DIANA MARÍA ROMERO CHACÓN

UNIVERSIDAD AUTONOMA DE BUCARAMANGA
FACULTAD DE MEDICINA
BUCARAMANGA
2015

DESCUBRIMIENTO DE PATRONES DE SUPERVIVENCIA EN PACIENTES CON TRASPLANTE RENAL UTILIZANDO MINERÍA DE DATOS

DIANA MARÍA ROMERO CHACÓN

Trabajo de Investigación para optar al título de Especialista en Epidemiología

UNIVERSIDAD AUTONOMA DE BUCARAMANGA
FACULTAD DE MEDICINA
BUCARAMANGA
2015

SIGLAS

TR: Trasplante Renal

CAC: Cuenta de Alto Costo

ERC: Enfermedad Renal Crónica

TFG: Tasa de Filtración Glomerular

NKF: National Kidney Foundation

EPIC: Norfolk (European Prospective Investigation of Cancer and Nutrition

TRRC: Terapia de reemplazo renal continúa

RIPS: Registro Individual de Prestación de Servicios

RUAF: Modulo de Nacimientos y Defunciones del Registro Único de Afiliados

NHANES III: (National Health and Nutrition Examination Survey)

EPS: Entidad Promotora de Salud

IPS: Institución Prestadora de Salud

SISPRO: Sistema Integral de Información de la Protección Social

ND – RUAF: Modulo de Nacimientos y Defunciones del Registro Único de Afiliados

HTA: Hipertensión Arterial

DM: Diabetes Mellitus

TABLA DE CONTENIDO

Lista de Figuras

Lista de Tablas

Resumen

El paciente con enfermedad renal crónica presenta un mayor riesgo de mortalidad Cardiovascular. Cuando la enfermedad alcanza una etapa avanzada es imprescindible reemplazar algunas de las funciones del órgano mediante las denominadas Terapias de Reemplazo Renal, que intentan evitar el estadio de intoxicación sanguínea conocido como uremia y prolongar la vida del paciente. El trasplante renal es la mejor opción para el paciente con Insuficiencia Renal Crónica, debido a que en la mayoría de casos logra una recuperación completa y mejora la calidad de vida al prescindir de la dependencia de diálisis y las dietas rigurosas. En Colombia el primer trasplante renal se llevó a cabo por el Hospital San Vicente de Paul en el año 1973. En América Latina, Colombia ocupó el quinto puesto en el año 2007 en el número de trasplantes renales realizados.

El objetivo del presente estudio es conocer los patrones de supervivencia en pacientes con trasplante renal a partir de la información de la Cuenta de Alto Costo del año 2009, se utilizaron técnicas de minería de datos para encontrar relaciones subyacentes que no se pueden identificar mediante un tratamiento estadístico clásico. La edad y el diagnostico principal de Diabetes Mellitus fueron identificados como factores condicionantes que influyen en la predicción del desenlace final del paciente.

Se escogió la minería de datos como herramienta para el análisis de la información y la generación de conocimiento debido a que se planea encontrar relaciones no explicitas entre los atributos del conjunto de datos. La aplicación de métodos de agrupamiento y clasificación de minería de datos permitió construir un modelo de conocimiento para identificar patrones de comportamiento entre las variables analizadas.

La información que se utilizó para la generación de los modelos de minería de datos, fue dispuesta por la Bodega de datos de SISPRO – Sistema Integral de Información de la Protección Social, del Ministerio de Salud y Protección Social, utilizando las Fuentes de la Cuenta de Alto Costo – ERC (Enfermedad Renal Crónica), Registro Individual de Prestación de Servicios - RIPS y el Módulo de Nacimientos y Defunciones del Registro Único de Afiliados - RUAF. Esta información fue transformada para la aplicación de técnicas de Minería de Datos que permitieron la extracción de conocimiento para la definición de factores que inciden en la supervivencia de los pacientes con trasplante renal. La edad y el

diagnostico principal de Diabetes Mellitus fueron identificados como factores condicionantes que influyen en la predicción del desenlace final del paciente.

Para los datos analizados se encontró una tasa de supervivencia del 66% a los 72 meses de transcurrido el trasplante, el tiempo medio de supervivencia es de 61.6 meses, con un intervalo de confianza del 95% (IC 95) entre 61,1 y 62,13 meses. Se realizó un análisis de las curvas de supervivencia por la variable edad, para lo cual se definieron cuatro agrupaciones de pacientes, pacientes con edad inferior a 43 años, entre 43 y 54 años, entre 55 y 72 años y mayores a 72 años, la diferencia es claramente significativa, con tasas de supervivencia a los 6 años de 90,9%, 76,5%, 59,5% y 37,1% respectivamente.

Se propone como trabajo futuro la incorporación de los modelos de minería de datos sobre el reporte de la información de la Cuenta de Alto Costo - Enfermedad Renal Crónica realizado en el marco de la resolución 2463 de 2014, donde se incluyen nuevas variables que podrían aportar otros factores determinantes en la supervivencia de pacientes con trasplante renal.

Palabras Clave: Trasplante Renal, Sobrevida, Minería de Datos, weka, perceptrón multicapa, clúster, Microsoft Excel, arboles de decisión

Abstract

The patient with Chronic Kidney Disease presents a high risk of cardiovascular mortality. When the illness gets an advanced stage, it is unpredictable to replace some functions of the organ by the denominated Kidney Replacement Therapy, this one tries to avoid the blood intoxication known as uremia and to extend the life of the patient. The renal transplant is the best choice for the patient with Chronic Kidney Insufficiency, because most of the cases get a complete recovery and a better quality of life since the patients do not depend of the dialysis and strict diets. In Colombia the Transplant Group at the Antioquia University and San Vicente Paul Hospital did the first renal transplant in 1973. Colombia is the fifth country in Latin America related with the high frequency per million inhabitants with Renal Transplant.

The objective of this study is to recognize in the patients with Kidney Transplant the patterns of survival based in the information of the High Cost Account in 2009, using mining techniques of dates excepting to find underlying relations that cannot be identified by a classic statistical process. Observing risk factors that affect

patients with transplant, they can be prevented, detected and treated. Different morbidities that affected them can be treated in the same way.

Data mining was chosen as a tool to analyze the information to generate knowledge due to the fact we plan to find non-explicit relationships between characteristics of data sets. The application of grouping methods and data mining classification allowed us to build a knowledge model to identify behavioral patterns between the analyzed variables.

The SISPRO Data Warehouse – Integral information System of Social protection, Ministry of health and Social protection, using the Sources of High Costs Account - Chronic Kidney Failure (CKF), provided the information used to generate data mining methods. Individual services register and national birth and death registration. This information was modified to apply data mining techniques that allowed extraction of knowledge to get the definition of factors that have a bearing on the survival rate of patients with renal transplant. The age and main diagnosis of Diabetes Mellitus were identified as factors influence the prediction of the patient's final outcome.

On the analyzed data it was found a survival rate of 66% 72 months after the transplant surgeon, survival average time is 61.6 months, with a confidence interval of 95% between 61.1 and 62.13 months. Survival curves analyses was applied by the age variable, so four different groups of patients were defined, patients younger than 43, patients between 43 and 54, patients between 55 and 72 and patients older than 72. The difference is clearly meaningful, with survival rates of 6 years of 90,9%, 76,5%, 59,5% and 37,1%.

Incorporating data mining models is proposed as a future work, about the report on the Sources of High costs Invoice- Chronic Renal Failure (CRF), resolution 2463 of 2014, where new variables are included and could be also become meaningful and influence on the survival rate of renal transplant patients.

Keywords: Kidney Transplant, Survival, Dataminig, weka, multilayer percetron, cluster, Microsoft Excel, desicion tree.

1. Planteamiento del problema y justificación

Cuáles son los principales factores socio económicos y clínicos asociados a la supervivencia en pacientes con trasplante renal?

Cuál es la tasa de supervivencia en los primeros seis años en los pacientes con trasplante renal realizado en el año 2009?

El trasplante renal es la mejor terapia de reemplazo renal para los pacientes con Insuficiencia Renal Crónica Avanzada y se ha constituido como una actividad medica consolidada en muchas instituciones de salud en el mundo. Garantiza mejores resultados en términos de una mayor calidad de vida, supervivencia y relación coste-efectividad.

Para reducir la posibilidad de rechazar el riñón donado se tiene la necesidad de tomar medicamentos diariamente, especialmente inmunosupresores que disminuyen las defensas propias del organismo, el paciente además debe seguir las indicaciones dadas por el medico nefrólogo y someterse a una valoración post trasplante. Los efectos secundarios inespecíficos de la inmunodepresión comprenden un mayor riesgo de neoplasias malignas e infecciones, en particular, de infecciones oportunistas (1).

El riesgo de cáncer y cardiopatías es varias veces mayor en los pacientes sometidos a trasplante que en la población general. El cáncer es una causa de morbimortalidad importante en la población trasplantada. Las enfermedades cardiovasculares son la causa más frecuente de muerte en los receptores de un aloinjerto renal (2).

El TR requiere un seguimiento cuidadoso inicialmente en UCI o Unidad de Trasplante y posteriormente en planta o consulta externa. Durante el seguimiento pueden surgir diversas complicaciones quirúrgicas o médicas, algunas precoces y otras tardías, tal como aparece en la Tabla 1 (3).

Caracterizar demográficamente a los pacientes con trasplante renal permitirá enfocar los cuidados post trasplante a la población más vulnerable a tener complicaciones. Los hallazgos encontrados serán importantes para soportar la toma de decisiones en los organismos gubernamentales y privados del sector salud en lo relacionado con el planteamiento de políticas públicas y programas que permitan mejorar la atención a las pacientes con trasplante renal, con el fin

de disminuir las tasas de mortalidad y como consecuencia aumentar el tiempo de supervivencia.

Tabla 1. Complicaciones del trasplante renal

Precoces	Tardías
Generales (dehiscencia herida, infección rotura injerto renal, etc....) -Necrosis tubular aguda -Vascular: sangrado, trombosis arterial o de vena renal -Linfática: Linfocele -Urológica: obstrucción, fistula urinaria -Rechazo: hiperagudo, acelerado, agudo -Nefrotoxicidad por anticalcineurínicos	-Estenosis de la arteria renal - Infecciones -Enfermedad cardiovascular (Hipertensión arterial, coronariopatía,...) -Nefropatía crónica del injerto -Recurrencia de la nefropatía -Tumores -Enfermedad ósea -Gastrointestinales, pancreáticas y hepáticas -Cutáneas -Hematológicas -Oculares -Neurológicas -Metabólicas (diabetes mellitus, aumento ácido úrico, etc.)

2. Objetivo

2.1 General

Conocer los patrones de supervivencia en pacientes con trasplante renal a partir de la información de la Cuenta de Alto Costo del año 2009, utilizando técnicas de minería de datos esperando encontrar relaciones subyacentes que no pueden identificarse mediante un tratamiento estadístico clásico.

2.2 Específicos

- Caracterizar demográficamente los pacientes con trasplante renal, teniendo en cuenta la condición de vivo o muerto al transcurrir los primeros seis años de la terapia de reemplazo renal.
- Identificar la morbilidad en los primeros cinco años de los pacientes con Trasplante Renal.
- Conocer las causas de Mortalidad en los pacientes que recibieron Terapia de Reemplazo Renal.

- Calcular la tasa de Sobrevida en el primer, tercer y sexto año posterior a la realización del Trasplante Renal.

3. Marco conceptual, teórico y estado del arte

3.1 Definición

La enfermedad renal crónica se define como un proceso fisiopatológico con múltiples causas, cuya consecuencia es la pérdida inexorable del número y el funcionamiento de nefronas, y que a menudo desemboca en insuficiencia renal terminal. A su vez, la ERC es un estado o situación clínico en que ha ocurrido la pérdida irreversible de función renal endógena, de una magnitud suficiente para que el sujeto dependa en forma permanente del tratamiento sustitutivo renal (diálisis o trasplante) con el fin de evitar la uremia, que pone en peligro la vida. La uremia es el síndrome clínico y de laboratorio que refleja la disfunción de todos los sistemas orgánicos como consecuencia de la insuficiencia renal aguda o crónica no tratada (5).

La enfermedad renal crónica ha sido considerada en el Sistema General de Seguridad Social en Salud como una patología de alto costo, por generar un fuerte impacto económico sobre las finanzas del Sistema General de Seguridad Social en Salud y por causar un dramático efecto sobre la calidad de vida del paciente y su familia, incluidas las repercusiones laborales. Se estima que el manejo de la enfermedad renal crónica ha comprometido en los últimos años aproximadamente el 2% del gasto en salud del país y el 4% del gasto en Seguridad Social en Salud (6).

3.2 Clasificación

A continuación se muestra la clasificación de la IRC en 5 estadios dada por la NKF (National Kidney Foundation), en función del filtrado glomerular renal.

Tabla 2. Tabla obtenida de National Kidney Foundation. K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification.

Estadio	Descripción	FRG
I	Daño Renal con FRG	> 90 ml / min

Estadio	Descripción	FRG
	normal	
II	Leve	60 - 89 ml/min
III	Moderado	30 - 59 ml/min
IV	Severo	15 – 29 ml/min
V	Fallo Renal (diálisis)	<15 ml/min

La enfermedad Renal Crónica se divide en cinco estadios, con el estadio mayor representando la peor función renal. En el estadio 1 están los pacientes que no tiene un claro déficit de filtración y se define como una función renal normal o elevada en presencia de daño renal evidente, el cual se define ampliamente, pero más frecuentemente se encuentra albuminuria persistente. El estadio 2 es una reducción leve de la función renal presencia de daño renal. Los estadios 3 y 4 corresponden a reducciones moderada y severa de la función renal. Esta disminución grande de la función renal se clasifica como enfermedad renal sin considerar la evidencia adicional de daño renal.

Se escogió el umbral de 60 ml/min/1.73m² porque representa una reducción de 50% de la función renal, comparada con el nivel normal de un adulto joven, y además porque se aumentan las complicaciones cuando la tasa de filtración glomerular cae por debajo de 60 ml /min /1.73m².

El estadio 5 es la falla renal con TFG menor de 15 ml/min /1.73m² o la necesidad de diálisis o trasplante renal.

3.3 Factores de riesgo que afectan la progresión de la enfermedad renal crónica

Los factores de riesgo que favorecen la aparición o afectan la progresión de la ERC, pueden ser demográficos, modificables y no modificables o desarrollarse durante la vida de un individuo, susceptible por lo tanto de prevención. Los factores sociodemográficos y clínicos que condicionan el riesgo en ERC, se muestran en la tabla 3 (23).

Tabla 3. Tabla de Riesgo de Enfermedad Crónica

Tipo	Definición	Ejemplos
Factores de susceptibilidad	Aumentan susceptibilidad a daño renal	• Mayor edad • Historia familiar de

Tipo	Definición	Ejemplos
		enfermedad renal • Bajo peso de nacimiento • Reducción de masa renal • Raza
Factores de Iniciación	Inician directamente el daño	• Diabetes • Hipertensión arterial • Enfermedades autoinmunes • Infecciones sistémicas • Infección del tracto urinario • Cálculos urinarios • Obstrucción del tracto urinario • Toxicidad a drogas
Factores de Progresión	Causan empeoramiento del daño renal y declinación más rápida de la función renal	• Control pobre de glicemia en diabetes • Tabaquismo • Proteinuria • Hipertensión arterial

3.4 Aspectos Epidemiológicos de la ERC

La verdadera incidencia y prevalencia de la ERC en una población es difícil de evaluar, ya que los estados iniciales usualmente son asintomáticos; sin embargo, diversas investigaciones han evaluado este aspecto. En Estados Unidos el NHANES III (National Health and Nutrition Examination Survey III) (1.988 – 1.994) estimó que cerca del 11% de la población presentaba algún grado de ERC (alrededor de 19 millones de habitantes); el 3% de la población tenía elevación de la creatinina y el 70% de estos pacientes tenían hipertensión arterial (HTA); la prevalencia de albuminuria fue del 12%. En el Reino Unido la cohorte EPIC- Norfolk (European Prospective Investigation of Cancer and Nutrition) en 23.964 pacientes entre los 29 y los 40 años, encontró una prevalencia de microalbuminuria y macroalbuminuria del 11.8 % y 0.9% respectivamente (7). La edad, el género femenino, la presión arterial sistólica y el

tabaquismo, fueron predictores independientes de la aparición de albuminuria.

