuerza en el Servicio de Hospitalización

4.2.3. Emergencia

El servicio de Emergencia contempla toda contingencia de gravedad que afecte a la salud del ser humano con inminente peligro para la conservación de la vida o de la integridad física de la persona, como consecuencia de circunstancias imprevistas e inevitables, independientemente del lugar de su acontecimiento. (Salud S.A., 2017)**.** Según el usuario experto en auditoría médica, una emergencia se define como la condición de una persona en la que está en riesgo su vida o algún signo vital.

Al igual que el servicio de Hospitalización, los pagos en el servicio de Emergencia muestran una tendencia incremental. Al finalizar el año 2017 (ver Figura 76) la predicción indica un incremento del 84% en comparación al 2012. En el año 2017 se proyecta el valor pagado más alto en los últimos 6 años. Según lo que se predice en el año 2017 se pagará un 6% más que el año 2016 en patologías músculo – esqueléticas. Esta tendencia indica un incremento en el número de enfermedades músculo esqueléticas, lo que según el criterio de expertos del negocio, podría considerarse como una alerta en la forma del entrenamiento al que está sujeto el personal militar.

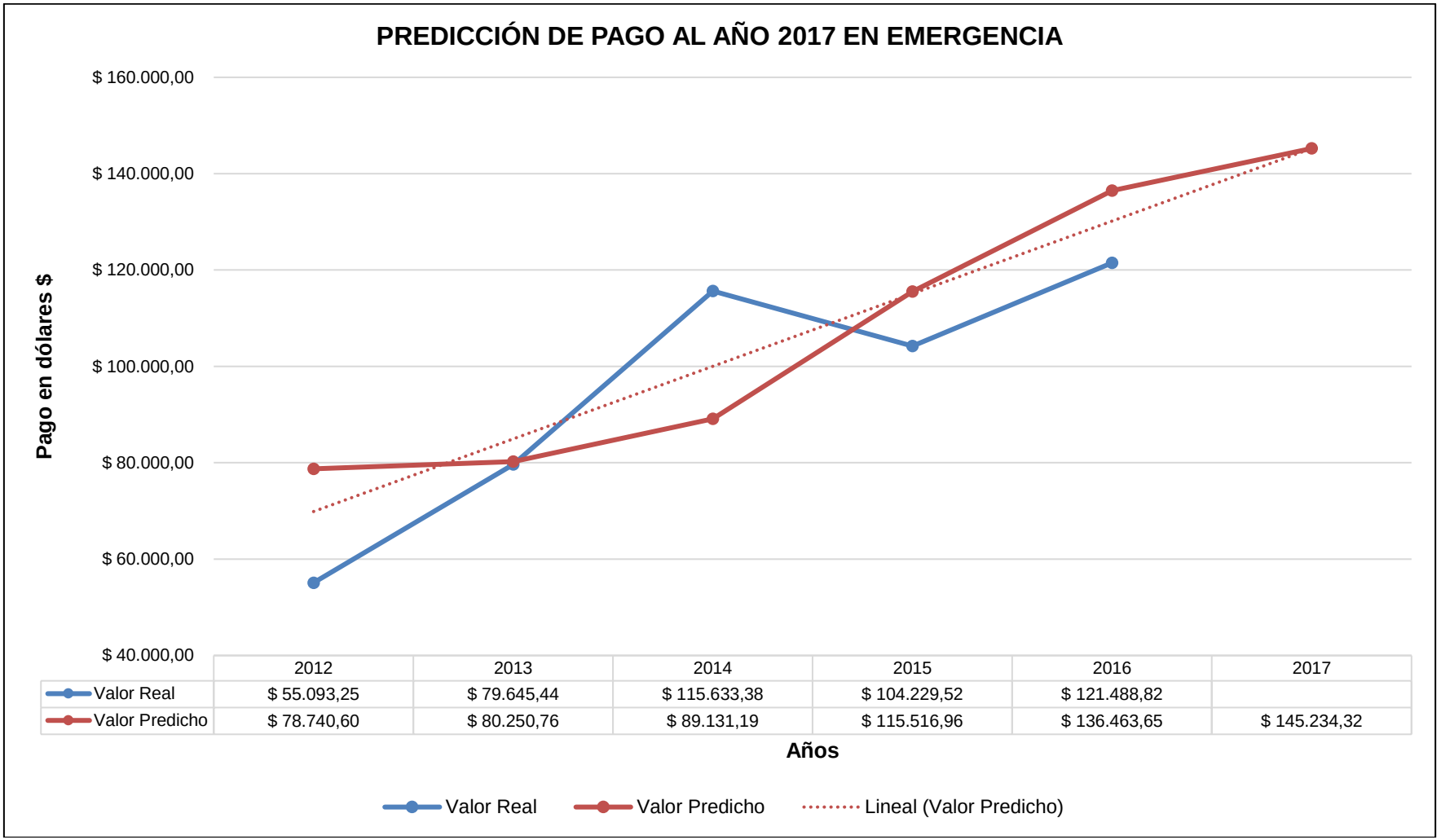


Figura 76 Predicción de pago al 2017 en Emergencia

Se observa en la Figura 77 que en al año 2016, el consumo por enfermedades músculo esqueléticas no supera al 1% de lo presupuestado, sin embargo el año 2017 incrementa al 14,35%. En base a este número significativo, Dirección de Salud debería destinar un monto mayor en este servicio para los años siguientes.