En Colombia se calculó la prevalencia de ERC en TRRC en 181 ppm en el año 2001 y en 450 ppm en 2007. Tomando como base las estadísticas del NHANES III, se podría estimar que en Colombia hay 4.600.000 pacientes con ERC. En Colombia para el año 2012, existían 763.918 casos de ERC, de los cuales el 52,2% (398.771 casos) se encontraban en estadio 3 (CAC, 2013). La prevalencia de ERC terminal en Colombia es de 621 pacientes por millón, (27.637) de los cuales el 56% se encuentra en hemodiálisis, 25% en diálisis peritoneal, el 17% con trasplante renal y el 2% en tratamiento médico no dialítico (CAC, 2013) (8).

En el país se encuentran registrados según los datos reportados a la CAC - 2013, 575.174 Diabéticos, 2.272.642 con Hipertensión Arterial y de esta población, 765.000 en cualquier estadio de ERC. Sin embargo, aún hay varios pacientes sin estudio y diagnóstico de HTA, DM y sin descartar ERC.

En América Latina, en el año 2007, Colombia con una frecuencia de 14,6 trasplantes renales por millón de habitantes ocupó el quinto lugar después de Uruguay (36,7), Argentina (24,1), México (21,8) y Brasil (20,5). De acuerdo con el informe de la situación de la ERC en Colombia, en el año 2010 había en el país 23.707 personas con ERC estadio 5, la mayoría de ellos en terapia dialítica, 15,6% con trasplante renal como modalidad de tratamiento, correspondiente a 3.691 pacientes, 868 de los cuales se realizaron ese mismo año (9).

3.5 Supervivencia después del trasplante renal

Se han realizado varios estudios que intentan determinar la supervivencia de pacientes con trasplante renal, por ejemplo, “Long-term allograft survival after kidney transplantation” (16), cuyo objetivo fue identificar los factores asociados con la supervivencia del injerto 15 años después del trasplante. Para esto se realizó un estudio retrospectivo llevado a cabo en el Hospital Reina Sofía de Córdoba entre Febrero de 1979 y Diciembre de 1997. El análisis de la serie fue realizado aplicando análisis de Kaplan-Meier y el modelo de Cox de riesgos proporcionales Se estudiaron un total de 487 aloinjertos renales con un seguimiento medio de 114 meses, de los cuales el 37% (n = 180) sobrevivieron durante los primeros 15 años. De los 180 pacientes, las principales causas de fracaso del injerto fueron la nefropatía crónica del injerto en 29 (66 %) y la muerte del paciente en 13 (29,5 %). El análisis multivariado identificó como factores que están asociados de forma

significativa con la pérdida del injerto después de 15 años: el número de compatibilidades HLA (hazard ratio [HR] 1,25, IC del 95%: 1,01 a 1,56), el panel de anticuerpos reactivos (HR 2,61, IC del 95 % 1.28 a 5.26), y la función retardada del injerto (HR 11.25, IC 95% 1,33-95,28). Concluyendo que el alto riesgo inmunológico de los pacientes se asoció de forma independiente con la pérdida del injerto.

Con el fin de validar si existe asociación entre la condición socioeconómica del paciente con trasplante renal y un mayor riesgo de defunción se realizó el estudio, Socioeconomic deprivation is independently associated with mortality post kidney transplantation (17). Para lo cual, se obtuvieron datos de 19.103 procedimientos de trasplante renal, realizados en Inglaterra a partir de abril de 2001 a marzo de 2012. Los datos demográficos del paciente fueron: edad, sexo, tipo de donante (vivo o muerto), la etnia, el año del trasplante, el fracaso del injerto, comorbilidad médica y el nivel socioeconómico. Las medidas de resultado primarias y secundarias para el cálculo de la mortalidad a 1 y 5 años fueron calculadas con modelos de riesgos proporcionales de Cox para identificar los factores independientes asociados con la mortalidad. Los datos se dividen en quintiles de pacientes por nivel de carencia socioeconómica, de 1 al 5 (de más a menos carencias o limitaciones económicas, respectivamente). En el pos trasplante a 1 año, se registraron 566 muertes, en donde la causa más común de muerte fue infección. En comparación con las personas más necesitadas (punto de referencia), los receptores con mayores limitaciones económicas habían disminuido significativamente el riesgo de muerte a los 1 y 5 años después del trasplante renal (índice de riesgo 0,66, IC 95 % (0,57-0,76) y la razón de riesgo 0,65, IC del 95% (0,54 a 0,77), respectivamente). Por lo tanto, el nivel de privación socioeconómica se asocia de forma independiente con una mayor mortalidad post-trasplante renal.

3.6 Fuentes de Información

Las Fuentes de información almacenadas en la Bodega de Datos de SISPRO, del Ministerio de Salud y Protección Social, de las que se extraerá conocimiento son las siguientes:

- Cuenta de Alto Costo

Cuenta de Alto Costo de Colombia (CAC) es un Organismo Técnico no gubernamental del SGSSS que constituye un "Frente Común" para el abordaje del Alto Costo como fenómeno de gran impacto en el país, mediante articulación de Sociedades Científicas, Instituciones Prestadoras, Asociaciones de pacientes y Aseguradoras, que promueve la gestión de riesgos en EPS e IPS mediante pedagogía, difusión de experiencias exitosas, instrumentación y la creación de referentes técnicos, para disminuir la tendencia de nuevos casos de alto costo y procurar la calidad técnico-científica en el manejo de los casos existentes con el fin de disminuir los eventos evitables (9). La CAC se ha encargado de manera continua en la caracterización de la Enfermedad Renal Crónica (ERC) en el país, para aportar información epidemiológica confiable aplicando mecanismos de ajuste por riesgo, minimizando la distribución inequitativa de los costos de atención, promoviendo y premiando el enfoque preventivo al evaluar a los asegurados y prestadores.

- RIPS – Registro Individual de Prestación de Servicios

De acuerdo a la Resolución 3374 de 2000 expedida por el Ministerio de la Protección Social en el año 2000, los Registros Individuales de Prestación de Servicios de Salud (RIPS) se definen como: “El conjunto de datos mínimos y básicos que el Sistema de Seguridad Social en Salud requiere para los procesos de dirección, regulación y control” (20). Estos datos identifican las actividades o atenciones de salud que realizan las IPS públicas o privadas y profesionales independientes. Los RIPS son considerados una de las fuentes principales de información sobre la oferta y demanda de los servicios en el país.

A través de los RIPS, podemos obtener perfiles epidemiológicos de la población colombiana, frecuencia de uso y demanda atendida. Permitiendo identificar las necesidades de servicios, de recursos, políticas en salud y seguimiento a la gestión institucional entre otros.

- Módulo Nacimientos y Defunciones – RUAF

En el marco del Sistema Integral de Información de la Protección Social – SISPRO, en el año 2008, el Ministerio de la Protección Social desarrolló el Registro Único de Afiliados – RUAF, del cual hace parte el módulo de nacimientos y defunciones. Así mismo, de manera conjunta con el DANE, se rediseño el mejoramiento tecnológico del sistema de Estadísticas Vitales, que trasciende del diligenciamiento de los certificados de nacido vivo y de defunción en físico, a la gestión en tiempo real de las certificaciones en medios magnéticos, buscando

reducir de manera significativa el tiempo de obtención, análisis y producción estadística de datos de los hechos vitales de nacimientos y defunciones (11).

3.7 Minería de Datos

Tradicionalmente, el análisis estadístico ha sido una herramienta que a través de la evaluación de una hipótesis ha permitido dentro del campo de la salud generar conocimiento a favor del cuidado de las enfermedades y de la recuperación de la salud. Mediante el análisis estadístico, se consideran fundamentalmente variables y relaciones primarias, sin tener en cuenta las verdaderas interrelaciones, que por lo general están ocultas y que únicamente se pueden descubrir utilizando un tratamiento de los datos más complejo, que solo es posible con la minería de los mismos.

La minería de datos surge como una tecnología que intenta ayudar a comprender el contenido de una base de datos. De forma general, los datos son la materia prima bruta y en el momento que el usuario les atribuye algún significado especial pasan a convertirse en información. Cuando los especialistas elaboran o definen un modelo y realizan la interpretación de la información se genera un valor agregado que se refiere a conocimiento. En este contexto, la minería de datos emerge como el siguiente paso evolutivo en el proceso de su análisis.

Utilizando la información que los especialistas médicos consideran importante, con la minería de datos se pretende encontrar patrones y relaciones entre las variables, que permitan predecir, los factores que inciden en la supervivencia de los pacientes que padecen ciertas enfermedades.

En las aplicaciones médicas, donde no se puede obviar la importancia del componente temporal, las técnicas de minería de datos han adquirido gran relevancia. Las aplicaciones de estas técnicas van desde la visualización inteligente de grandes cantidades de datos médicos, hasta el control de diferentes indicadores de calidad en instituciones hospitalarias.

Un proceso típico de minería de datos consta de los siguientes pasos generales:

1. Selección del conjunto de datos, tanto en lo que se refiere a las variables objetivo, como a las variables independientes. En esta fase se eliminan o corrigen registros incorrectos y se decide la estrategia con los datos incompletos

2. Análisis de las propiedades de los datos, se utiliza en especial los histogramas, diagramas de dispersión, para identificar la presencia de valores atípicos y ausencia de datos.
3. Transformación del conjunto de datos de entrada, se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema, a este paso también se le conoce como pre procesamiento de los datos.
4. Seleccionar y aplicar la técnica de minería de datos, se construye el modelo predictivo, de clasificación o segmentación.
5. Extracción de conocimiento, mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un pre procesado diferente de los datos.
6. Interpretación y evaluación de datos, una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

Las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados. Las técnicas de minería de datos más utilizadas en el sector salud son:

Modelos de Regresión Logística

Son modelos de regresión que permiten estudiar si una variable dicotómica depende, o no, de otra u otras variables. El objetivo de esta técnica estadística es expresar la probabilidad de que ocurra un hecho como función de ciertas variables que se consideran potencialmente influyentes (12).

Redes bayesianas

Las redes bayesianas permiten construir modelos de minería de datos de tipo probabilístico ya que relacionan un conjunto de variables aleatorias. Basados en el teorema de Bayes y basados en la estimación de las probabilidades se pueden obtener nuevas evidencias de la ocurrencia de una variable (21).

Redes Neuronales

Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Las redes neuronales han sido empleadas en una gran cantidad de aplicaciones, especialmente dentro de la medicina en áreas de análisis de detección de cáncer de mama, análisis de señales de electrocardiogramas y de manera importante en clasificación y segmentación de imágenes médicas (13). Una gran cantidad de trabajos se han dirigido principalmente hacia la detección de cáncer de mama (que es el cáncer más común en países desarrollados), donde se han evaluado diversas arquitecturas con el objetivo de tener la mayor certeza en los diagnósticos médicos.

Algunos ejemplos de la aplicación de la minería de datos en campos de la medicina son, por ejemplo, el proyecto “Development of a new standard for polygraphic sleep analysis” (14), en donde se establecieron pautas en el electroencefalograma realizado durante el sueño (Polisomnografía) para poder establecer de esa forma el diagnostico de las alteraciones del sueño, empleando redes neuronales. Otro trabajo, de aplicación de redes neuronales es “The alarm detection in cardiotocograms (CTG) for fetal monitoring" (15). Realizado en la Clínica Ginecológica de la Universidad de Viena, para la búsqueda de los signos de alarma durante el desarrollo del Monitoreo Cardíaco Fetal de la paciente en el tercer trimestre de embarazo, mediante el empleo de redes neuronales.

Bellaachia y Guven realizaron un análisis de la predicción de la tasa de supervivencia de los pacientes con cáncer de mama usando técnicas de minería de datos (18). Los datos usados fueron tomados de la base de datos del SEER Public-Use Data y correspondieron a 151.886 registros y 16 atributos almacenados en el periodo de 1973 a 2002. Se aplicaron tres técnicas de minería

de datos: árboles de decisión con el algoritmo C45, Naive Bayes y redes neuronales. Los experimentos realizados mostraron que los árboles de decisión con C45 presentan mejor rendimiento y exactitud para predecir la tasa de supervivencia de estos pacientes que las otras dos técnicas. De acuerdo a ello, si el tiempo de supervivencia es mayor o igual a 60 meses, el paciente vive. El 76.8 % de todos los pacientes se clasifica de esta manera. En el caso contrario el paciente muere si la causa de la muerte es cáncer de mama.

Otro proyecto de investigación "La minería de datos aplicada al descubrimiento de patrones de supervivencia en mujeres con cáncer invasivo de cuello uterino" (19), cuyo objetivo fue extraer patrones de supervivencia en mujeres con diagnóstico de cáncer invasivo de cuello uterino, utilizando técnicas de minería de datos a partir de la información almacenada en el Registro Poblacional de Cáncer del Municipio de Pasto (Colombia), durante el periodo de 1998 a 2007 y monitoreada en diferentes bases de datos hasta el 2010. Teniendo en cuenta los resultados obtenidos en la etapa de minería de datos, en la cohorte 1998-2007 tomada para este estudio, el patrón general es que la supervivencia de las mujeres después de haber sido diagnosticadas con cáncer invasivo de cuello uterino sobrepasa los 37 meses. Teniendo en cuenta el número total de casos de cáncer de cuello uterino analizados en este estudio, el 65.9% sobrevive y de estos el 63.5% sobrepasa el umbral de 37 meses de diagnosticado el cáncer.

4. Metodología

4.1 Diseño

Se crearan tres algoritmos de minería de datos que en primer lugar analizara un conjunto de datos, buscando patrones y tendencias específicos. Después, el algoritmo utilizara los resultados de este análisis para definir los parámetros del modelo de minería de datos.

El modelo de minería de datos que crea un algoritmo puede tomar diversas formas, incluyendo:

- Un conjunto de reglas que describen como se agrupan los productos en una transacción.
- Un conjunto de clústeres que describe como se relacionan los escenarios de un conjunto de datos
- Un árbol de decisión que predice si un cliente determinado comprará un producto.

- Un modelo matemático de predicción.

En la minería de datos, se captan y procesan los datos con la esperanza de que de ellos surja una hipótesis apropiada. Se desea que los datos nos describan o indiquen por qué presentan determinada configuración y comportamiento. Como afirma Eduardo Morales: “La más inocente mirada a los datos puede inspirar una hipótesis. Recuérdese que los humanos tienen un gran poder para generalizar e identificar patrones. Luego entonces, validar una hipótesis inspirada por los datos en los datos mismos, será numéricamente significativa, pero experimentalmente inválida” (17).

Es importante insistir, en que las técnicas de minería de datos no pueden utilizarse para confirmar o rechazar hipótesis, porque puede conducir a errores fatales. Su función es otra, se trata de explorar datos, darles sentido, convertir un volumen de datos, que poco o nada aportan a la descripción, en información para interpretar un fenómeno, para adoptar decisiones de acuerdo con las necesidades (4). Sin embargo, la minería de datos si permite formular hipótesis.