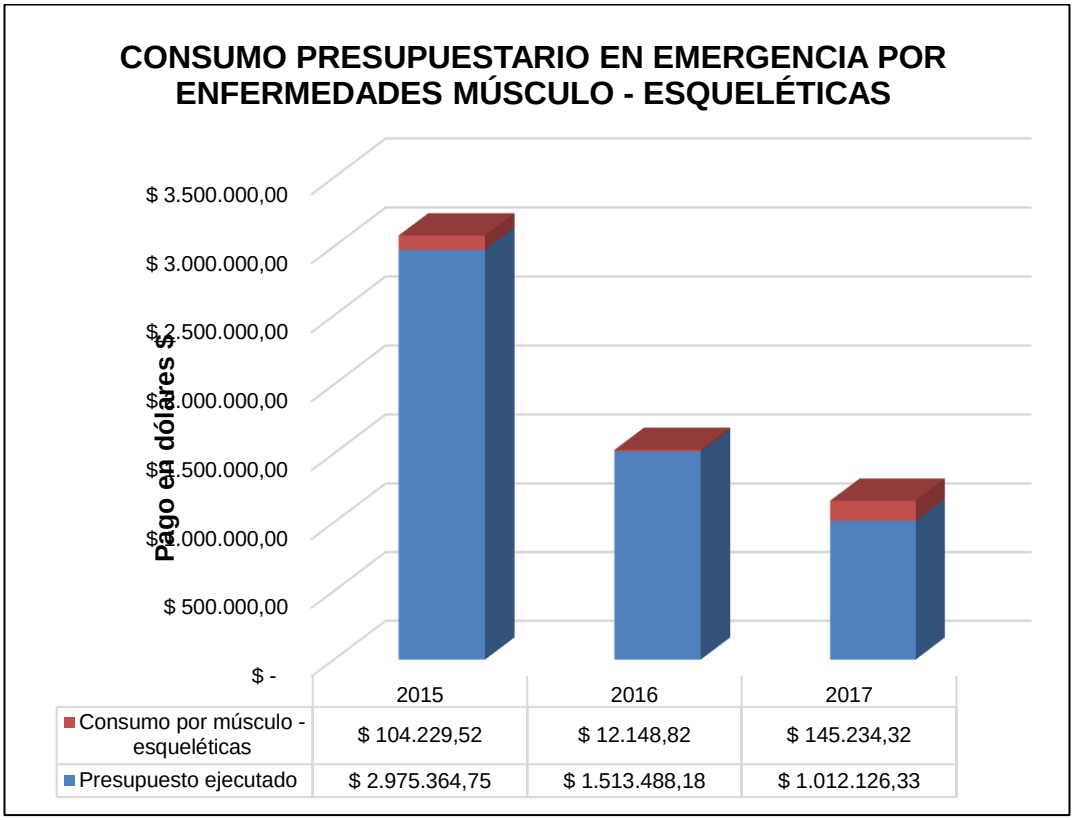
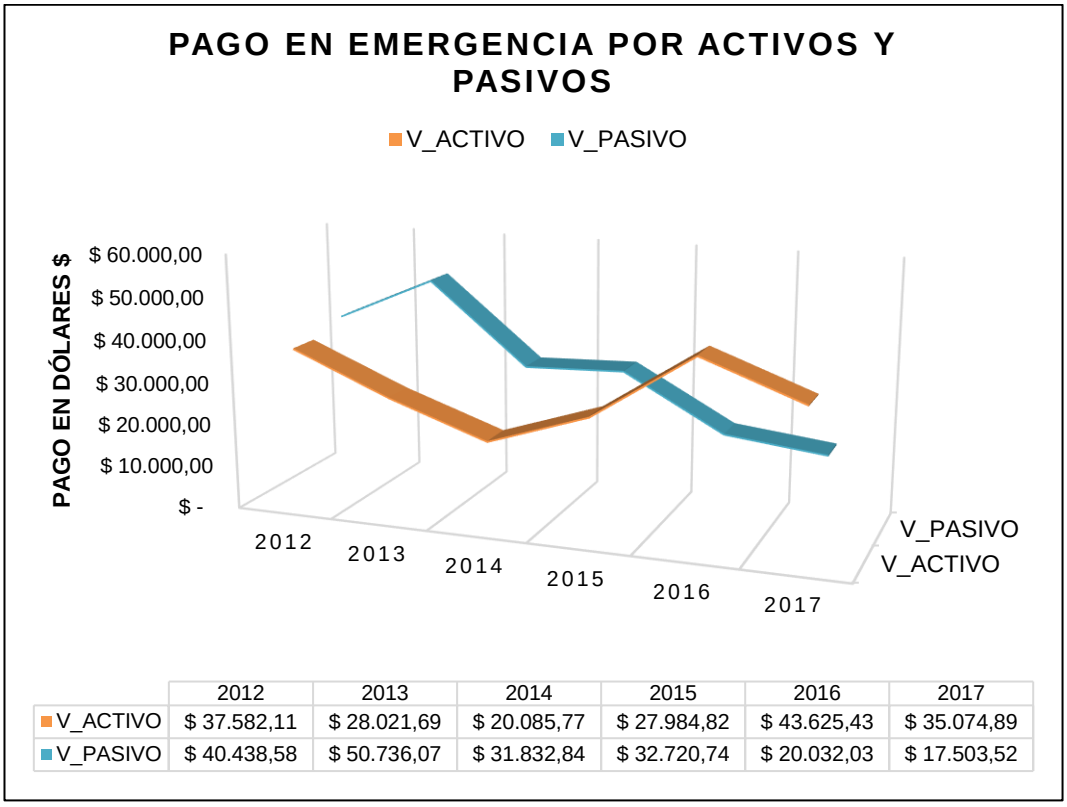


Figura 77 Consumo presupuestario en Emergencias por enfermedad músculo esqueléticas

Se evidencia, que en el comportamiento de Emergencias del año 2017 respecto al 2016, el valor pagado en el personal activo es el doble del pasivo, según criterio de expertos en el negocio, este factor podría deberse a que el personal activo sufre mayores daños por emergencia como consecuencias propias de sus actividades. (ver Figura 78)



	2012	2013	2014	2015	2016	2017
V_ACTIVO	$ 37.582,11	$ 28.021,69	$ 20.085,77	$ 27.984,82	$ 43.625,43	$ 35.074,89
V_PASIVO	$ 40.438,58	$ 50.736,07	$ 31.832,84	$ 32.720,74	$ 20.032,03	$ 17.503,52

Figura 78 Pago en Emergencia por activos y pasivos

Pese a que se evidencia un alto gasto en militares en servicio activo, se observa (ver Figura 79 y Figura 80) que los dependientes son quienes consumen en mayor proporción este servicio y se predice que existirá un incremento del 10% con respecto al año 2016, lo que explica por la cantidad de dependientes, que evidentemente supera al número de titulares del ISSFA.

CONTINUA

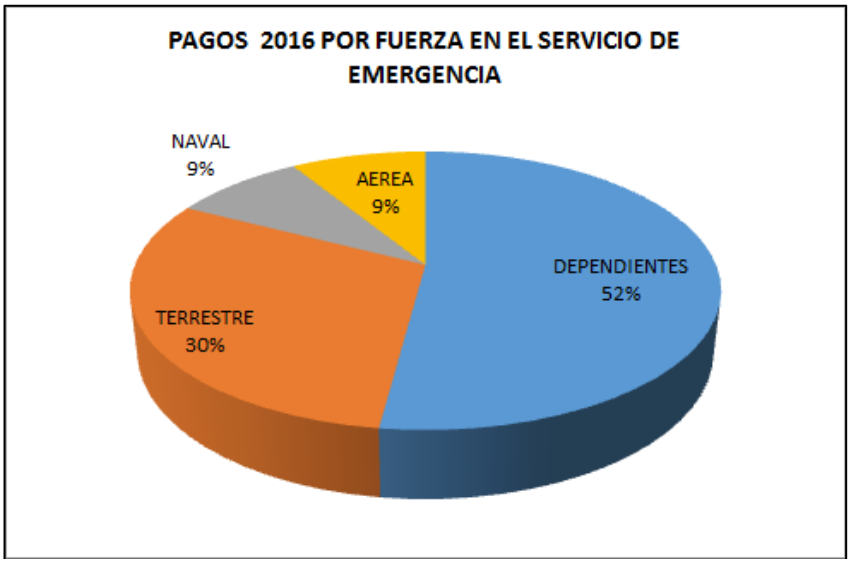


Figura 79 Pagos 2016 por fuerza en el servicio de Emergencia

Figura 80 Pagos 2017 por fuerza en el servicio de Emergencia

En relación al nivel del prestador (ver Figura 81), el servicio de Emergencia en prestadores de segundo nivel muestra una tendencia incremental, mientras que en prestadores de tercer nivel disminuye el valor pagado, a diferencia del primer nivel que mantiene un comportamiento constante.

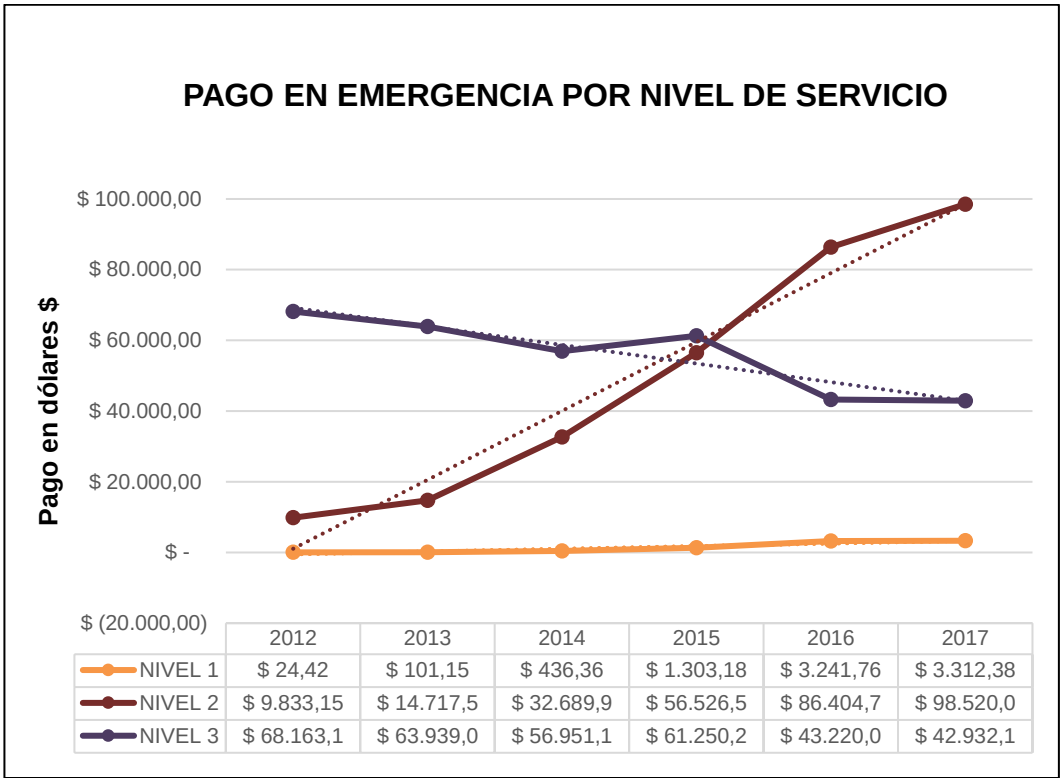


Figura 81 Pago en Emergencia por nivel de servicio

4.2.4. Atenciones médicas por consulta externa

Es un espacio del establecimiento de salud destinado a la atención ambulatoria de individuos, dentro de cada especialidad médica, incluyendo acciones de prevención. Esta atención es realizada por el médico, odontólogo, psicólogo y obstetriz a un paciente ambulatorio. Si un paciente recibe varias atenciones en un mismo día, ya sea en la misma o distinta consulta, deberá computarse tantas consultas como atenciones médicas recibidas. (Ministerio de Salud Pública del Ecuador, 2013)

En el servicio de Atenciones Médicas por Consulta Externa se presenta un comportamiento constante en los últimos 6 años (Ver Figura 82), en donde el pago no sobrepasa a los $21270 y no se encuentra una disminución significativa.

Según criterio de expertos del negocio, este hecho es un indicativo de no ha existido variación en las medidas tomadas para tratar padecimientos de enfermedades músculo-esqueléticas.

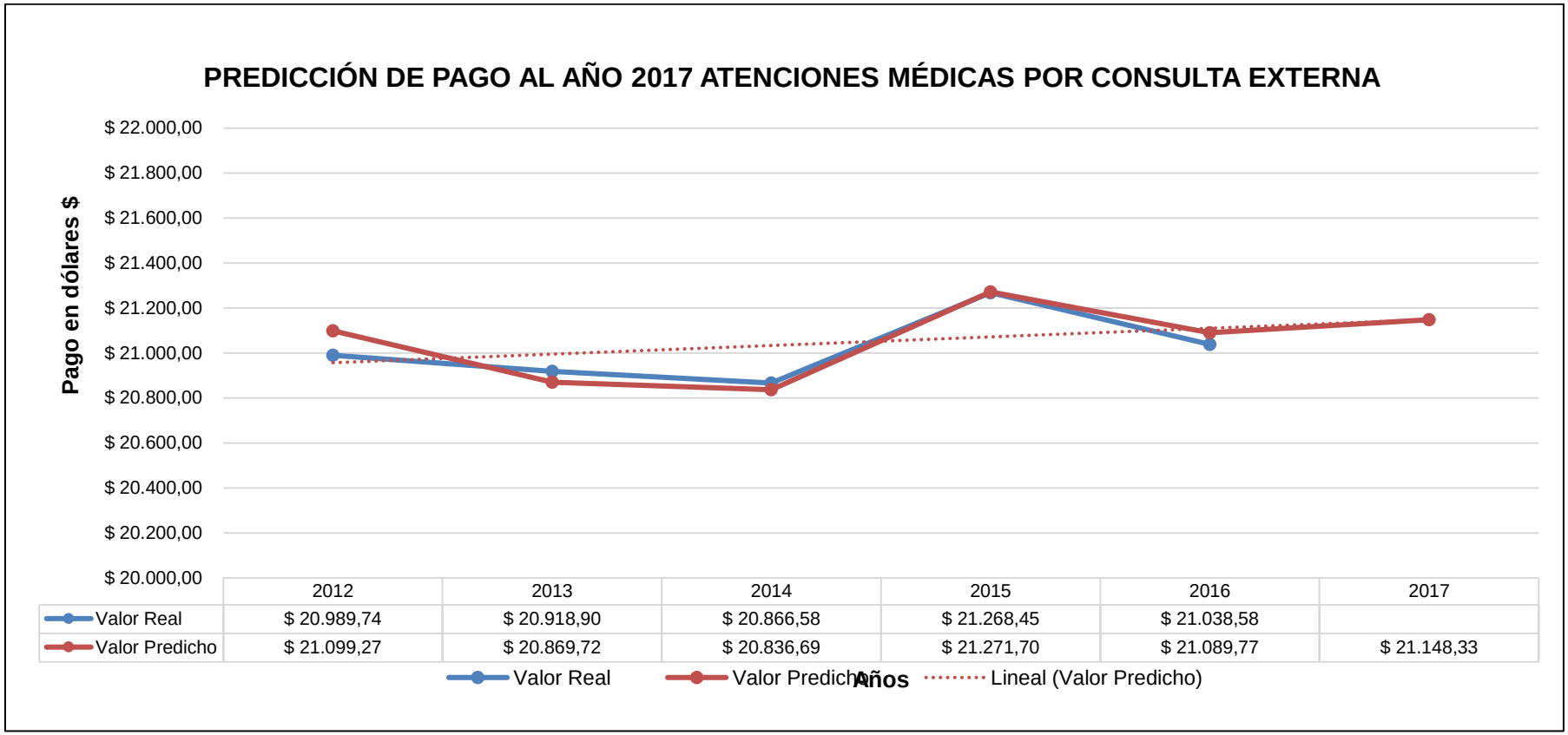


Figura 82 Predicción de pago al año 2017 Atenciones médicas por consulta externa

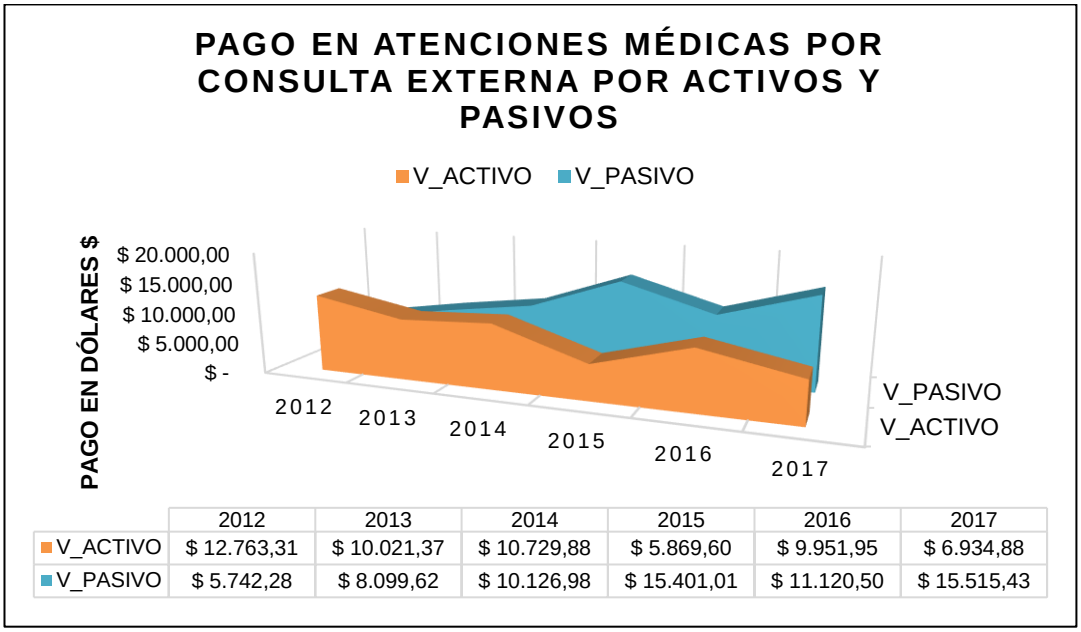
Debido al comportamiento lineal, el incremento del consumo del presupuesto no supera el 1%, como se muestra en la Tabla 21