Para el desarrollo del estudio se cuenta con la información histórica almacenada en la Bodega de Datos de SISPRO, que realiza la integración de diversas Fuentes, relacionadas con el sector Salud. En la Bodega de Datos de SISPRO están contenidas las bases de datos de la Cuenta de Alto Costo – ERC y del Registros Individual de Prestación de Servicios - RIPS. De la Base de Datos de la Cuenta de Alto Costo – ERC, se filtraran los pacientes con trasplante renal en el año 2009. Esta información será cruzada con la información de RIPS disponible en la Bodega de Datos, para los años 2009 – 2013 y el Registro de Defunciones almacenado en el sistema ND – RUAF del Ministerio de Salud y Protección Social de los años 2010 -2014.

La CAC a través de los procesos de auditoria de la información, garantiza la calidad de la información reportada por las aseguradoras extendiendo la auditoria a Unidades Renales para verificación de consistencia y coincidencia de soportes clínicos aportados por las EPS/EOC. Contiene 48 variables que abarcan los datos de identificación y ubicación del paciente, variables paraclínicas de seguimiento tanto de la evolución de la enfermedad como del manejo según estadio clínico, diagnostico de enfermedad precursoras y de enfermedad renal crónica y costos terapéuticos.

La información almacenada en la Bodega de Datos sobre los RIPS, corresponde a la información reportada por las EPSs al Ministerio de Salud, del régimen contributivo y subsidiado, en el marco de la Resolución 3374 de 2000. Las EPSs

que han reportado información al Ministerio tienen la cobertura del 80% de los afiliados del país. En este registro se identifica al prestador del servicio de salud, al usuario que lo recibe, la prestación del servicio propiamente dicho y el motivo que origino su prestación: diagnóstico, finalidad y causa externa.

En la Bodega de Datos de SISPRO, se encuentra también la información socio económico de la persona, relacionado con el Nivel de SISBEN, esta información es recopilada y procesada por el Departamento Nacional de Planeación - DNP y enviada a la Bodega de Datos de SISPRO.

El cruce de estas bases de datos, permite conocer los factores clínicos y socioeconómicos que pueden estar asociados a la supervivencia de pacientes con trasplante renal. El cruce de información entre las bases de datos se realizó por los campos tipo y número del documento de identificación.

Los campos de la CAC que se utilizaron son:

- Tipo y Número de Identificación
- Régimen
- Fecha de Nacimiento
- Sexo
- Estadio de ERC (KDOQI)
- Fecha de Trasplante
- Hipertensión arterial
- Diabetes Mellitus

De la información de RIPS, se toman los siguientes campos:

- Archivo de Usuarios:
  - Tipo y Número de Identificación
  - Sexo
  - Edad
- Archivo de Procedimientos
  - Tipo y Número de Identificación
  - Código del Procedimiento
  - Diagnostico
- Archivo de Urgencias
  - Tipo y Número de Identificación

    - Diagnostico
- Archivo de Hospitalizaciones
    - Tipo y Número de Identificación
    - Diagnóstico de Ingreso
- Archivo de Consultas
    - Tipo y Número de Identificación
    - Código de la Consulta
    - Diagnóstico

De la información de Estadísticas Vitales – Defunciones, se utilizaran los campos a continuación:

- Tipo y Número de Identificación
- Causa Básica de Muerte
- Fecha de Defunción

Para el proceso de diseño y desarrollo de los modelos de minería de datos se realizó primero la selección, limpieza y transformación de los datos. En esta etapa se realizó el preprocesado y filtrado de los datos (de forma que se eliminan valores incorrectos, no válidos, desconocidos). Otra labor que se realizó en la preparación de los datos fue la construcción de atributos, que consiste en construir automáticamente nuevos atributos aplicando alguna operación o función a los atributos originales con objeto de que estos nuevos atributos hicieran más fácil el proceso de minería. Por último en esta etapa se hizo la selección de atributos relevantes.

4.2 Población y muestra

Población de Referencia:
Pacientes que tienen alguna enfermedad precursora para ERC, afiliadas a las aseguradoras que reportan a la Cuenta de Alto Costo, en el marco de la Resolución 4700 de 2008.

Población de estudio:
Pacientes que recibieron trasplante renal durante el año 2009.

4.3 Variables: Operacionalización de variables

4.3.1 Variables Independientes

- Procedimientos realizados
- Diagnostico
- Régimen
- Genero
- Grupos etarios
- Régimen de afiliación al SGSS
- Nivel de SISBEN
- Causa de Muerte
- Hipertensión Arterial
- Diabetes Mellitus

4.3.2 Variable Dependiente o Desenlace

- Supervivencia a 72 meses de pacientes con Trasplante Renal

4.3.3 Operacionalización de Variables

A continuación se detalla la definición de las variables independientes.

Tabla 4. Operacionalización de las variables

Nombre de la Variable	Definición	Dimensiones	Fuente	Escala de Medición
Procedimientos realizados	La clasificación única de procedimientos en salud (c.u.p.s.) corresponde a un ordenamiento lógico y detallado de los procedimientos e intervenciones que se realizan en Colombia, identificados por un código y descritos por una nomenclatura validada por los expertos del país, independientemente de la profesión o disciplina del		RIPS	Nominal

Nombre de la Variable	Definición	Dimensiones	Fuente	Escala de Medición
	sector salud que los realice así como del ámbito de realización de los mismos.			
Régimen	Tipifica la clase de participación a la cual tiene derecho todo colombiano en el servicio esencial de salud que presta el Sistema General de Seguridad Social en Salud; la participación se puede hacer en condición afiliados al régimen contributivo o subsidiado y de forma temporal como participantes vinculados.		RIPS	Nominal
Género	Se refiere a la condición biológica de un conjunto de seres humanos (diferenciación biológica de constitución entre la mujer y el hombre)		RIPS	Nominal
Nivel Socioeconómico	Es una medida total económica y sociológica combinada de la preparación laboral de una persona y de la posición económica y social individual o familiar en relación a otras personas, basada en sus ingresos, educación, y empleo.	Nivel de SISBEN	SISBEN	Ordinal
Grupos etarios	Clasificación de cada uno de los tipos de edades en que se puede catalogar el tiempo de vida de una persona		RIPS	Nominal
Causa Directa de Muerte	Enfermedad o lesión que inició la cadena de acontecimientos patológicos que condujeron directamente a la muerte, o las circunstancias del accidente o violencia que produjeron la lesión fatal.		EEVV	Nominal
Hipertensión Arterial	Indica si el paciente tiene diagnóstico de Hipertensión arterial. Los pacientes con HTA son los que tienen alguno de los diagnósticos de HTA definidos por CIE10.		CAC	Nominal

Nombre de la Variable	Definición	Dimensiones	Fuente	Escala de Medición
Diabetes	Indica si el paciente tiene diagnóstico de Diabetes Mellitus. Los pacientes con DM son los que tienen alguno de los códigos de diagnóstico de Diabetes Mellitus definidos por CIE10.		CAC	Nominal

4.4 Procedimiento

La mineria de datos no es más que un paso esencial de un proceso más amplio cuyo objetivo es el descubrimiento de conocimiento en bases de datos (del inglés Knowledge Discovery from Databases, KDD). El proceso del KDD se organiza en torno a cinco fases como se ilustra en la Figura 1. Estas fases son iterativas ya que la salida de alguna puede hacer volver a pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad.

Figura 1. Proceso de Extracción del Conocimiento en Bases de Datos

4.4.1 Integración y Recopilación

La Bodega de Datos de SISPRO almacena información histórica del sector salud,

reúne diversas Fuentes que son variantes en el tiempo y están organizadas de tal manera que permiten la toma de decisiones a nivel gerencial y estratégico. Que los datos estén alojados en una Bodega de Datos agiliza los procesos de análisis y la extracción de conocimiento, ya que permite la integración de diversas Fuentes heterogéneas y la manipulación de la información.

La información generada por la Bodega de Datos de SISPRO, se dividió en dos archivos, uno con la información demográfica y los datos de defunción de los pacientes con trasplante renal realizado en el año 2009. El otro archivo contiene las atenciones reportadas en la Fuente del Registro Individual de Prestación de Servicios – RIPS para estos mismos pacientes luego del trasplante. A continuación se detalla la estructura de los archivos entregados:

- Archivo de Información proveniente de la Cuenta de Alto Costo - CAC cruzada con la información del Registro de Estadísticas Vitales – Defunciones, con 4.518 registros y la siguiente estructura:

  PersonaID, Regimen, NivelSisbenDesc, Edad, Sexo, TFG, HTA, DM, FechaTrrID, CausaMuerte, FechaDefuncion, CausaBasicaMuerteA, CausaMuerteB, CausaMuerteC, CausaMuerteD, TiempoSobrevida (Meses), Condicion, SubgrupoDiagnostico

- Archivo de Información de RIPS, con 38.117 Registros y la siguiente estructura:

  Personaid, Fecha, TiempoSobrevida (Meses), Condicion, SubgrupoDiagnostico, CodigoProcedimiento, ProcedimientoDesc, CodigoDiagnostico, DiagnosticoDesc, NumeroAtenciones

4.4.2 Selección, Limpieza y Transformación

En todos los sistemas de información existen problemas de calidad de datos, es por esto que el paso previo para iniciar el proceso de Minería de datos es la preparación óptima de los datos. Para esto se deben identificar los datos incompletos, inconsistentes, con valores no permitidos o irrelevantes para el análisis que se pretende realizar

Las reglas aplicadas durante la fase de limpieza de los datos fueron las siguientes:

- Fecha de defunción debe ser superior a la fecha de trasplante
- Registros con Condición igual a Muerte, deben tener un valor valido para los campos Fecha de defunción y Causa de Defunción.

Aplicando estas reglas se excluyeron 214, para un total de 4.304 registros correctos.

Durante la transformación de atributos se modifica la forma de los datos, para ello algunas variables se transforman en otras, derivan a nuevas variables o cambian el tipo o el rango. Adicionalmente se crearon nuevas variables con el objetivo de reducir la dimensionalidad y obtener un menor número de dominios por variable y de esta manera garantizar un mejor comportamiento en las técnica de minería de datos a aplicar. A partir de la variable causa de muerte que corresponde al estándar de la Clasificación Internacional de Enfermedades - CIE 10 se generó una nueva variable utilizando la lista de causas agrupadas 6/67 de la Organización Panamericana de Salud - OPS obteniendo un conjunto de dominios reducido.

4.4.3 Exploración y selección

Durante esta etapa se aplican técnicas estadísticas que permiten conseguir un entendimiento básico de los datos y muestran las características generales de las variables. Para el análisis de las frecuencias se aplicaron histogramas para la generación de frecuencias que permitan la identificación de datos anómalos que están por fuera de la distribución normal de los datos. Se realizaron análisis bidimensionales para conocer la influencia de las variables seleccionadas en la situación o condición final del paciente (vivo, muerto). Para el análisis de la edad se utilizó un diagrama de cajas-bigotes para visualizar la simetría y dispersión de los datos. Los resultados detallados de estos métodos se pueden ver en el Anexo A.

A partir de las técnicas estadísticas aplicadas se realizaron los siguientes análisis:

Género: Del total de pacientes con trasplante renal realizado en Colombia en el año 2009, el 43 % corresponde a mujeres, de las cuales el 35 % de ellas su condición final durante los primeros 72 meses fue muerto, frente a un 33 % de hombres con esta misma condición final para el periodo evaluado.

Edad: El 39 % de la población con trasplante renal realizado en el año 2009 tenía entre 55 y 73 años al momento del trasplante. El 41% era menor de 55 años y un 18 % mayor de 73 años. Para la condición final muerto al cabo de 6 años, la mayor proporción de pacientes se encontraban entre los 58 y 78 años al momento del trasplante.

Régimen de Afiliación al Sistema de Salud: El 62 % de la población con trasplante renal realizado en el año 2009, pertenece al régimen contributivo, el 36 % al régimen subsidiado y tan solo el 2% corresponde al régimen de excepción o especial.

Nivel de SISBEN: Del total de la población, con trasplante renal realizado en el año 2009, tan solo el 25 % pertenece al Nivel I del SISBEN, el 5 % al Nivel II del SISBEN y la población restante no se encuentra en la encuesta del SISBEN o su clasificación es superior a los niveles I y II del SISBEN.

Hipertensión Arterial: Del total de pacientes con trasplante renal realizado durante el año 2009, el 80% de personas padece de HTA, de las cuales el 30% su condición final fue la muerte, al cabo de los 6 años siguientes.

Diabetes: Del total de la población con trasplante renal realizado en el año 2009 el 38% padece de algún tipo de diabetes, de esta proporción el 47 % de casos terminó en muerte, transcurridos los primeros 6 años.

Condición: El 34% de pacientes con trasplante renal realizado en el año 2009, falleció durante el transcurso de los años 2009 a 2014. El 36 % de esta población falleció durante el año 2012, el 23% en el año 2013, el 20% en el año 2011 y otro 20% falleció en el año 2014.

Sobrevida después del trasplante: El 14% de los pacientes del total de personas con trasplante renal realizado en el año 2009 murió durante los tres primeros años luego de realizado el trasplante. Un 30% de los pacientes, murió durante los cinco primeros años de la realización del trasplante y un 34 % falleció al cabo de los 7 años.

Causa de muerte: Las tres principales causas de muerte de las personas que fallecieron al cabo de los 6 primeros años del trasplante renal realizado durante el año 2009 son: Enfermedades del sistema urinario (21%), Diabetes Mellitus (18%) y Enfermedades isquémicas del corazón (15%).

RIPS - Registro Individual de Prestación de Servicios. Del total de personas con trasplante renal realizado en el año 2009 (4.304), 2.136 personas recibieron algún tipo de atención posterior al trasplante, durante los años 2009 – 2013, que corresponden a 85.814 atenciones reportadas en RIPS. Estos registros pertenecen a atenciones de consultas, hospitalizaciones, procedimientos y urgencias. El promedio de atenciones corresponde a 40,2 atenciones por paciente,

realizadas durante los cinco años siguientes al trasplante, es decir en promedio 8 atenciones al año por paciente.

Las atenciones recibidas por los pacientes con trasplante renal están asociadas principalmente a los diagnósticos: Enfermedades del sistema urinario, Enfermedades Hipertensivas y Diabetes Mellitus.

Los 10 procedimientos principales realizados a esta población fueron:

1. Laboratorio Clínico
2. Consulta, Mediciones Anatómicas, Fisiológicas, Exámenes Manuales Y Anatomopatologicos
3. Otros Procedimientos En Vasos Sanguíneos
4. Imagenologia Radiológica
5. Banco De Sangre Y Medicina Transfusional
6. Procedimientos E Intervenciones En Desempeño Funcional, Rehabilitación Y Relacionados +
7. Imagenologia Con Otras Técnicas No Radiológicas Incluye: Estudios De Imagen Diagnostica Con Ultrasonido O Ultrasonografía O Ecografía, Resonancia Magnética
8. Otros Procedimientos En Pared Abdominal Incluye: Cavidad Pélvica Masculina, Epiplón, Espacio Retroperitoneal, Flancos, Hipocondrios, Mesenterio, Peritoneo, Región Costal, Crural, Epigastrica, Inguinal Y Lumbar Excluye: Piel Y Tejido Celular Subcutánea
9. Procedimientos Profilácticos, Terapéuticos Y Otros Procedimientos Misceláneos + Incluye: Procedimientos E Intervenciones Profilácticos Y/O Terapéuticos No Quirúrgicos Relacionados Con Las Fases De Promoción De La Salud, Prevención De La Enfermedad
10. Procedimientos En Vasos Sanguíneos (Incisión, Escisión Y Oclusión) Simultaneo: Cuando Se Realice Derivación Cardiopulmonar [Circulación Extracorporea] (39.6.1) Excluye: Aquellos De Vasos Coronarios (36)

Selección de las Variables

La vista minable debe recoger toda la información necesaria para realizar la tarea de minería de datos. Para la implementación del modelo de minería a aplicar se definió una vista minable con las siguientes variables:

- Procedimientos realizados

- Diagnóstico
- Régimen
- Genero
- Grupos etarios - Edad
- Régimen de afiliación al SGSS
- Nivel de SISBEN
- Causa de Muerte
- Hipertensión Arterial
- Diabetes

4.5 Construcción del Modelo de Minería de Datos

El modelo de minería de datos es una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos o para explicar situaciones pasadas. Para la definición del modelo es necesaria la selección del tipo de tareas más apropiado para entender mejor una situación, un tipo de modelo y un algoritmo que resuelva la tarea y obtenga el tipo de modelo que este está buscando.