Tabla 21

Consumo del presupuesto en Atenciones médicas por enfermedades músculo esqueléticas

Año	Porcentaje de consumo por enfermedades músculo esqueléticas
2015	1,15
2016	2,03
2017	2,89

Al año 2017, la predicción del valor pagado en el personal pasivo duplica el valor de activos. Este factor podría indicar que el personal adquiere enfermedades que requieren mayor control una vez que se jubilan, por el cambio de estilo de vida o hábitos alimenticios. También podría revelar que no existe una cultura de prevención mediante revisiones periódicas, sino que el personal acude a atenciones médicas cuando ya ha enfermado. (ver Figura 83).



	2012	2013	2014	2015	2016	2017
V_ACTIVO	$ 12.763,31	$ 10.021,37	$ 10.729,88	$ 5.869,60	$ 9.951,95	$ 6.934,88
V_PASIVO	$ 5.742,28	$ 8.099,62	$ 10.126,98	$ 15.401,01	$ 11.120,50	$ 15.515,43

Figura 83 Pago en Atenciones médicas por consulta externa por activos y pasivos

En la **Figura 84**, se observa la predicción al 2017 del valor pagado por fuerza. El mayor porcentaje será destinado a la Fuerza Terrestre y el menor a la Fuerza Naval.

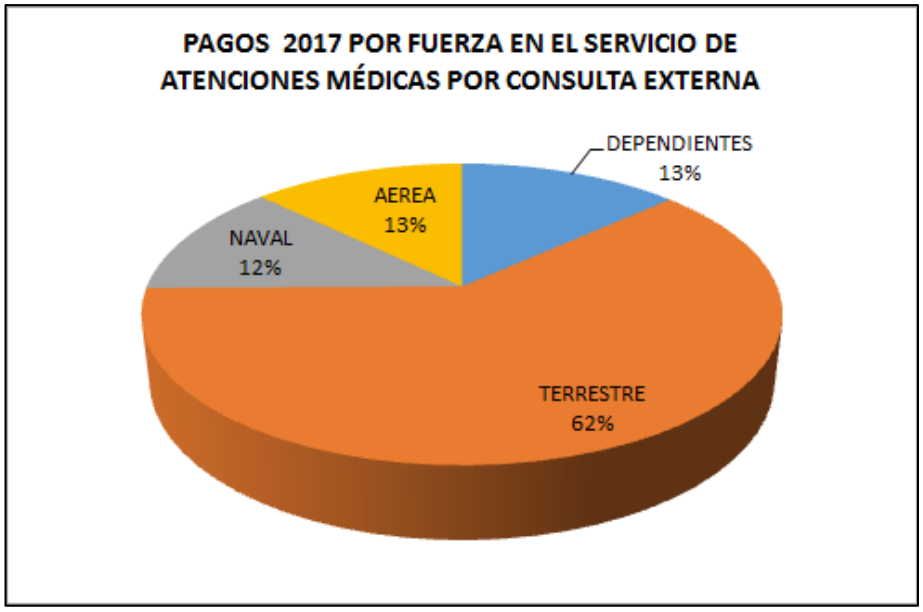


Figura 84 Pagos 2017 por fuerza en el servicio de Atenciones médicas por consulta externa

4.2.5. Reposición de gastos hospitalarios

Proceso que permite al Afiliado recuperar el valor por un servicio de salud hospitalario, en el caso de haberse atendido en un prestador privado. (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

En Reposición de Gastos Hospitalarios, se paga el mayor monto en el año 2016. La predicción para el año 2017 indica una disminución (ver Figura 85). Es importante considerar que recientemente, el Ministerio de Salud Pública del Ecuador, prohibió los pagos por reposiciones, sin embargo, se predice que a fines del año en curso se pagará únicamente 4% menos que el año pasado.