Los modelos pueden ser de dos tipos, los predictivos que pretenden estimar valores futuros o desconocidos de la variable de interés, utilizando otras variables independientes. Los modelos descriptivos identifican patrones que explica o resumen los datos, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos.

4.5.1 Clustering o Agrupamiento

K – means

El algoritmo k - medias (K-means) es un método de agrupamiento por vecindad en el que se parte de un número determinado de prototipos y de un conjunto de ejemplo a agrupar sin etiquetar. Calcula el centroide de cada partición como el punto medio del clúster y asigna cada dato al clúster cuyo centroide sea el más próximo. Los centroides son recalculados para los grupos nuevos y el algoritmo se repite hasta la convergencia, la cual es obtenida cuando no haya más datos que cambien de grupo de una iteración a otra. Para la aplicación del método de agrupamiento en los datos se utilizó el complemento de Minería de Datos para Microsoft Excel, que permite la aplicación del algoritmo K – means, el cual para su funcionamiento requiere parámetros previamente especificados (10) (24).

Para calcular el centroide más cercano a cada punto se debe utilizar una función de distancia. Para datos reales se suele utilizar la distancia euclídea. Para datos categóricos la distancia se calcula como la probabilidad de que un punto de datos pertenezca al clúster.

Los datos fueron divididos en un conjunto de entrenamiento, de 3013 filas y un conjunto de test con 1021 filas. Se realizaron varias iteraciones del algoritmo, encontrando la agrupación óptima al definir el número de conglomerados o clústers igual a 4 (K = 4).

Los dos primeros grupos resultantes están compuestos por pacientes que permanecen vivos a diciembre de 2014 y los dos siguientes grupos a los pacientes que fallecieron durante la serie temporal evaluada. El detalle de los centroides calculados, así como el tamaño de cada uno de los clúster formados se aprecia en la tabla No. 5.

Cluster 1 (34 %): Pacientes con estado final Vivo, edad entre 55 y 72 años, que padecen de HTA y Diabetes Mellitus y pertenecen al Nivel I de la encuesta del SISBEN.

Cluster 2 (33%): Pacientes con estado final Vivo, edad entre 0 y 42 años, que padecen de HTA y no presentan Diabetes Mellitus y pertenecen al Régimen contributivo del Sistema de Seguridad Social en Salud.

Cluster 3 (21%): Pacientes con estado final muerto, entre 73 y 100 años y pertenecen al Régimen contributivo del Sistema de Seguridad Social en Salud, en su mayoría corresponde a hombres.

Cluster 4 (12%): Pacientes con estado final muerto, entre 55 y 72 años y pertenecen al Régimen subsidiado del Sistema de Seguridad Social en Salud, en su mayoría corresponde a mujeres, con Nivel I en la encuesta del SISBEN.

Tabla 5. Tabla de Centroides

Variables	Estados	Población (Todo)	Clúster 1	Clúster 2	Clúster 3	Clúster 4
Tamaño		3013	1016	979	645	373
CausaBasicaDesc_DAN	NULL	1993	100 %	100 %	0 %	0 %
CausaBasicaDesc_DAN	INSUFICIENCIA RENAL CRONICA, NO E	142	0 %	0 %	12 %	17 %
CausaBasicaDesc_DAN	INFARTO AGUDO DEL MIOCARDIO, SIN	127	0 %	0 %	12 %	14 %
CausaBasicaDesc_DAN	DIABETES MELLITUS, NO ESPECIFICAD	110	0 %	0 %	10 %	13 %
CausaBasicaDesc_DAN	ENFERMEDAD RENAL HIPERTENSIVA	97	0 %	0 %	10 %	9 %
CausaBasicaDesc_DAN	ausente	51	0 %	0 %	6 %	4 %
CausaBasicaDesc_DAN	INSUFICIENCIA RENAL TERMINAL	46	0 %	0 %	5 %	3 %
CausaBasicaDesc_DAN	NEUMONIA, NO ESPECIFICADA	30	0 %	0 %	3 %	2 %
CausaBasicaDesc_DAN	...	...	...	...	...	...
Condicion	Vivo	1993	100 %	100 %	0 %	0 %
Condicion	Muerto	1020	0 %	0 %	100 %	100 %
DM	0	1818	39 %	97 %	45 %	49 %
DM	1	1195	61 %	4 %	55 %	51 %
GrupoEtareo	55 - 72	1215	53 %	19 %	43 %	55 %
GrupoEtareo	0 - 42	624	13 %	44 %	3 %	11 %
GrupoEtareo	43 - 54	616	20 %	28 %	11 %	20 %
GrupoEtareo	73 - 105	516	12 %	8 %	42 %	12 %
GrupoEtareo	> 106	41	2 %	1 %	1 %	2 %
HTA	1	2436	86 %	74 %	82 %	83 %
HTA	0	577	14 %	26 %	18 %	17 %
NivelSisbenDesc	NO REPORTADO	1682	36 %	73 %	74 %	35 %
NivelSisbenDesc	NIVEL 1	814	52 %	4 %	10 %	50 %
NivelSisbenDesc	NO APLICA	355	8 %	17 %	12 %	9 %
NivelSisbenDesc	NIVEL 2	162	5 %	6 %	4 %	5 %
RangoSobrevidaDesc	5	1993	100 %	100 %	0 %	0 %
RangoSobrevidaDesc	6	180	0 %	0 %	18 %	17 %
RangoSobrevidaDesc	7	156	0 %	0 %	17 %	13 %
RangoSobrevidaDesc	4	148	0 %	0 %	14 %	15 %
RangoSobrevidaDesc	9	141	0 %	0 %	14 %	13 %
RangoSobrevidaDesc	8	117	0 %	0 %	10 %	14 %
RangoSobrevidaDesc	10	108	0 %	0 %	11 %	11 %
RangoSobrevidaDesc	...	75	0 %	0 %	6 %	9 %
RangoSobrevidaDesc		...	...	...	...	...
Regimen	CONTRIBUTIVO	1848	47 %	73 %	95 %	11 %
Regimen	SUBSIDIADO	1103	52 %	24 %	3 %	87 %
Regimen	EXCEPCION	62	1 %	2 %	3 %	2 %
Sexo	MASCULINO	1695	56 %	58 %	63 %	41 %
Sexo	FEMENINO	1318	44 %	42 %	37 %	60 %
SubgrupoMortalidad	ausente	1993	100 %	100 %	0 %	0 %
SubgrupoMortalidad	610 - ENFERMEDADES DEL SISTEMA U	211	0 %	0 %	20 %	23 %
SubgrupoMortalidad	601 - DIABETES MELLITUS	188	0 %	0 %	18 %	20 %
SubgrupoMortalidad	303 - ENFERMEDADES ISQUÉMICAS DE	159	0 %	0 %	16 %	16 %
SubgrupoMortalidad	302 - ENFERMEDADES HIPERTENSIVA	107	0 %	0 %	11 %	10 %
SubgrupoMortalidad	609 - RESTO DE ENFERMEDADES DEL	63	0 %	0 %	7 %	4 %
SubgrupoMortalidad	307 - ENFERMEDADES CEREBROVASC	52	0 %	0 %	5 %	5 %
SubgrupoMortalidad	108 - INFECCIONES RESPIRATORIAS AI	33	0 %	0 %	4 %	2 %
SubgrupoMortalidad	...	...	...	...	...	...

La relación entre los clúster se visualiza a continuación, Figura 2, aparecen sombreados los clústers con información para la variable condición igual a Muerto, durante los primeros 6 años. Es importante anotar que las agrupaciones 1 y 2 comparten características similares, como son el género y presencia de HTA.

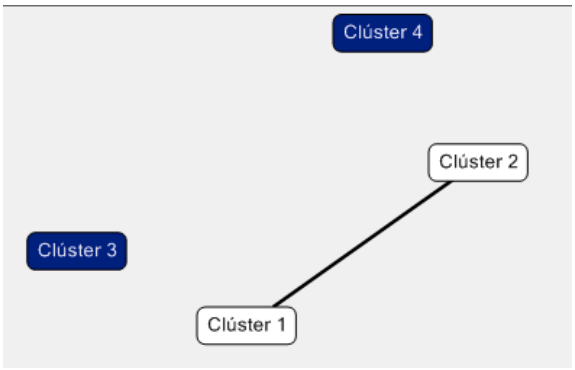


Figura 2. Relación entre los clústers

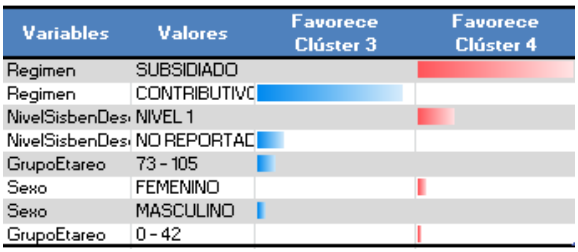
Distinción del clúster

Las agrupaciones 1 y 2 que corresponden a la condición final vivo, transcurrido el periodo evaluado, se diferencian entre sí por las siguientes características únicas para cada clúster: El clúster 1, está conformado por los pacientes con Diabetes Mellitus, que pertenecen al Nivel 1 de SISBEN, conformado en su mayoría por pacientes de edades entre 55 y 72 años. El clúster 2, por pacientes que no presentan Diabetes Mellitus, que no aplican para Nivel de SISBEN, y tienen entre 0 y 42 años. En la Figura 3, se muestran las diferencias más importantes entre los clústeres 1 y 2, mostrando por orden de importancia las características diferenciales más importantes entre los estados de los atributos.

Variables	Valores	Favorece Clúster 1	Favorece Clúster 2
DM	0		
DM	1		
NivelSisbenDesc	NIVEL 1		
NivelSisbenDesc	NO REPORTADO		
GrupoEtareo	0 - 42		
GrupoEtareo	55 - 72		
Regimen	SUBSIDIADO		
Regimen	CONTRIBUTIVO		
HTA	1		
HTA	0		
NivelSisbenDesc	NO APLICA		

Figura 3. Distinción entre los clústers 1 y 2

Los clúster 3 y 4 para pacientes cuya condición final es muerto, se diferencia principalmente por que el clúster 3, reunió a personas del Régimen Contributivo con edades entre 73 y 105 años, y el clúster 4 a pacientes del Régimen Subsidiado y algún porcentaje con Nivel I en la encuesta del SISBEN, tal como se muestra en la Figura 4.



Variables	Valores	Favorece Clúster 3	Favorece Clúster 4
Regimen	SUBSIDIADO		
Regimen	CONTRIBUTIVO		
NivelSisbenDes	NIVEL 1		
NivelSisbenDes	NO REPORTAD		
GrupoEtareo	73 - 105		
Sexo	FEMENINO		
Sexo	MASCULINO		
GrupoEtareo	0 - 42		

Figura 4 Distinción entre los clúster 3 y 4

Los datos analizados son estables y uniformes ya que no se encontraron grandes diferencias entre los clúster formados.

4.5.2 Árbol de Decisión con el algoritmo C45

Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. Cada nodo hoja tiene una etiqueta de clase asociada. Los arboles de decisión utilizan la clasificación como tarea de aprendizaje. Para optimizar el modelo implementado con el algoritmo C45, se aplicó la técnica de análisis de componentes para reducir el número de atributos y un método para el balanceo de clases.

Reducción de la dimensionalidad

La alta dimensionalidad de las variables es un gran problema a la hora de aprender de los datos, ya que pueden existir altos grados de libertad haciendo poco robustos los patrones extraídos. De las variables seleccionadas la variable Causa de Muerte poseía una alta dimensionalidad al estar definida por 49 dominios diferentes. Por esta razón fue necesaria la transformación de los datos iniciales en un nuevo conjunto de datos, mediante la aplicación de la técnica de análisis de componentes principales, que permitió la generación de 15 nuevas variables independientes entre sí y relevantes para la obtención de nuevo conocimiento. Luego de la aplicación de esta técnica la vista minable, insumo para la aplicación del algoritmo quedo definida de la siguiente forma:

- Régimen
- Nivel Sisben
- Edad

- Sexo
- HTA
- DM
- Tiempo SobreVida
- Condición
- 108 - Infecciones respiratorias agudas
- 309 - Las demás enfermedades del sistema circulatorio
- 306 - Insuficiencia cardíaca
- 610 - Enfermedades del sistema urinario
- 701 - Signos, síntomas y afecciones mal definidas
- 606 - Resto de enfermedades del sistema respiratorio
- 601 - Diabetes mellitus
- 106 - Septicemia, excepto neonatal
- 301 - Fiebre Reumática Aguda
- 302 - Enfermedades hipertensivas
- 304 - Enfermedad cardiopulmonar, enfermedades de la circulación pulmonar
- 307 - Enfermedades Cerebrovasculares
- 614 - Resto de las enfermedades
- 303 - Enfermedades isquémicas del corazón
- 609 - Resto de enfermedades del sistema digestivo

A continuación, Figura 5, podemos visualizar el desglose de todos los atributos, con el fin de encontrar los atributos más efectivos para la definición de la clase vivo o muerto. De acuerdo a la gráfica observamos que ningún atributo define separadamente la clase (vivo - muerto).

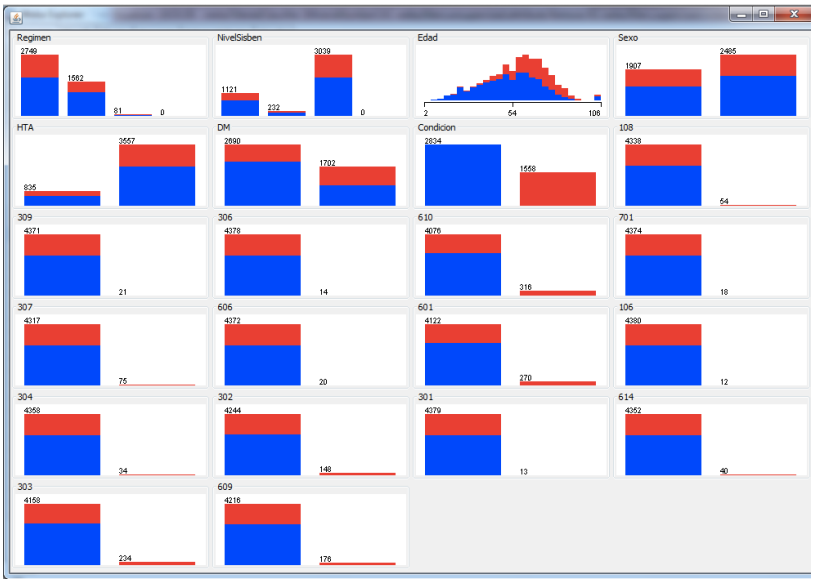


Figura 5. Desglose de las variables para la clase condición final

Balanceo de Clases

Al realizar continuas iteraciones del algoritmo sobre los datos se encontró un bajo porcentaje de aciertos para la clase minoritaria (muerto), comportamiento opuesto al obtenido con la clase vivo. Por esta razón y con el fin de mejorar el número de aciertos sobre la clase minoritaria, fue necesario aplicar el método SMOTE (Syntetic Minority Over-Sampling Technique), incluido en el programa WEKA para re - equilibrar la muestra. Este algoritmo genera instancias "artificiales o sintéticas" para equilibrar la cantidad de datos. Las nuevas muestras se generan realizando interpolación de las muestras originales, favoreciendo en su mayoría a las clases minoritarias.