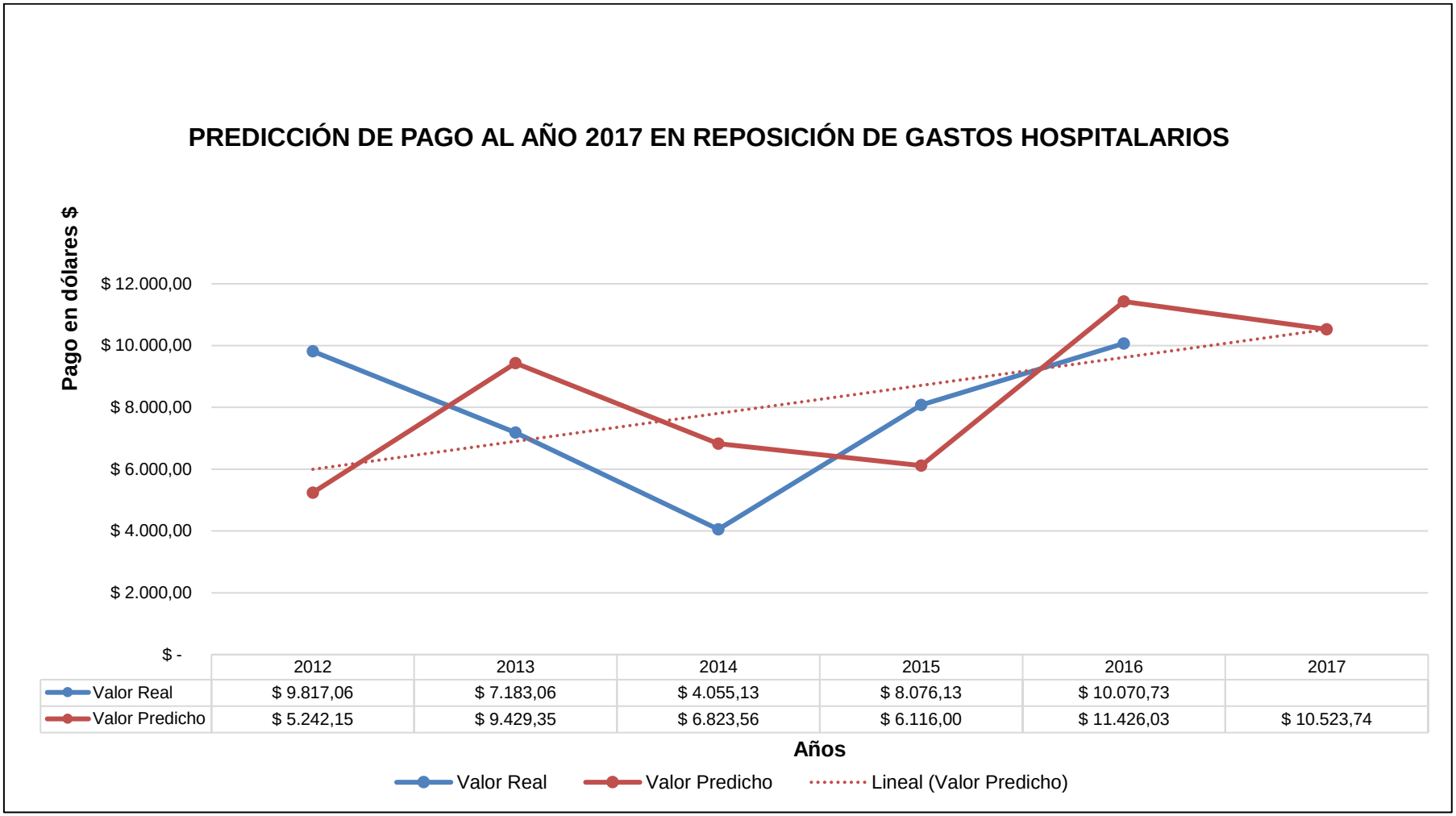


Figura 85 Predicción de pago al año 2017 en Reposición de gastos hospitalarios

En relación al personal activo y pasivo, se observa (ver Figura 86) que el año 2016 registra un monto mayor pagado personal activo, mientras que para el año 2017 este valor se disminuye en un 21%. En cuanto al personal pasivo, el valor de pago en reposiciones incrementa en un 73%. Lo antes mencionado podría ser un indicador de que la población activa recurre con mayor frecuencia a clínicas privadas, para recibir el servicio en períodos de tiempo más cortos en comparación de los tiempos de agendamiento de prestadores de tercer nivel. Esto con la necesidad de reincorporarse a sus actividades lo más pronto posible.

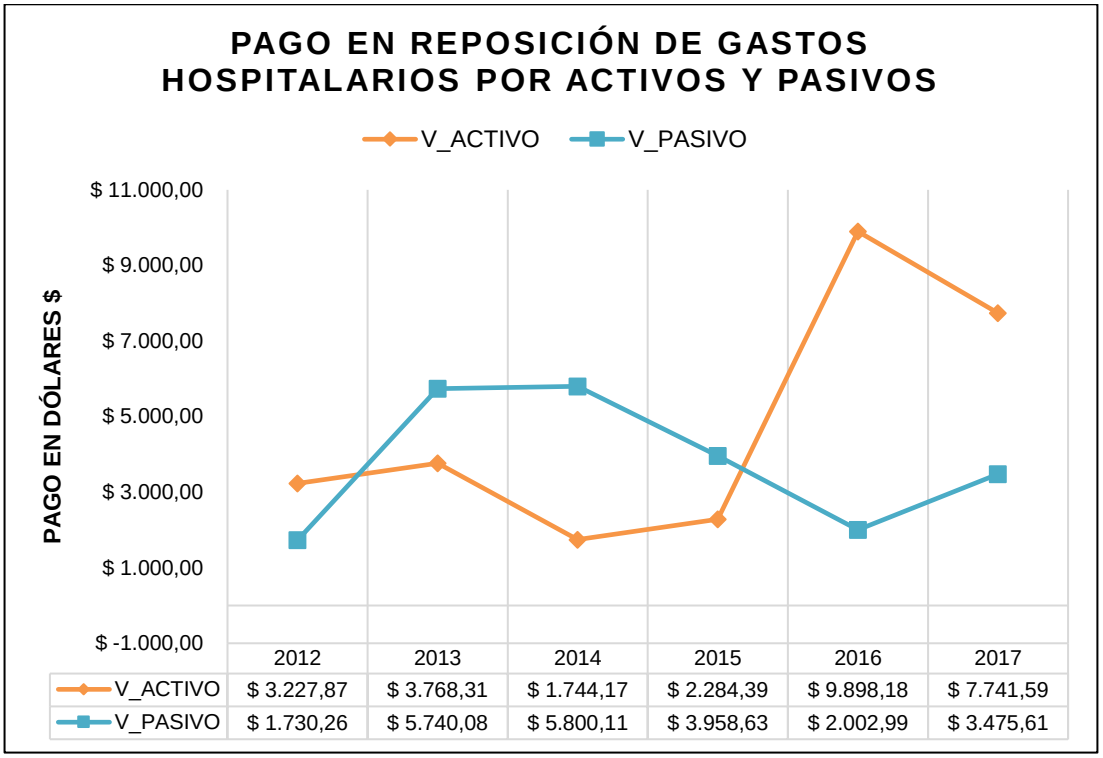


Figura 86 Pago en reposición de gastos hospitalarios por activos y pasivos

CAPITULO 5: CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Se desarrollaron satisfactoriamente modelos enfocados a la predicción del pago en servicios de salud esqueléticos utilizando cuatros técnicas de minería de datos, siendo las más precisas el árbol de decisión y regresión lineal. El árbol de decisión se utilizó para predecir aspectos cualitativos mientras que regresión lineal valores numéricos.
- La definición correcta del alcance del proyecto depende del conocimiento del usuario del negocio, como parte fundamental para la especificación de tareas.
- La exploración de los datos fue clave para encontrar los valores atípicos (outliers) en los atributos. Mediante el diagrama de cajas se observó mayor dispersión, únicamente en el atributo clase valor total, razón por la cual fue necesario realizar el tratamiento previo a la construcción del modelo.
- Un mínimo porcentaje del conjunto inicial de datos requería proceso de limpieza, tarea que no fue compleja con la utilización de componentes de la herramienta RapidMiner que incorporan estas funcionalidades.
- A más de la tasa de error RMSE que es el indicador más común para la evaluación del modelo, es importante considerar otros criterios para complementar la evaluación como son la precisión, costo computacional, indicador kappa.
- La ejecución de validación cruzada da como resultado un incremento del 40% en la tasa de error RMSE cuando hay presencia de alores atípicos.
- Pese a que en el estado del arte la técnica de SVM tiene un mayor grado de precisión en comparación de otras, en este proyecto se

confirma que la mejor técnica es la de Regresión lineal demostrando hasta un 5% más de precisión en todos los servicios.

- Pese al alto grado de precisión que muestra SVM, esta técnica demanda mayor costo computacional, que es directamente proporcional al número de registros con un coeficiente de crecimiento mucho más alto que las otras técnicas.
- La presencia de resultados con valores negativos en redes neuronales, indican un sobre ajuste del modelo lo cual se refleja en una alta tasa de error RMSE. Motivo por el cual esta técnica se descarta en los servicios de emergencia y exámenes y procedimientos.

5.2. Recomendaciones

- Identificar correctamente la clase del conjunto de datos antes de seleccionar la técnica de minería de datos a utilizar.
- Identificar los requerimientos del negocio previo a la definición del alcance del proyecto.
- Realizar la exploración de todos los atributos del conjunto de datos o al menos de la clase.
- Verificar la presencia de valores atípicos y su tratamiento con la finalidad de disminuir la tasa de error.
- Conocer todos los indicadores de precisión y rendimiento de cada técnica de modelo para seleccionar el más efectivo.
- Se recomienda realizar varias ejecuciones del modelo con diferentes parametrizaciones para reducir la tasa de error.
- Realizar la ejecución de los modelos en un equipo diferente al servidor de base de datos para no afectar la transaccionalidad.

BIBLIOGRAFÍA

Agencia Europea para la Seguridad y la Salud en el Trabajo. (Julio de 2017). *EU-OSHA: Trastornos musculoesqueléticos*. Obtenido de https://osha.europa.eu/es/themes/musculoskeletal-disorders

Agrawal, R., Imielinski, T., & Swami, A. (1993). *Mining association rules between sets of items in large databases.* Washington.