Luego de la aplicación del método SMOTE se obtuvo un conjunto de datos que favorece a la clase minoritaria (muerto), tal como se aprecia en la Tabla 6:

Tabla 6. Resultado de la aplicación del Método SMOTE

Clases	Datos iniciales	Datos finales luego de la aplicación del método SMOTE
Vivo	2.834	2.834
Muerto	1.470	1.558

Construcción del Modelo

Weka implementa el algoritmo de Arboles de Decisión C4.5 a través del clasificador J48, se utilizó como factor de confianza durante la poda el valor de más alta precisión permitido (2.5), para obtener un árbol aunque de mayor tamaño con un alto porcentaje de aciertos. La variable a predecir es la condición final (Vivo, muerto).

Se generó un árbol de 63 nodos y 34 hijos. Con un porcentaje de casos clasificados correctamente del 97%. El nodo principal del árbol es la causa básica de la muerte: "610 - Enfermedades del sistema urinario", este nodo es de tipo booleano.

Entrenamiento: Se usó el método de validación cruzada con 10 particiones, técnica que incrementa la fiabilidad de los modelos generados, una partición es la división aleatoria del fichero original de datos en otros dos, uno para la etapa de entrenamiento (training) y el otro para la etapa de prueba (test).

```
Number of Leaves  :     34

Size of the tree :      63


Time taken to build model: 3.98 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4258               96.949  %
Incorrectly Classified Instances       134                3.051  %
Kappa statistic                          0.9326
Mean absolute error                      0.1066
Root mean squared error                  0.185
Relative absolute error                 23.2881 %
Root relative squared error             38.6608 %
Total Number of Instances             4392

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.99      0.069      0.963     0.99      0.977      0.971    VIVO
                 0.931     0.01       0.982     0.931     0.956      0.971    MUERTO
Weighted Avg.    0.969     0.048      0.97      0.969     0.969      0.971
```

Figura 6. Resultados del clasificador Arboles de Decisión

Como se observa el modelo clasifica correctamente el 97% de los casos, considerando el resultado obtenido se afirma que el nivel de concordancia es moderado. Igualmente, de acuerdo al valor obtenido de área bajo la curva ROC, se pude afirmar que el modelo tiene un buen nivel de exactitud. Lastimosamente debido al tamaño del árbol hace que la lectura de las reglas sea una tarea compleja.

Para analizar el modelo más a fondo, se extrajeron arboles de decisión más pequeños contenidos en el árbol principal, en los que intervienen las variables edad, Nivel de SISBEN, Sexo y Régimen de Afiliación. Ver Figura 7, 8 y 9.

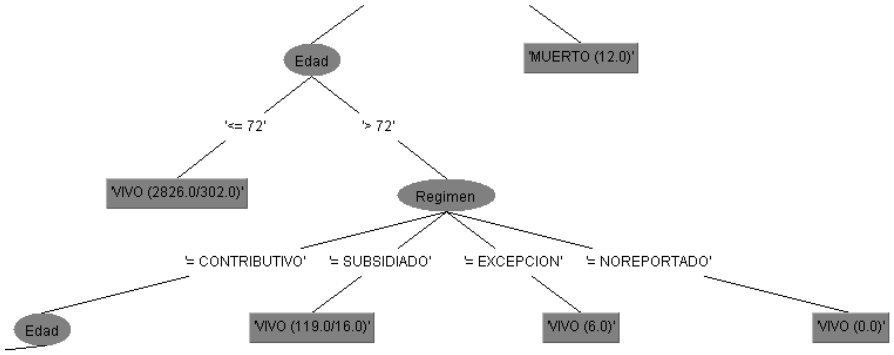


Figura 7. Árbol de decisión Edad, Régimen

Con un alto porcentaje de precisión (90%) los pacientes con una edad igual o menor a 72 años que recibieron trasplante renal, fueron clasificados como vivos. Si el paciente es mayor de 72 años y pertenece al Régimen Subsidiado o Excepción su estado final es Vivo (87% de precisión).

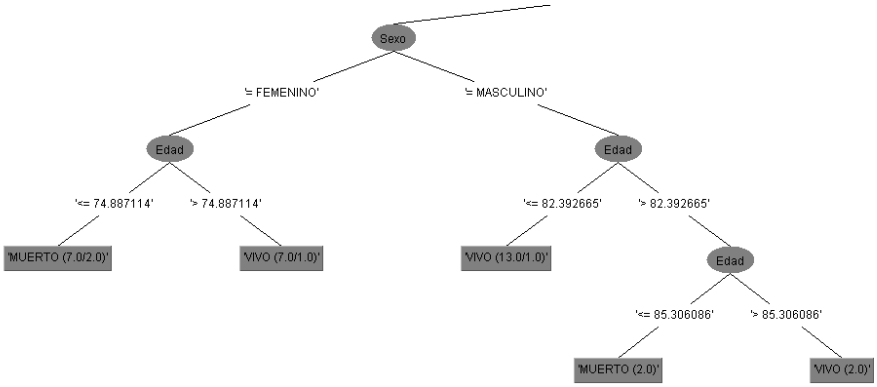


Figura 8. Árbol de decisión Sexo, Edad

Si el paciente que recibió el trasplante renal, es mujer con Nivel SISBEN I y cuya edad es igual o superior a 75 años es clasificado como muerto, si por el contrario la edad del paciente es inferior a 75 años es clasificado como vivo (71% de precisión). Si se trata de un hombre, con Nivel I de SISBEN y cuya edad es menor igual de 82 años la persona con trasplante renal se clasifica como vivo.

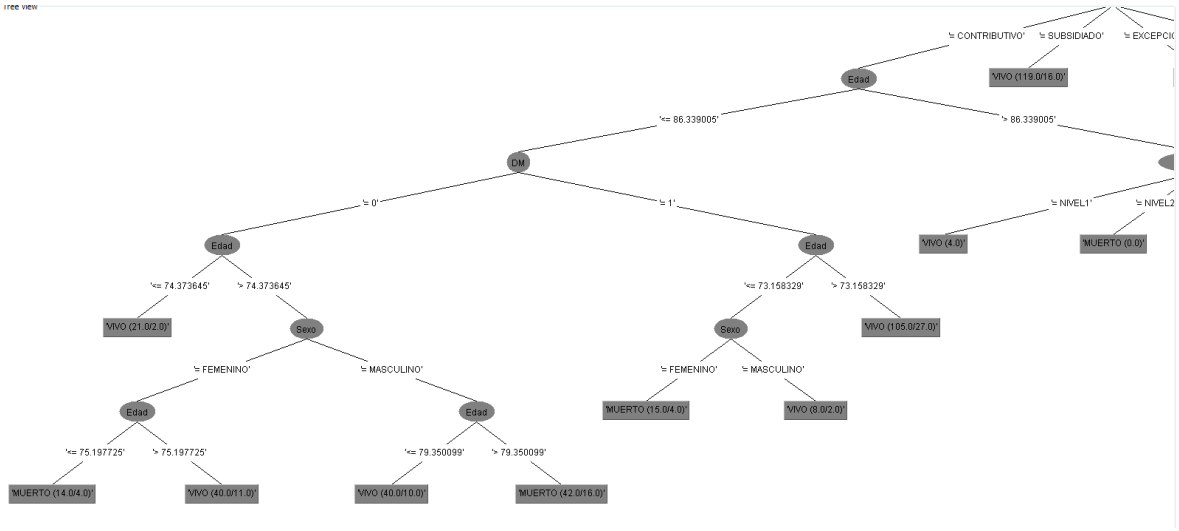


Figura 9. Árbol de decisión Edad, DM – Diabetes Mellitus

Si el paciente con trasplante renal es de sexo femenino presenta Diabetes y su edad es inferior o igual a 73 años se clasifica como muerto. Si presenta Diabetes y tiene entre 74 y 86 años su estado final es vivo (precisión del 75%). Si el paciente no tiene Diabetes y su edad es igual o menor a 74 años se cataloga como vivo.

Validación del Modelo:
Para la validación del modelo se midieron las siguientes escalas: Casos Clasificados correctamente e incorrectamente, estadística kappa, la media de error absoluto, la raíz cuadrada media de error, el error relativo, la raíz cuadrada del error y la media de error absoluto, Área bajo la curva: El área bajo la curva ROC, matriz de confusión: Para cada clasificador se encontró la matriz de confusión. Los parámetros obtenidos se presentan a continuación:

Matriz de confusión:
En la matriz de confusión se pueden observar las instancias clasificadas correctamente, se reconocen porque están en la diagonal, estos son los aciertos y el resto de valores son los errores, es decir, muestra cuantas instancias son predichas a cada uno de los valores posibles de cada clase. A continuación se muestra la matriz de confusión resultante del modelo, Figura 10, en forma general se puede ver que la diagonal claramente tiene los valores bien definidos, es por esto que se puede decir que el modelo es confiable. Para la clase “VIVO” se clasifican correctamente 2.807 instancias y 27 incorrectamente, para la clase “MUERTO” las correctas fueron 1451 y las incorrectas 107. Verdaderos positivos: 2807, falsos negativos: 27, falsos positivos: 107 y verdaderos negativos: 1451.

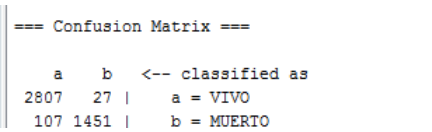


Figura 10 Matriz de confusión del clasificador Árboles de decisión

4.5.3 Redes Neuronales

Las redes neuronales artificiales son un método de aprendizaje cuya finalidad inicial era la de emular los procesadores biológicos de información. Las RNA parten de la presunción de que la capacidad humana de procesar información se debe a la naturaleza biológica de nuestro cerebro. Hay dos tipos principales de aprendizaje en RNA, supervisado y no supervisado. Para la construcción del clasificador se utilizó el método perceptrón multicapa que utiliza el aprendizaje supervisado, para esto se le debe proporcionar a la red un conjunto de datos de entrada y la respuesta correcta. El conjunto de datos de entrada es propagado hacia adelante hasta que la activación alcanza las neuronas de la capa de salida. De esta forma se puede comparar la respuesta calculada por la red con aquella que se desea obtener, el valor objetivo. Para la aplicación de este método se implementó el algoritmo Perceptrón Multicapa de WEKA.

El algoritmo fue aplicado a los casos de trasplante renal, con el objetivo de predecir la relación entre la condición final y las variables dependientes: Causa de muerte, Régimen, NivelSisben, Edad, Sexo, HTA y DM.

Para la implementación del algoritmo se definieron los siguientes parámetros

- Capas Ocultas: Define el número de neuronas en cada capa oculta. Normalmente a más neuronas el coste en tiempo y procesamiento se aumenta ya que tiene que calcular los pesos de más neuronas. Se obtuvo un mejor desempeño del algoritmo al definir una capa oculta con 27 neuronas.

- Tasa de aprendizaje: Este parámetro define cuanto de rápido se desea avanzar en la dirección opuesta al gradiente del error (dirección que minimice el error). Por tanto con una tasa de aprendizaje alta se irá rápidamente hacia el mínimo del error y los pesos de la red se actualizarán más rápidamente. (0.3).

- Momento: Permite a la red ignorar características pequeñas o de mínimo significado en la superficie del error. Si en el momento una red puede

estancarse en un mínimo local, con la definición de este parámetro puede deslizarse y evitar caer en un mínimo local (0.2).

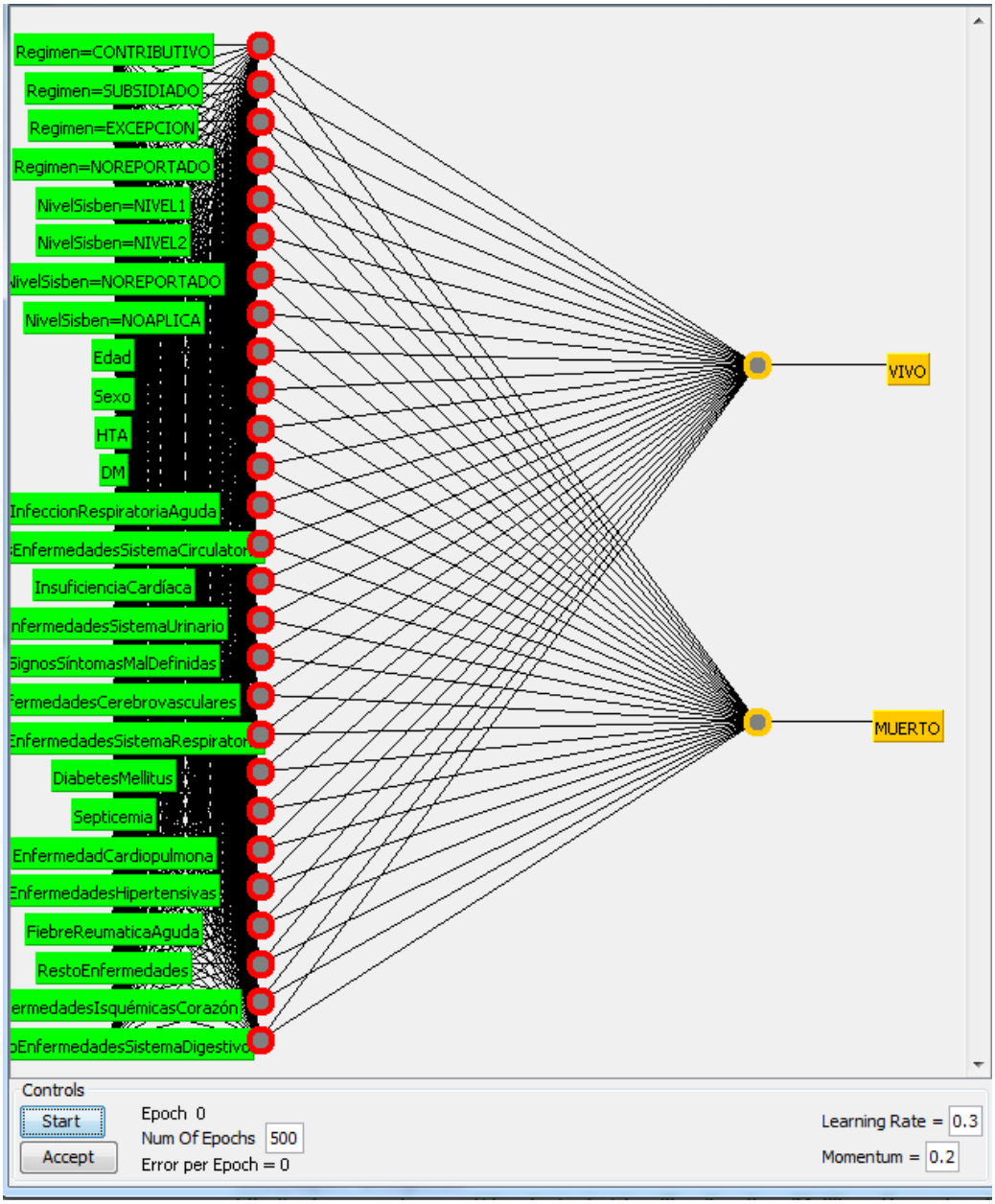


Figura 11. Topología del Perceptrón Multicapa utilizado

Los resultados obtenidos de la aplicación del algoritmo, nos permite identificar fácilmente las variables con una notable relevancia para la predicción del desenlace final, luego de la realización del trasplante renal, para esto se deben comparar los pesos obtenidos para cada una de las variables.