Avellano. (2016). *Redes Neuronales*. Recuperado el Abril de 2017, de http://avellano.usal.es/~lalonso/RNA/

Barnston, A. (1992). *Correspondence among the Correlation and Heidke Verification Measures.*

Bontempi, G. (2005). Structural feature selection for wrapper methods. *European Symposium on Artificial Neural Networks.*

Bouza, C., & Santiago, A. (Marzo de 2014). *La minería de datos: arboles de decisión y su aplicación en estudios médicos.* Obtenido de https://www.researchgate.net/publication/268516570_la_mineria_de_datos_arboles_de_decision_y_su_aplicacion_en_estudios_medicos

Breiman, L., & Cutler, A. (2016). *Random Forests*. Recuperado el Abril de 2017, de http://www.stat.berkeley.edu

Breuning, M., Kriegel, H., & Sander, J. (2000). LOF: Identifying Density-based Local Outliers. *ACM SIGMOD International Conference on Management of Data*, (págs. 93-104).

Chandra Pandey, S., & Allahabad, N. (2017). Data Mining Techniques for Medical Data: A Review. *IEEE*.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2016). *Step-by-step data mining guide.*

Clínica Sebastián de Benálcazar. (2017). *Hospitalización*. Obtenido de http://portal.colsanitas.com/portal/web/clinica-sebastian/hospitalizacion-y-cirugia

Comando Conjunto de las Fuerzas Armadas. (Junio de 2017). *Comando Conjunto de las FF.AA.* Recuperado el Junio de 2017, de https://www.ccffaa.mil.ec/

Cristianini, N., & Shawe-Taylor , J. (2017). *Support Vector Machines*. Recuperado el Abril de 2017, de http://www.support-vector.net

Díaz, J., Gallego, B., & León, A. (2006). *El diagnóstico médico: bases y procedimientos.*

Dirección del Seguro de Salud del ISSFA. (2016). *Formulario MRL -SCP-01 Descripción y Perfil del Puesto Jefe de Economía de la Salud.*

Dirección del Seguro de Salud del ISSFA. (2016). *Formulario MRL-SCP-01 Descripción y Perfil del Puesto Director del Seguro de Salud.*

Dirección del Seguro de Salud del ISSFA. (2016). *Procedimiento de Pertinencia y Liquidación de Servicios de Salud del ISSFA.*

Dirección del Seguro de Salud del ISSFA. (2016). *Procedimiento para cobertura de servicios de salud a los aspirantes a oficiales, aspirantes a tropa y conscriptos siniestrados en actos de servicio.*

Dirección Financiera y Departamento de presupuesto del ISSFA. (2017). *Cédula presupuestaria de egreso.* Quito.

Divya, T., & Agarwal , S. (2013). A survey on data mining approaches for healthcare, International Journal of Bio-Science and Bio-Technology. *IEEE*.

Diz, J., Marreiros, G., & Freitas, A. (2016). Applying Data Mining Techniques to Improve Breast Cancer Diagnosis. *Springer Science+Business Media.* New York.

Durgadevi, M., & Kalpana, R. (2017). Medical Distress Prediction Based on Classification Rule Discovery Using Ant-Miner Algorithm. *International Conference on Intelligent Systems and Control (ISCO).*

ECC, IAA, & UTPL. (2008). *Clasificación supervisada y no supervisada*. Recuperado el Abril de 2017, de https://advancedtech.wordpress.com/2008/04/14/clasificacion-supervisada-y-no-supervisada/

Fonasa. (2017). *Procedimientos Médicos (Ambulatorios y Hospitalarios)*. Obtenido de https://www.fonasa.cl/sites/fonasa/beneficiarios/coberturas/plan-general/procedimientos-medicos

Frost, J. (Diciembre de 2013). *Minitab.com*. Obtenido de Regression Analysis Tutorial and Examples: http://blog.minitab.com/

García, C., & Gómez, I. (2014). *algoritmos de aprendizaje: knn & kmeans.*

Gartner. (Julio de 2017). *Gartner: IT Glossary*. Obtenido de https://www.gartner.com/it-glossary/predictive-modeling

Gartner. (2017). *Gartner: IT Glossary*. Obtenido de Predictive Analytics: https://www.gartner.com/it-glossary/predictive-analytics

Gartner. (2017). *Glossary: Descriptive Analytics*. Obtenido de https://www.gartner.com/it-glossary/descriptive-analytics

Gartner. (14 de Febrero de 2017). *Magic Quadrant for Data Science Platforms.* Obtenido de https://rapidminer.com/resource/gartner-magic-quadrant-data-science-platforms/

Gómez, H., Hernandez, G., & Martinez, A. (2016). Comparativa entre crisp-dm y semma para la limpieza de datos en productos modis en un estudio de cambio de cobertura y uso del suelo. *ieee 11ccc*.

Gonzalez, F. (2014). *Aprendizaje de Máquina*. Obtenido de http://dis.unal.edu.co/profesores/fgonza/courses/2007-I/ml/index.html

Hospital de Especialidades "Eugenio Espejo". (2017). *Emergencias*. Obtenido de http://hee.gob.ec/emergencias/

Huang, W., McGregor, C., & James, A. (2014). A Comprehensive Framework Design for Continuous Quality Improvement within the Neonatal Intensive Care Unit: Integration of the SPOE, CRISP-DM and PaJMa Models. *Biomedical and Health Informatics (BHI)*.

IBM. (2012). *Manual CRISP-DM de IBM SPSS.*

IBM. (Junio de 2017). *IBM SPSS Modeler*. Recuperado el Junio de 2017, de https://www.ibm.com/co-es/marketplace/spss-modeler

Instituto de Seguridad Social de las Fuerzas Armadas ISSFA. (Septiembre de 2017). *ISSFA*. Recuperado el Abril de 2017, de http://www.issfa.mil.ec/

ISSFA. (2017). *Reporte de cuentas de la Dirección de Salud 2017.* Quito.

Jabbar, M., Deekshatulu, B., & Chndra, P. (2014). Alternating decision trees for early diagnosis of heart disease. *Proceedings of International Conference on Circuits, Communication, Control and Computing (I4C 2014).*

Jacob, S., & Ramani, G. (2012). Data Mining in Clinical Data Sets: A Review. *International Journal of Applied Information Systems (IJAIS).* New York.

Jacob, S., Ramani, G., & Nancy, P. (2011). Feature Selection and Classification in Breast Cancer Datasets through Data Mining Algorithm.

Jain , A., Duin, P., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Trans. on Pattern Analysis and Machine Intelligence.*

Khemphila, A., & Veera, B. (2010). Comparing performances of logistic regression, decision trees , and neural networks for classifying heart disease patients. *International Conference on Computer Information Systems and Industrial Management Applications .*

Kunwar, V., Khushboo, C., Sabitha, S., & Bansal, A. (2016). Chronic kidney disease analysis using data mining classification techniques. *International Conference - Cloud System and Big Data Engineering (Confluence).*

Lantares. (2015). *Las estrategias del análisis predictivo y sus ventajas*. Obtenido de http://www.lantares.com/blog/las-estrategias-del-analisis-predictivo-y-sus-ventajas

Lantares. (2016). *Prescriptive Analytics: de análisis predictivo a análisis prescriptivo*. Obtenido de http://www.lantares.com/blog/prescriptive-analytics-de-analisis-predictivo-a-analisis-prescriptivo

Lukáčová, A., Babič, F., & Paraličová, Z. (2015). How to Increase the Effectiveness of the Hepatitis Diagnostics by Means of Appropriate Machine Learning Methods. *Springer International Publishing Switzerland.* Switzerland.

Mathematics. (2017). *Mathematics*. Obtenido de Regression vs Classification: https://math.stackexchange.com/questions/141381/regression-vs-classification

Medline. (2017). *Medline*. Recuperado el Abril de 2017, de https://www.ncbi.nlm.nih.gov/pubmed/

Ministerio de Salud Pública del Ecuador. (Agosto de 2013). *Instructivo para el llenado del Registro Diario Automatizado de Consultas y Atenciones Ambulatorias.* Obtenido de https://aplicaciones.msp.gob.ec/salud/archivosdigitales/documentosDirecciones/dnn/archivos/instructivo-rdaca__final_04_09_2013.pdf

Ministerio de Salud Pública del Ecuador. (Enero de 2017). *Norma técnica de relacionamiento para la prestación de servicios de salud entre instituciones de la red pública integral de salud y de la red privada complementaria y su reconocimiento económico.* Obtenido de http://www.issfa.mil.ec/descargas/2017/Normativas/norma-de-relacionamiento-2017.pdf

Ministerio de Trabajo y Empleo. (2015). *Código de trabajo: Régimen Laboral Ecuatoriano.*

Moine, J., & Haedo, A. (2016). Una herramienta para la evaluación y comparación de metodologías de minería de datos.

Mora, R., & Rodriguez , M. (2001). Estadística Informática: casos y ejemplos con el SPSS.

Motaz, K., Saad, & Nabil, M. (2009). A comparative Study of Outlier Mining and Class Outlier Mining. *ISSR Journals.*

Muñoz, J., & Amón, I. (2013). Técnicas para detección de outliers multivariantes. *Revista en Telecomunicaciones e Informática*, 11-25.

Novás, J., Gallego, B., & León, A. (2006). El diagnóstico médico: bases y procedimientos.

Paliyawan, P., Nukoolkit, C., & Mongkolnam, P. (2014). Prolonged Sitting Detection for Office Workers. *2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON).*

Panwong, P., & lam-On, N. (2016). Predicting Transitional Interval of Kidney Disease Stages 3 to 5 Using Data Mining Method. *Second Asian Conference on Defence Technology (ACDT).*

Paredes Chicaiza, P. (2017). *Incidencia de lesiones músculo esqueléticas en tren superior en personal militar.* Obtenido de

http://repositorio.uta.edu.ec/bitstream/123456789/25837/2/tesis%20lesiones%20musculo%20esqueleticas%20final.pdf

Parenteau, J., Sallam, R., Howson, C., Tapadinhas, J., Kchiegel, K., & Oestreich, T. (2016). *Magic Quadrant for Business Intelligence and Analytics Platforms.*

Pérez, N., Guevara, M., Silva, A., Ramos, I., & Loureiro, J. (2014). Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection. *Computer Science and Information Systems (FedCSIS).*

Phuong, T., Lin, Z., & Altman, R. (2005). Choosing SNPs using feature selection. *IEEE Computational Systems Bioinformatics Conference.*

Piatetsky, G. (2017). *Kdnuggets*. Recuperado el Junio de 2017, de http://www.kdnuggets.com/

Piffaut, P. (2015). *Inteligencia de Negocios y Análisis de Negocios: ¿Cuál es el nexo?* Recuperado el Abril de 2017, de http://langeron.com/white-papers/business-intelligence-analytics.html

Princy, T., & J, T. (2016). Human Heart Disease Prediction System using Data Mining Techniques. 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT].

Pudil, P., Novovicová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognigion Letters.*

Random Forests. (2014). *A Linear Support Vector Machine*. Obtenido de https://randomforests.wordpress.com/2014/01/29/a-linear-support-vector-machine/

RapidMiner. (2017). Obtenido de Forward Selection: https://docs.rapidminer.com/studio/operators/modeling/optimization/feature_selection/optimize_selection_forward.html

RapidMiner. (2017). *Decision Tree*. Obtenido de https://docs.rapidminer.com/studio/operators/modeling/predictive/trees/parallel_decision_tree.html

RapidMiner. (2017). *RapidMiner Documentation*. Obtenido de https://docs.rapidminer.com/studio/operators/validation/cross_validation.html

RapidMiner Community. (2017). *Interpretation/Meaning of Performance Measure*. Obtenido de https://community.rapidminer.com/t5/Original-Rapid-I-Forum/Interpretation-Meaning-of-Performance-Measure/td-p/8749

RapidMiner Documentation. (2017). *Performance*. Obtenido de https://docs.rapidminer.com/studio/operators/validation/performance/predictive/performance_classification.html

Reyes, M. (2016). *¿Cómo pueden las empresas monetizar sus datos?* . Obtenido de http://www.elmundo.es/economia/2015/05/27/5565986522601d4b378b4583.html

Rodriguez, O. (Junio de 2017). *Aprendizaje Supervisado*. Obtenido de http://www.oldemarrodriguez.com/

Roslina, & Noraziah. (2010). Prediction of Hepatitis Prognosis using support vector machine and wrapper method. *IEEE Fuzzy Systems and Knowledge Discovery.*

Rouse, M. (Juniox de 2010). *TechtTarget: real-time business intelligence (BI).* Obtenido de http://searchbusinessanalytics.techtarget.com/definition/real-time-business-intelligence-BI

Rousseeuw, P., & Kaufman, L. (1990). *Finding Groups in Data: An Introduction to Clúster Analysis.*

Rousseeuw, P., & Kaufman, L. (1990). *inding Groups in Data: An Introduction to Clúster Analysis. Wiley.*

Sailesh, S., & Lu, K. (2016). Context Driven Data Mining to Classify Students of Higher Educational Institutions. *International Conference on Inventive Computation Technologies (ICICT)*.

Salud S.A. (2017). *Glosario Salud sa*. Obtenido de https://www.saludsa.com/glosario/

salud.ccm.net. (2014). *Músculo esquelético- Definición*. Obtenido de http://salud.ccm.net/faq/definiciones-48#18554

Sathyadevi, G. (2011). *Application of Cart Alorithm in Hepatitis Disease Diagnosis.* Tamil: IEEE.

Shapiro, G. P. (2015). *Microarray Data Mining: Facing the Challenges.*

Shekhar, M., Chikka, V., & Thomas, L. (2015). Identifying Medical Terms Related to Specific Diseases. *IEEE 15th International Conference on Data Mining Workshops.*

Shomona, J., & Ramani, G. (2012). Mining of Classification Patterns in Clinical data through data mining methods and techniques.

Stuart, G., Bienenstock, E., & Doursat, R. (1992). *Neural Networks and the bias/variance dilemma.*

Tangient LLC. (2017). Recuperado el 06 de 2017, de https://metodosemma.wikispaces.com/comparativa

Tapia, V., Pérez, A., & Pérez, R. (2016). Automating the analysis and evaluation of occupational risk factors accumulated in the flower industry. *2016 11th Iberian Conference on Information Systems and Technologies (CISTI).*

Tirumala, S., & Narayanan, A. (2016). Attribute Selection and Classification of Prostate Cancer Gene Expression Data Using Artificial Neural Networks. *Springer International Publishing Switzerland.*

Vaca, R., & Alarcón, J. (2015). Importancia de Implementar da Direccion de Gestión de Riesgos en El Ejército.

Venkataraman, A., Kubicki, M., Westin, C.-F., & Golland, P. (2010). Robust Feature Selection in Resting-State fMRI Connectivity Based on Population Studies.

Villada, F., Muñoz, W., & Henao, A. (2015). Aplicación de las redes neuonales en ingenieria y economía.

Wikipedia. (2017). *Diagrama de Cajas*. Obtenido de https://es.wikipedia.org/wiki/Diagrama_de_caja

Wikipedia. (2017). *Wikipedia*. Obtenido de Traumatismos: https://es.wikipedia.org/wiki/Traumatismo

World Health Organization. (Octubre de 2017). *CIE 10 Classifications*. Obtenido de http://www.who.int/classifications/icd/icdonlineversions/en/#

Yun, Z., & Chen, L. (2014). Applying Balanced ScordCard Strategic Performance Management to CRISP-DM. *International Conference on Information Science, Electronics and Electrical Engineering (ISEEE),*.