Tabla 7. Selección de variables con Redes Neuronales

Variable	**Peso**
Edad	4,940574928
601 - Diabetes mellitus	4,2759352
Regimen=EXCEPCION	3,232236652
Sexo	2,900964967
NivelSisben=NIVEL2	1,856753971
HTA	1,543279471
301 - Fiebre reumática aguda y enfermedades cardíacas reumáticas crónicas	1,318146641
609 - Resto de enfermedades del sistema digestivo	1,099931219
307 - Enfermedades Cerebrovasculares	0,902335815
106 - Septicemia, excepto neonatal	0,87441519
610 - Enfermedades del sistema urinario	0,856775647
306 - Insuficiencia cardíaca	0,719399936
606 - Resto de enfermedades del sistema respiratorio	0,688363135
701 - Signos, síntomas y afecciones mal definidas	0,622967097
309 - Las demás enfermedades del sistema circulatorio	0,459843535
DM	0,387816495
304 - Enfermedad cardiopulmonar, enfermedades de la circulación pulmonar	0,353730832
108 - Infecciones respiratorias agudas	0,25006233
614 - Resto de las enfermedades	0,180037391
NivelSisben=NOAPLICA	-0,025997542
Regimen=NOREPORTADO	-0,038045028
302 - Enfermedades hipertensivas	-0,097274882
NivelSisben=NOREPORTADO	-0,278464186
NivelSisben=NIVEL1	-0,600071098
303 - Enfermedades isquémicas del corazón	-0,857825487

Variable	Peso
Regimen=SUBSIDIADO	-0,969521691
Regimen=CONTRIBUTIVO	-1,332833579

De acuerdo a los resultados obtenidos las variables que predicen la condición final del paciente con trasplante renal son: la Edad del paciente al momento del trasplante, el diagnóstico que condujo al desenlace final correspondiente a Diabetes Mellitus y estar afiliado al sistema de Salud en el régimen de excepción.

Validación del modelo:
El porcentaje de instancias clasificadas correctamente corresponde a 4.279, es decir un 97.4%. Para la clase vivo, todas las instancias fueron correctamente clasificadas, mientras que para la clase muerto un 7.2 % fue clasificado erróneamente. De acuerdo al valor obtenido del área bajo la curva ROC, se pude afirmar que el modelo tiene un buen nivel de exactitud, Figura12.

```
Time taken to build model: 34.71 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances        4279               97.4271 %
Incorrectly Classified Instances       113                2.5729 %
Kappa statistic                          0.9429
Mean absolute error                      0.0409
Root mean squared error                  0.1516
Relative absolute error                  8.9354 %
Root relative squared error             31.6772 %
Total Number of Instances             4392

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.073      0.962     1         0.98       0.985    VIVO
                 0.927     0          1         0.927     0.962      0.985    MUERTO
Weighted Avg.    0.974     0.047      0.975     0.974     0.974      0.985
```

Figura 12. Resultados de la aplicación del algoritmo Redes Neuronales

Matriz de confusión:
En la matriz de confusión se pueden observar que 113 instancias fueron clasificadas erróneamente, para la clase muerto. En la diagonal de los valores acertados o clasificados correctamente se puede definir la bondad del clasificador. A continuación se muestra la matriz de confusión resultante del modelo, Figura 13.

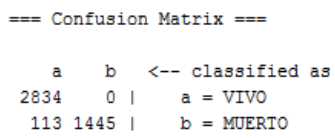


Figura 13. Matriz de confusión del clasificador Redes Neuronales

4.6 Validación de los Modelos

A continuación se presenta los resultados obtenido en los procesos de entrenamiento, validación y análisis de los modelos de clasificación propuestos. Para la validación del modelo se tienen en cuenta varios parámetros, que no son más que medidas de validez estadística, coeficiente Kappa, la precisión del modelo, la sensibilidad y la especificidad. El índice de Kappa muestra la concordancia entre los datos de prueba y la clasificación hecha por el modelo, es decir, cuando el resultado del coeficiente de Kappa es 1, significa que todas las instancias son clasificadas correctamente, es allí donde se dice que tiene máxima concordancia, por el contrario, cuando el valor es igual a cero, la concordancia se debe al azar. La precisión mide el número de términos correctamente reconocidos respecto al total de términos predichos, sean estos verdaderos o falsos términos. (VP/VP+FP). La sensibilidad es la probabilidad de clasificar correctamente a un individuo cuyo estado real sea el definido como positivo respecto a la condición que estudia la prueba (Fracción verdaderos positivos FVP). La especificidad es la probabilidad de clasificar correctamente a un individuo cuyo estado real sea el definido como negativo. Es igual al resultado de restar a uno la fracción de falsos positivos (FFP).

En este trabajo se exponen aquellos modelos para los que se obtuvo un alto nivel de precisión en la clasificación. La tabla a continuación resume las medidas estadísticas arrojados con la aplicación de los dos modelos de clasificación implementados.

Tabla 8. Resumen resultados de los modelos con validación cruzada tomando 10 particiones

Algoritmo	Arboles de Decisión		Redes Neuronales	
Casos Clasificados correctamente	4.258	96.94%	4279	97.4%
Casos Clasificados incorrectamente	134	3.06%	113	2.57%
Estadística de kappa	0.93		0.94	
Media del error absoluto	0.1066		0.04	

Raíz cuadrada del error	0.185		0.15	
Error relativo absoluto	23.281		8.93	
Área bajo la curva ROC	Vivo	Muerto	Vivo	Muerto
	0.971	0.971	0.985	0.985
Precisión	Vivo	Muerto	Vivo	Muerto
	0.963	0.982	0.962	1
Especificidad: VN/(VN+FP) = FVN = 1-FFP	0.93		0.93	
Sensibilidad: VP/(VP+FN) = FVP	0.99		1	

Los porcentajes de acierto de los dos clasificadores son altos, lo cual demuestra que los dos modelos tienen un nivel correcto de predicción. Dada la complejidad en la lectura de las reglas resultantes del árbol de decisión, es más apropiado implementar el algoritmo de Redes neuronales para la identificación de las variables principales que influyen y predicen la condición final del paciente luego del trasplante.

4.7 Análisis de Supervivencia

Se realizó un análisis de supervivencia con las curvas de Kaplan-Meier, con el fin de estimar la supervivencia de los pacientes trasplantados, durante el transcurso de los años 2009 a diciembre de 2014. Para esto se calculó el tiempo de sobrevida o duración de la supervivencia, calculado como la diferencia en meses entre la fecha de realización del trasplante y la fecha de defunción del paciente. A los pacientes que continúan con vida durante el periodo de tiempo observado se les denomina casos censurados. Cabe anotar, que se descartaron los casos con mortalidad temprana, es decir aquellos pacientes que fallecieron antes de los tres primeros meses, que en los datos correspondían a 3 casos.

La supervivencia global encontrada en los datos fue de 97.5%, 86,3%, 77%, 69% y 66% a los 24, 48, 60 y 72 meses de haber recibido el trasplante renal, respectivamente. El tiempo medio de supervivencia fue de 61.6 meses, con un intervalo de confianza del 95% (IC 95) entre 61,1 y 62,13 meses. La Figura 14, muestra el análisis de la sobre vida de los pacientes, transcurridos 72 meses del trasplante.

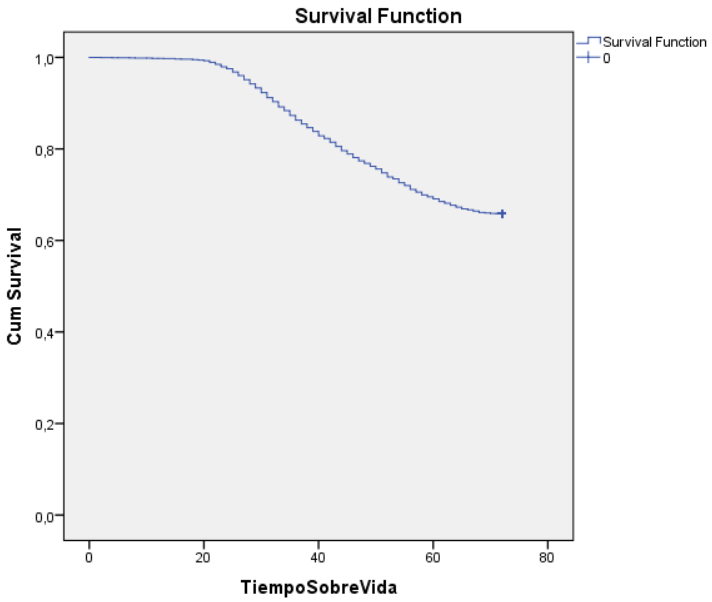


Figura 14. Curva de sobrevida a 72 meses de los pacientes con trasplante renal

Adicionalmente y teniendo en cuenta que la edad del paciente al momento del trasplante representa una variable que influye significativamente en la condición final (vivo o muerto), se incorporó esta variable en los análisis de supervivencia.

En relación a la edad, si se comparan las curvas de supervivencia de cuatro agrupaciones de pacientes, los que recibieron trasplante renal con una edad inferior a 43 años, entre 43 y 54 años, entre 55 y 72 años y mayores a 72 años, la diferencia es claramente significativa, con tasas de supervivencia a los 6 años de 90,9%, 76,5%, 59,5% y 37,1% respectivamente, ver Figura 15.

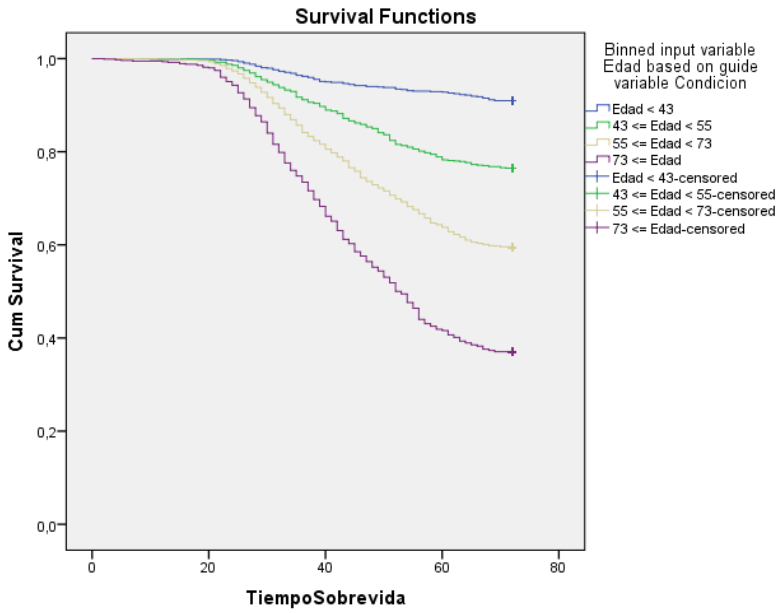


Figura 15. Curva de sobrevida a 72 meses de los pacientes con trasplante renal por agrupaciones de edad.

El porcentaje de pacientes menores de 43 años que sobreviven luego de 6 años del trasplante es el 91%. Para los pacientes entre 45 y 55 años la supervivencia disminuye a 77% en el mismo periodo de seguimiento. Para este grupo la tasa sobrevida a los 4 años es del 85%. Para los pacientes trasplantados entre 55 y 72 años presentan una tasa de supervivencia del 73 % y 59 % a los 4 y 6 años respectivamente.

Como se observa en la Tabla 9, el grupo de edad correspondiente a mayores de 72, presenta la menor probabilidad de supervivencia. A los 3 años la sobrevida corresponde al 73% y luego de 6 años disminuye al 37%, la diferencia se atenúa considerablemente

Tabla 9. Tasa de supervivencia a 72 meses para los pacientes con trasplante renal según el grupo etario

Grupo Etario	3 Años	4 Años	5 Años	6 Años
Edad < 43	96%	94%	93%	91%
43 <= Edad < 55	91%	85%	78%	77%
55 <= Edad < 73	84%	73%	64%	59%
73 <= Edad	73%	55%	42%	37%
Total	86%	77%	69%	66%

Al igual que los resultados obtenidos con la implementación del modelo de Redes Neuronales, el análisis de supervivencia evidencia claramente que la edad al momento de trasplante es un factor fuertemente ligado a la supervivencia del paciente.

5. Resultados

Se implementaron varios algoritos de mineria de datos, que utilizan tecnicas de aprendizaje supervisado y no suprevisado. Los modelos de mineria de datos fueron aplicados sobre la informacion demografica de los pacientes que recibieron trasplante renal en el año 2009 y sobre la informacion de las atenciones reportadas para estos pacientes en el RIPS – Registro Individual de Prestacion de Servicios. En el presente trabajo se exponen aquellos modelos que obtuvieron una mayor precisión a la hora de predecir la variable objetivo. Los algoritmos que fueron aplicados sobre la información de la Fuente de RIPS – Registro Individual de Prestación de Servicios de Salud, arrojaron un alto porcentaje de casos clasificados erróneamente, por tal razón, el detalle de la implementación de los algoritmos aplicados a esta información no fue expuesta en el presente trabajo, de lo cual se puede concluir que el número y tipo de atenciones recibidas luego del trasplante no tiene ninguna relación con la condición final del paciente.

Los modelos de mineria de datos que presentaron una mayor nivel de precision y exactitud en la prediccion de la variable
Se implementaron tres algoritmos de minería de datos, el primer algoritmo implementado fue el cluster o agrupamiento, que utiliza tenicas de aprendizaje no supervisado y otros dos algoritmos supervisados definidos como clasificadores, Arboles de decisión y Redes Neuronales. Para el diseño e implementacion de los modelos de mineria de datos se siguieron las etapas para la generación de conocimiento definidas en la literatura.

Al aplicar el algoritmo de agrupamiento o clúster se evidenció que los datos analizados son estables ya que no se encontraron grandes diferencias entre los clúster formados que definen la variable objetivo. Se generaron 4 clusters , los dos primeros corresponden a pacientes cuya condición final transcurrido 72 meses del trasplante es vivo y las 2 siguientes agrupaciones cuya condición final es muerto. Las variables prioritarias de agrupacion para los cluster 1 y 2 son: la

presencia de Diabetes Mellitus, Nivel de SISBEN y Edad. Los clúster 3 y 4 para pacientes cuya condición final es muerto, se diferencia principalmente por que el clúster 3, reunió a personas del Régimen Contributivo con edades entre 73 y 105 años, y el clúster 4 a pacientes del Régimen Subsidiado y algún porcentaje con Nivel I en la encuesta del SISBEN.

La aplicación del clasificador J48 generó un árbol con 63 nodos y 34 hijos, con un porcentaje de casos clasificados correctamente del 97%. El nodo raíz del árbol, que es el atributo a través del cual se empiezan a dividir los datos, es la causa básica de la muerte, "610 - Enfermedades del sistema urinario". Los resultados obtenidos permiten concluir que el nivel de concordancia es moderado y que el modelo tiene un buen nivel de exactitud. Debido al tamaño del árbol, la lectura de las reglas de clasificación se convirtió en una tarea compleja, y se presentaron contradicciones al evaluar las reglas creadas.

También se utilizó un algoritmo de redes neuronales, implementando el método perceptrón multicapa que utiliza el aprendizaje supervisado. El porcentaje de instancias clasificadas correctamente, luego de la aplicación y parametrizacion de este algoritmo, correspondió a un 97.4%. Para la clase vivo, todas las instancias fueron correctamente clasificadas, mientras que para la clase muerto un 7.2 % fue clasificado erróneamente. De acuerdo al valor obtenido se pude afirmar que el modelo tiene un buen nivel de exactitud. Los resultados obtenidos de la aplicación del algoritmo, nos permite identificar fácilmente las variables relevantes en la predicción del desenlace final trascurridos 72 meses de la realización del trasplante renal, estas variables son: la edad del paciente al momento del trasplante, la causa básica de muerte correspondiente a Diabetes Mellitus y estar afiliado al sistema de Salud en el régimen de excepción

Al evaluar los resultados obtenidos de la aplicación de los algoritmos que obtuvieron un nivel adecuado de precision y exactitud se puede concluir que la tecnica más apropiada para realizar este tipo de analisis son las redes neuronales artificiales, aplicando el método perceptron multicapa. El cual demostró que existe una relación fuerte entre la edad al momento del trasplante y la presencia de Diabetes Mellitus en la condición final del paciente. Permitieron igual concluir que factores como el género, nivel de SISBEN (característica socio económica), la morbilidad presentada luego del trasplante y los procedimientos realizados no ejercen ninguna influencia sobre el desenlace del paciente con trasplante renal.

6. Discusión

Al revisar la literatura de los últimos años, se encuentran un sinfín de factores ligados a la supervivencia de los pacientes sometidos a terapia renal sustitutiva, en muchas ocasiones describiéndose un factor en unos trabajos como protector y en otros como perjudicial, tal es el caso del estudio titulado "Factores predictores de supervivencia en terapia sustitutiva de la insuficiencia renal terminal (IRT)" de C. Díaz Corte, F. Ortega, P. Rebollo (25), donde se expone que la supervivencia global de la muestra evaluada, que corresponde a personas con terapia sustitutiva es de 90%, 82%, 61% y 50%, a los 12, 24, 60 y 96 meses, respectivamente. El tiempo medio de supervivencia fue de 80 meses con un intervalo de confianza del 95% (IC 95) entre 75 y 85 meses. El estudio concluye que "Las variables que tienen influencia significativa respecto a la supervivencia son la edad al inicio de la terapia sustitutiva, la DM como causa de la IR, la patología concomitante, la Situación Funcional, el hecho de haber recibido un trasplante (especialmente si su duración fue superior a un año) y los niveles de hemoglobina y de creatinina pre diálisis a los tres meses del inicio de la terapia sustitutiva", que concuerda con la conclusión del presente estudio. De los resultados de nuestro estudio quizá enfatizar que lo que parece realmente importante es tener menos de 43 años y que las diferencias se atenúan considerablemente en los rangos de edad de los pacientes con más de 55 años.

Otro estudio que evalúa el tiempo de sobrevida de los pacientes con trasplante renal es, el titulado "Sobrevida de los pacientes trasplantados de riñón en un hospital público de chile", de Sandra Mardones Carreño (26), donde se concluye que, "La sobrevida global del injerto desde 1978 a 2009, es de 94% al año y de 88% a los 5 años", estas tasas difieren de los encontrados en el presente estudio, 90% en el primer año y 69% a los cinco años de realizado el trasplante renal.

Al igual que en el estudio titulado Socioeconomic deprivation is independently associated with mortality post kidney transplantation, en donde se obtuvieron datos de 19.103 procedimientos de trasplante renal, realizados en Inglaterra a partir de abril de 2001 a marzo de 2012, en el que se analizaron los siguientes datos demográficos del paciente: edad, sexo, tipo de donante (vivo o muerto), la etnia, el año del trasplante, el fracaso del injerto, comorbilidad médica y el nivel socioeconómico, en este trabajo también se concluyó que el nivel de privación socioeconómica se asocia de forma independiente con una mayor mortalidad

7. Conclusiones y Recomendaciones

En el proceso de implementación de minería de datos se hace necesario seguir cada uno de los pasos del proceso de extracción de conocimiento, la fase de preparación de los datos es de gran importancia ya que puede representar el éxito o fracaso en la implementación de las técnicas de minería de datos, para esto se pueden utilizar procedimientos que permiten la depuración y transformación de los datos.

La aplicación del modelo de clúster se realizó con un complemento para minería de datos de Microsoft Excel, lo que evidenció que este complemento que es de uso frecuente en las organizaciones provee una gran herramienta a los usuarios para el análisis de la información y la toma de decisiones. Excel, mediante este complemento proporciona un conjunto de utilidades que permite realizar desde la limpieza y depuración de los datos hasta la generación de modelos de predicción.

Los resultados obtenidos a traves de la técnica de clasificación por redes neuronales permiten corroborar el éxito de la aplicación de este tipo de métodos matemáticos en el campo de la medicina, ya que los resultados obtenidos se ajustan bastante a la realidad. La aplicación de los modelos de minería de datos sobre las variables evaluadas provenientes de diferentes Fuentes de Información permitió la selección de las variables relevantes que influyen en la predicción final del paciente con trasplante renal.

Aunque la minería de datos no permite comprobar una hipotesis, si premite su formulación y planteamiento y está dirigida a la exploración de los datos en la busqueda de patrones que generen conocimiento. De acuerdo a los resultados obtenidos, la edad del paciente al momento del trasplante renal representa una variable que influye significativamente en la condición final (vivo o muerto) transcurrido el periodo evaluado, es por esta razón que esta variable fue incorporada en un análisis de supervivencia utilizando el método de Kaplan-Meier. El resultado de este análisis confirmó el resultado obtenido en la aplicación del algoritmo de redes neuronales, de esta forma se ratificó el planteamiento, en donde la minería de datos no es vista como un reemplazo de la estadística tradicional, sino como una extensión de la misma.

8. Futuros Desarrollos

Como futuras investigaciones complementarias al desarrollado del presente trabajo, se podrían implementar técnicas de minería de datos sobre el reporte de la información de la cuenta de alto costo para el año 2014, ya que para este periodo el número de variables a reportar aumento. Estas nuevas variables se encuentran definidas en la Resolución 2463 de 2014, donde se dan los lineamientos para el reporte de la información de Enfermedad Renal Crónica – ERC y sus precursoras Hipertensión Arterial y Diabetes Mellitus. Dentro de las variables adicionadas se incluyen, el tipo de donante, causa de muerte, utilización de medicamentos inmunosupresores, peritonitis, etiología de la ERC, pertenencia étnica, grupo poblacional y municipio de residencia, que permitirían encontrar otros factores para la predicción de la supervivencia del paciente con trasplante renal.

Bibliografía

(1) Halloran PF. Immunosuppressive drugs for kidney transplantation. N Engl J Med 2004 Dec;351(26): 2715-29. [Consultado 20 Febrero 2015] Disponible en: http://www.ncbi.nlm.nih.gov/pubmed/15616206

(2) T. Kälble, A. Alcaraz, K. Budde, U. Humke, G. Karam, M. Lucan, G. Nicita, C. Süsal. Guía clínica sobre el trasplante renal. 2009 mar: 1164 -1165. [Consultado 22 Febrero 2015] Disponible en: http://www.uroweb.org/gls/pdf/spanish/15-%20GUIA%20CLINICA%20SOBRE%20TRANSPLANTE%20RENAL.pdf pag 1164

(3) Martín P, Errasti P. Trasplante Renal – Kidney Transplant. 2006 pag 83 [Consultado 10 Abril 2015] Disponible en: http://www.sld.cu/galerias/pdf/sitios/trasplante/trasplante-renal.pdf

(4) Febles JP, González A. Aplicación de la minería de datos en la bioinformática. ACIMED (revista en internet). 2002; 10(2):69-76. [Consultado 10 Abril 2015]; Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024 -94352002000200003&lng=es&nrm=iso

(5) Harrison T. Principios de Medicina Interna. Parte XI. Trastornos del riñón y las vías urinarias. Editorial Mc Graw Hill.16a edición. 1824-1834

(6) Guía para el manejo de la enfermedad renal crónica – ERC - Basada en la Evidencia, Bogotá: Ministerio de la Protección Social y FEDESALUD; c2013. [Consultado 9 Diciembre 2014] Disponible en: Http://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/VSP/Gu%C3%ADa_para_el_manejo_de_la_enfermedad_renal_Cronica_ERC.pdf

(7) Coresh J, Wei GL, Maquillan G, et al. Prevalence of high blood pressure and elevated serum creatinine level in the united states: Findings from the Third National Health in the United States Survey, 1.988 - 1.994. Arch Intern Med 2.001; 161: 1207 - 1216.

(8) Rivera A, et al. Enfermedad Renal Crónica. En manejo integral de la Enfermedad Renal Crónica en Colombia. Rivera A et al, eds, Primera Edición,k Editorial Alkianza Producciones, 2009;29-39.

(9) Fondo colombiano de enfermedades de alto costo. Situación de la Enfermedad Renal en Colombia, Bogotá: Cuenta de Alto Costo. c2011. [Consultado 10 Diciembre 2014] Disponible en: http://www.cuentadealtocosto.org

(10) Microsoft Clustering Algorithm Technical Reference, Estados Unidos: Microsoft. c2015. [Consultado 17 Mayo 2015] Disponible en: https://msdn.microsoft.com/en-us/library/cc280445.aspx

(11) Nacimientos y Defunciones RUAF, Bogotá: Ministerio de Salud y Protección Social. C2013. [Consultado 10 Diciembre 2014]; Disponible en: http://www.sispro.gov.co/recursosapp/ruaf/pages/AplicaciónNacimientosyDefunciones-RUAF.aspx.

(12) Silva Ayçaguer. Excursión a la regresión logística en ciencias de la salud Excursión a la regresión logística en ciencias de la salud. Ed. Díaz de Santos. Madrid. 1ª Edición 1995

(13) Bello Maria Cristina. Creación de un DataWareHouse Financiero, y su Apoyo en la Gestion de la Compañía Universidad Icesi. Cali

(14) Porta Carlos Alberto,Development of a new standard for polygraphic sleep analysis‖. Proyecto SIESTA. Madrid 2005

(15) Porta Carlos Alberto. The alarm detection in cardiotocograms (CTG) for fetal monitoring. Buenos Aires 1997

(16) Gómez EG, Hernández JP, López FJ, Garcia JR, Montemayor VG, Curado FA, Vallejo ML. Long-term allograft survival after kidney transplantation. Elsevier INC, 2013; 45

(17) Stephens, M. R., Evans, M., Ilham, M. A., Marsden, A. and Asderakis, A. (2010), The Influence of Socioeconomic Deprivation on Outcomes Following Renal Transplantation in the United Kingdom. American Journal of Transplantation, 10: 1605–1612. doi: 10.1111/j.1600-6143.2010.03041.x

(18) Bellaachia A, Guven E. Predicting breast cancer survivability using data mining techniques. George Washington University. Washington D.C; 2005.

(19) Timarán R, Yépez M. La minería de datos aplicada al descubrimiento de patrones de supervivencia en mujeres con cáncer invasivo de cuello uterino. Rev Univ. Salud; 2012. Dec; [Consultado 10 Febrero 2015]; Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0124-71072012000200002&lng=en.

(20) Registro Individual de Prestación de Servicios, Bogotá: Ministerio de Salud y Protección Social. c2013. [Consultado 10 Mayo 2015]; Disponible en: http://www.sispro.gov.co/recursosapp/ruaf/pages/rips.aspx.

(21) López F R. Sistema para descubrir el comportamiento de los factores de riesgo en cáncer de cuello uterino utilizando técnicas de minería de datos. Rev. Univ. Nacional. 2011 Dic. [Consultado 15 Marzo 2015]; Disponible en: http://www.bdigital.unal.edu.co/3896/1/299511.2011.pdf

(22) Morales E. Descubrimiento de conocimientos en bases de datos. 2009. [Consultado 20 Abril 2015]; Disponible en: http://w3.mor.itesm.mx/~emorales/Cursos/KDD/node9.html

(23) Flores Juan C, Alvo Miriam, Borja Hernán, Morales Jorge, Vega Jorge, Zúñiga Carlos et al . Enfermedad renal crónica: Clasificación, identificación, manejo y complicaciones. Rev. méd. Chile [revista en Internet]. 2009 Ene; 137(1): 137-177. [Consultado 10 Mayo 2015]; Disponible en: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0034-98872009000100026&lng=es.

(24) Asistente para clúster (Complementos de minería de datos para Excel). Estados Unidos: Microsoft. c2015. [Consultado 24 Mayo 2015]; Disponible en: https://msdn.microsoft.com/es-es/library/dn282355(v=sql.120).aspx

(25) Díaz Corte C, Ortega F, Rebollo P. Factores predictores de supervivencia en terapia sustitutiva de la insuficiencia renal terminal (IRT). Rev Nefrología. Vol XIX. Número 3. [Revista en internet]. 1999. [Consultado 18 Mayo 2015]; Disponible en: http://www.revistanefrologia.com/es-publicacion-nefrologia-articulo-factores-predictores-supervivencia-terapia-sustitutiva-insuficiencia-renal-terminal-irt--X0211699599011407

(26) Mardones Carreño S. Sobrevida de los pacientes trasplantados de riñón en un hospital público de chile. 2013 Septiembre [Consultado 18 Mayo 2015]; Disponible en: http://www.saludpublicachile.cl:8080/dspace/bitstream/123456789/386/1/Tesis%20Sobrevida%20del%20Trasplante%20Renal.pdf

ANEXOS

ANEXO A

Análisis estadístico realizado durante el pre procesamiento de los datos

Durante la etapa de exploración y selección de los datos se realizó un análisis estadístico de las variables para obtener un entendimiento básico de los datos.

a) Estadística descriptiva de las variables continuas, características generales de las variables Edad, tasa de filtración glomerular y tiempo de sobrevida.