GLOSARIO

Afiliado en servicio activo: Personal militar desde el momento en que es dado de alta es decir: Subteniente, Alférez, Soldado, Marinero y Aerotécnico.

Afiliado en servicio pasivo: Personal militar que ya no está en ejercicio de trabajo militar.

Atención: “Es la asistencia sanitaria esencial basada en métodos y tecnologías prácticos, científicamente fundados y socialmente aceptables, puesta al alcance de todos los individuos y familias de la comunidad mediante su plena participación y a un costo que la comunidad y el país puedan soportar, en todas y cada una de las etapas de su desarrollo con un espíritu de autorresponsabilidad y autodeterminación. La atención primaria forma parte integrante tanto del sistema nacional de salud, del que constituye la función central y el núcleo principal, como del desarrollo social y económico global de la comunidad." (Ministerio de Salud Pública del Ecuador, 2013)

Atenciones por Consulta Externa: Es un espacio del establecimiento de salud destinado a la atención ambulatoria de individuos, dentro de cada especialidad médica, incluyendo acciones de prevención. Esta atención es realizada por el médico, odontólogo, psicólogo y obstetriz a un paciente ambulatorio. Si un paciente recibe varias atenciones en un mismo día, ya sea en la misma o distinta consulta, deberá computarse tantas consultas como atenciones médicas recibidas. (Ministerio de Salud Pública del Ecuador, 2013)

Atributo clase o label: campo que corresponde a la variable que se va a predecir en función de otras variables. (Mathematics, 2017)

Bioinformática: Es el campo de la ciencia en donde se aplican ciencias de la computación y estadística sobre datos biológicos o médicos. (Jacob & Ramani, Data Mining in Clinical Data Sets: A Review, 2012).

Categoría: Es la clasificación que se le da a un afiliado dependiendo del estado militar en el que se encuentre o su grado de dependencia con un afiliado. En el ISSFA las categorías más comunes son: activo, pasivo, esposa activo, esposa pasivo, hijo activo, hijo pasivo, padre activo, padre pasivo, montepío esposa.

Dependientes: Son asegurados del ISSFA que cuentan con cobertura debido a un parentesco que cuentan con un militar en servicio activo o pasivo, se considera como dependientes a esposas, hijos, padres y madres.

Diagnóstico médico: Procedimiento por el cual se identifica una enfermedad, es una de las tareas fundamentales de los médicos y la base para una terapéutica eficaz. Hay quienes lo señalan como la parte más importante del trabajo médico. (Díaz, Gallego, & León, 2006). Un diagnóstico médico se codifica utilizando la Clasificación Internacional de Enfermedades, décima versión, conocida como CIE10. (World Health Organization, 2017)

Economía de la Salud: Es una parte de economía que compete la realización de análisis estadísticos, estudios y construcción de modelos que permitan mejorar las condiciones en el ámbito de la salud.

Emergencia: Toda contingencia de gravedad que afecte a la salud del ser humano con inminente peligro para la conservación de la vida o de la integridad física de la persona, como consecuencia de circunstancias imprevistas e inevitables, independientemente del lugar de su acontecimiento. (Salud S.A., 2017)

El proceso de atención médica en el Servicio de Emergencia se inicia en el Área de Triage donde se determina la prioridad de atención a través de la valoración de signos vitales y sintomatología, después de esto se traslada al paciente al área que requiera dentro del Servicio de acuerdo a su patología para continuar la valoración médica y exámenes complementarios para llegar al diagnóstico y aplicar el tratamiento adecuado. (Hospital de Especialidades "Eugenio Espejo", 2017)

Enfermedades músculo esqueléticas (EME): Son enfermedades de origen laboral más comunes que se adquieren por: incrementos en el ritmo laboral, concentración de fuerzas en manos, muñecas y hombros, posturas forzadas y mantenidas causantes de esfuerzos estáticos en diversos músculos. Normalmente afectan a la espalda, cuello, hombros y extremidades superiores, aunque también pueden afectar a las extremidades inferiores. Comprenden cualquier daño o trastorno de las articulaciones y otros tejidos. Los problemas de salud abarcan desde pequeñas molestias y dolores a cuadros médicos más graves que obligan a solicitar la baja laboral e incluso a recibir tratamiento médico. En los casos más crónicos, pueden dar como resultado una discapacidad y la necesidad de dejar de trabajar. (Agencia Europea para la Seguridad y la Salud en el Trabajo, 2017). En el **ANEXO A** se encuentra el listado de diagnósticos de las enfermedades músculo esqueléticas.

Exámenes y Procedimientos Especiales: Los procedimientos médicos son prestaciones de salud (atenciones unitarias o en grupo) que se otorgan a un paciente para efectos diagnósticos, terapéuticos o quirúrgicos, implican el uso de equipamiento, instrumental, instalaciones y profesionales especializados, dependiendo de la complejidad del procedimiento y de las condiciones clínicas del paciente. (Fonasa, 2017)

Hospitalización: Este servicio está orientado a proporcionar cuidados básicos y especializados seguros en ambiente hospitalario confortable, que genere la satisfacción de nuestros usuarios y sus familias, además que propicie su participación en el proceso de atención con respeto de la autonomía y dignidad humana, así como el derecho a la intimidad y confidencialidad, garantizando de este modo la prestación de servicios asistenciales con altos estándares de calidad para el atención integral de los pacientes. Desarrolla actividades médicas y de enfermería encaminadas al tratamiento y recuperación de la salud del paciente hospitalizado, ofreciendo la prestación de los servicios con recurso humano calificado, comprometido y humanitario. (Clínica Sebastián de Benálcazar, 2017)

Información clínica: Es un conjunto datos utilizados para representar el estado de salud de una persona. La información clínica comprende los signos vitales, dolencias, enfermedades, procedimientos, tratamientos y resultados de haberlos aplicado en un paciente. (Jacob & Ramani, Data Mining in Clinical Data Sets: A Review, 2012)

Músculo esquelético: Son tipos de músculos estriados unidos al esqueleto formados por células o fibras alargadas. Forman parte del 90% del músculo que tiene el cuerpo humano mientras que el 10% restante corresponde al músculo cardíaco y visceral. (salud.ccm.net, 2014)

Norma del Proceso de Relacionamiento para la atención de pacientes: Norma técnica de relacionamiento para la prestación de servicios de salud entre instituciones de la red pública integral de salud y de la red privada complementaria, y su reconocimiento económico. (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

Paciente Ambulatorio: Es la persona que utiliza los servicios de diagnóstico y/o tratamiento de un hospital o unidad operativa pero que no ocupa cama hospitalaria.

Patología: Parte de la medicina que estudia los trastornos anatómicos y fisiológicos de los tejidos y los órganos enfermos, así como los síntomas y signos a través de los cuales se manifiestan las enfermedades y las causas que las produzcan.

Reposición de gastos Hospitalarios: proceso que permite al Afiliado recuperar el valor por un servicio de salud hospitalario, en el caso de haberse atendido en un prestador privado. (Instituto de Seguridad Social de las Fuerzas Armadas ISSFA, 2017)

Servicios de Salud: Los servicios de salud, por lo tanto, son aquellas prestaciones que brindan asistencia sanitaria. Puede decirse que la articulación de estos servicios constituye un sistema de atención orientado al mantenimiento, la restauración y la promoción de la salud de las personas.

Unidad de Salud Calificada: Es una casa de salud que ha presentado todos los requisitos legales necesarios al ISSFA para brindar los servicios de salud a los asegurados del ISSFA.