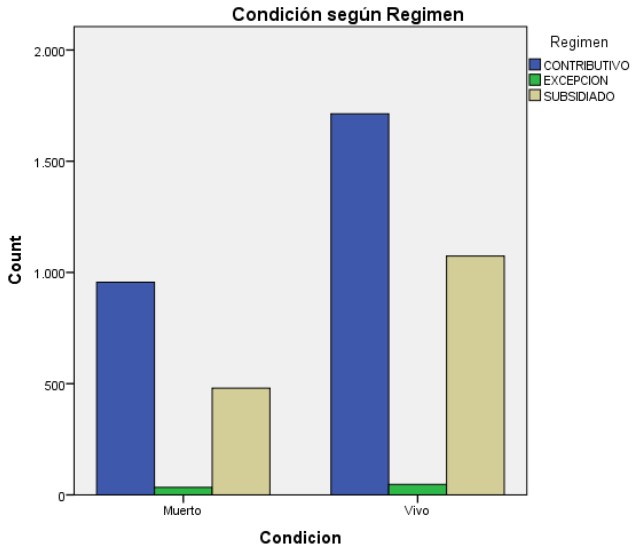
Statistics

	N									
	Valid	Missing	Mean	Std. Error of Mean	Median	Mode	Std. Deviation	Variance	Minimum	Maximum
Edad	4304	0	56,75	,273	58,00	61	17,889	320,028	2	106
TFG	4288	16	12,12	,34	5,79	0,00	22,42	502,52	0,00	125,83
TiempoSobrevida	1470	2834	41,68	,339	40,00	31[a]	12,999	168,986	0	72

a. Multiple modes exist. The smallest value is shown

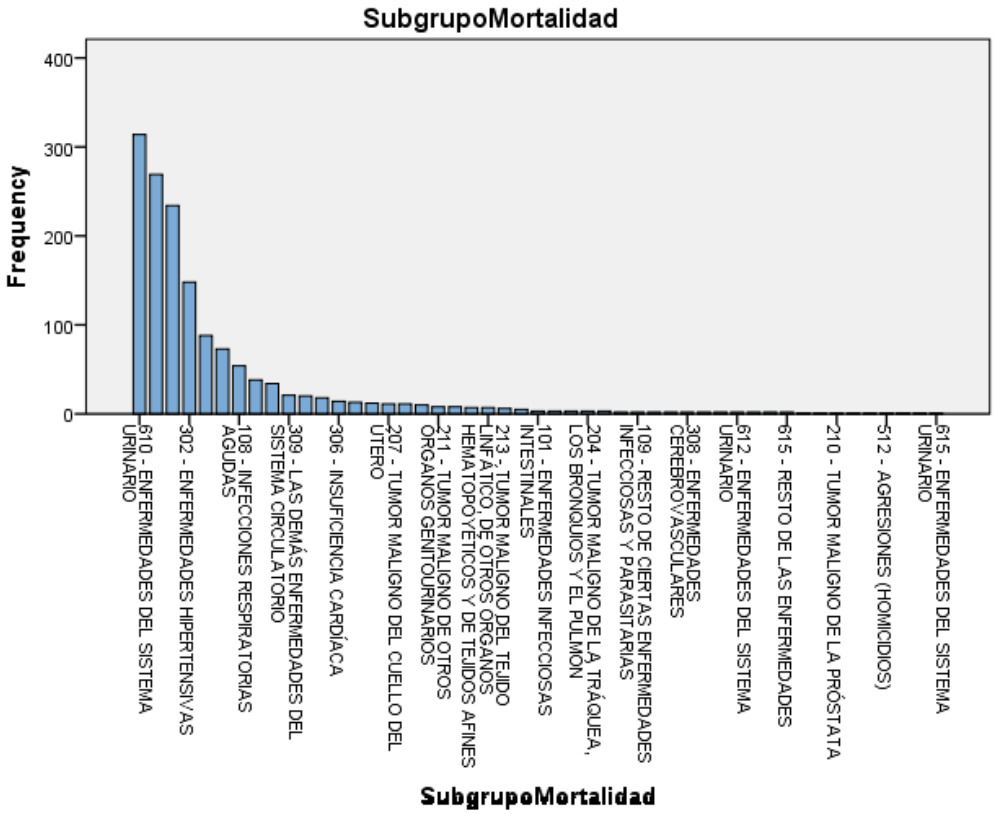
b) Régimen de Afiliación

Condición Final de los pacientes con trasplante renal realizado en el año 2009 en el transcurso de los primeros 72 meses, según Régimen de Afiliación.



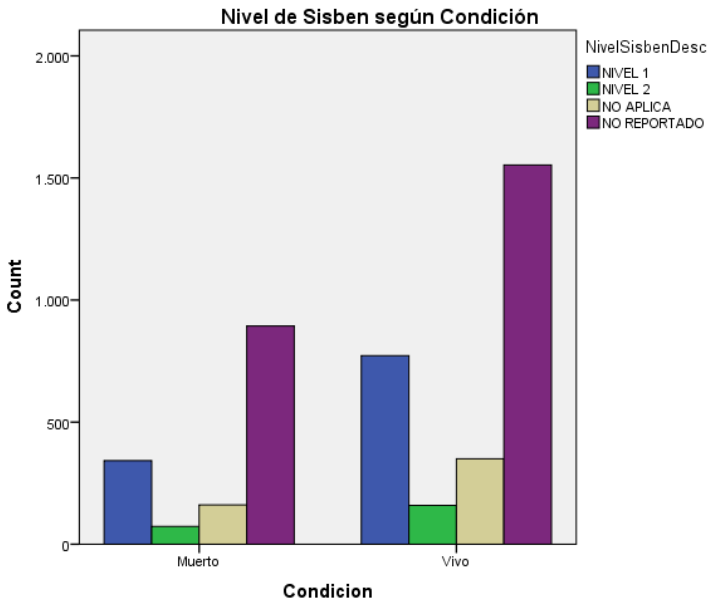
c) Sub Grupo Mortalidad

Condición Final de los pacientes con trasplante renal realizado en el año 2009 en el transcurso de los primeros 72 meses, según la causa directa de defunción agrupada.



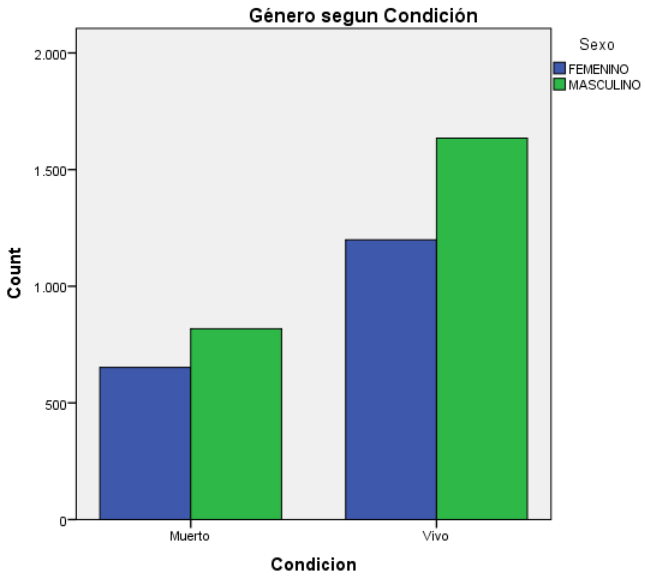
d) Nivel de SISBEN

Condición Final de los pacientes con trasplante renal realizado en el año 2009 en el transcurso de los primeros 72 meses, según el Nivel asignado en la Encuesta SISBEN.



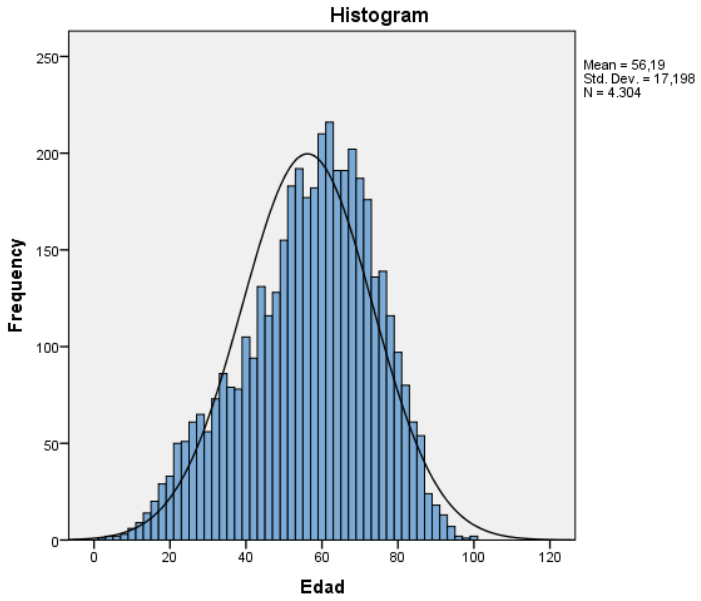
e) Genero

Condición Final de los pacientes con trasplante renal realizado en el año 2009 en el transcurso de los primeros 72 meses, según género.

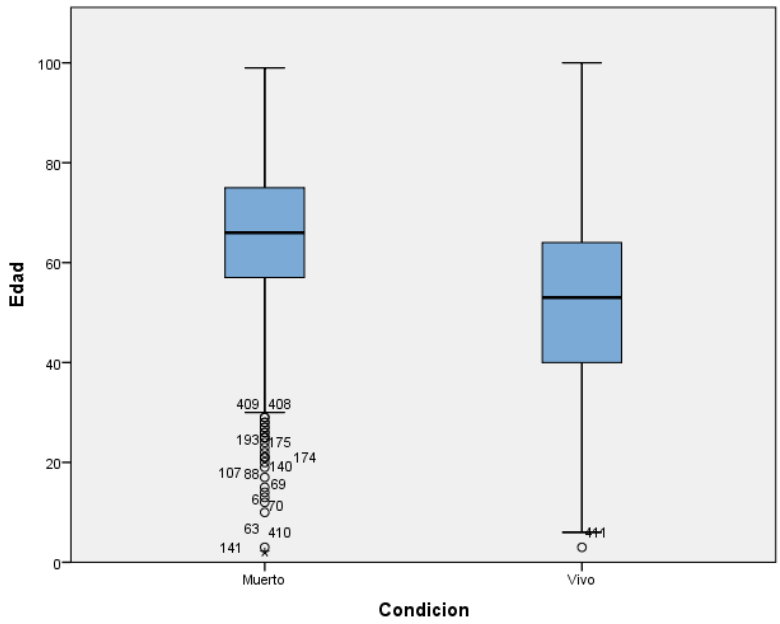


f) Edad

Histograma de la edad del paciente al momento del trasplante renal, 2009

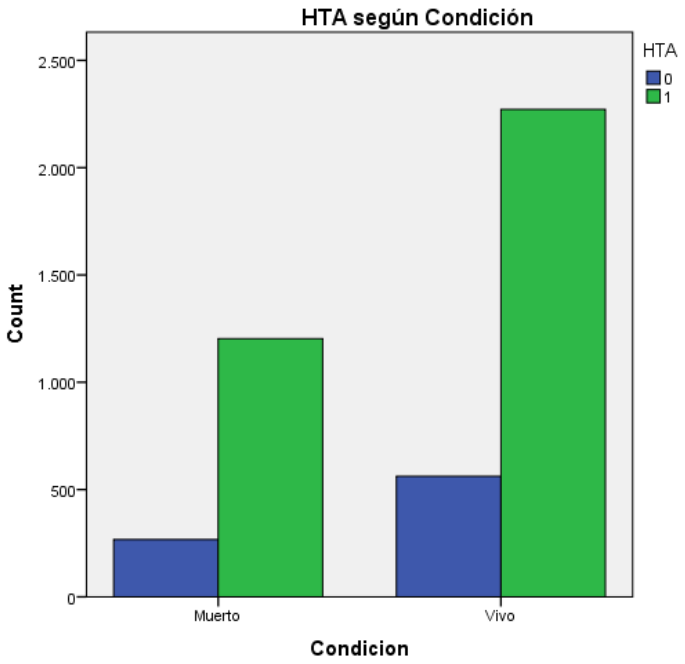


Edad del paciente al momento del trasplante renal realizado en el año 2009, según la condición final luego de los primeros 72 meses.



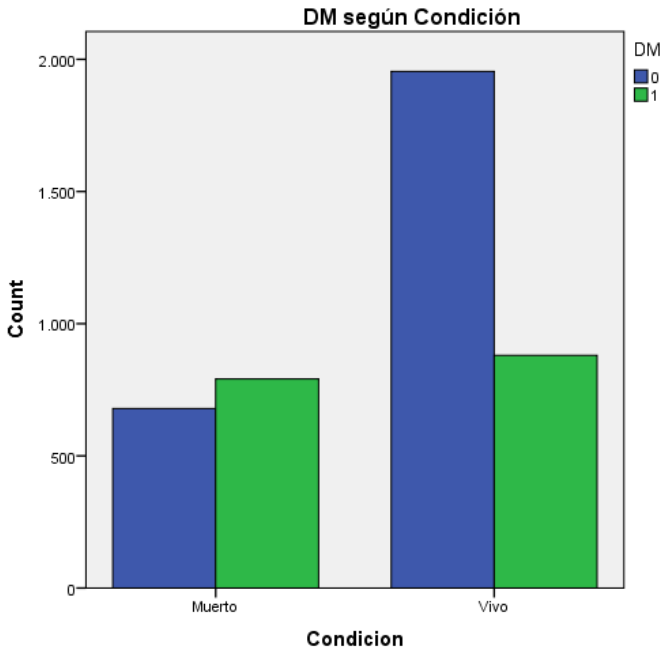
g) Hipertensión Arterial

Condición Final de los pacientes con trasplante renal realizado en el año 2009 en el transcurso de los primeros 72 meses, según presencia de Hipertensión arterial.



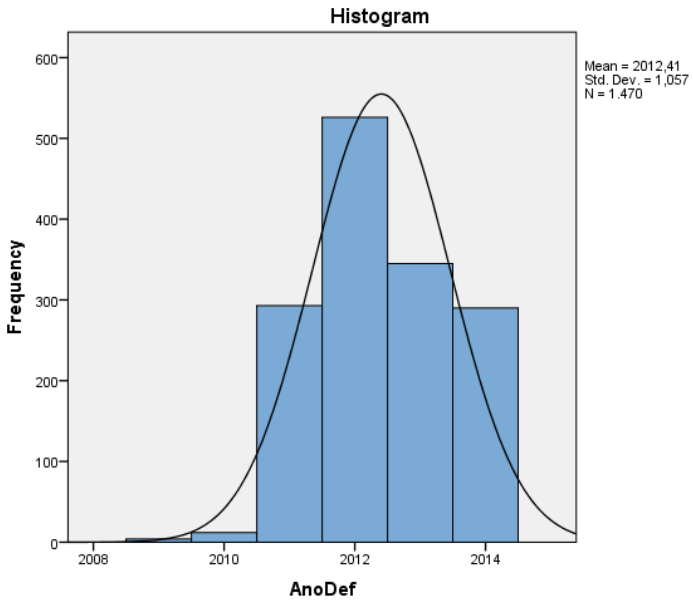
h) Diabetes Mellitus

Condición Final de los pacientes con trasplante renal realizado en el año 2009 en el transcurso de los primeros 72 meses, según presencia de Diabetes Mellitus.



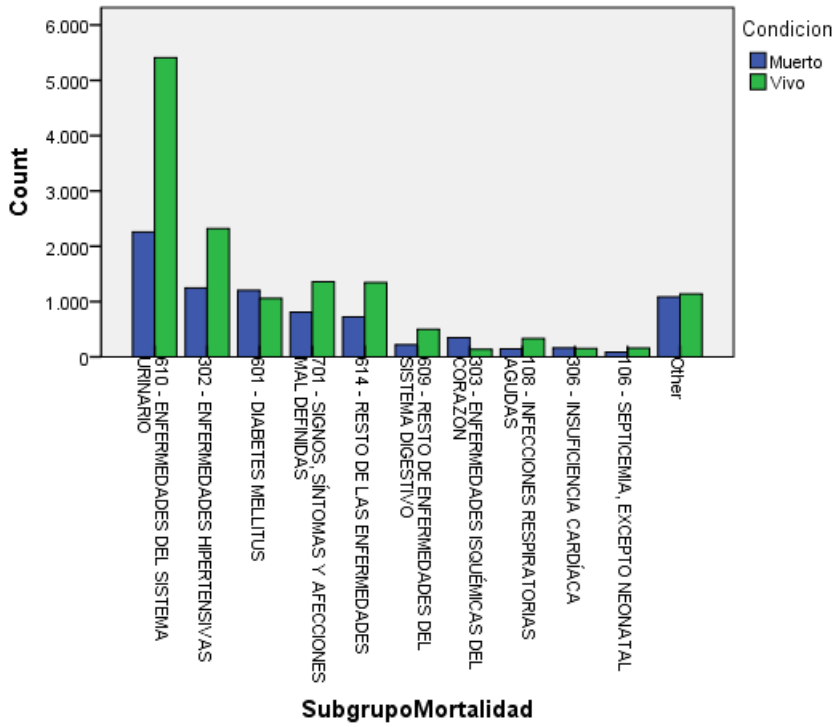
i) Año de la Defunción

Histograma de la variable año de defunción de los pacientes con trasplante realizado en el año 2009



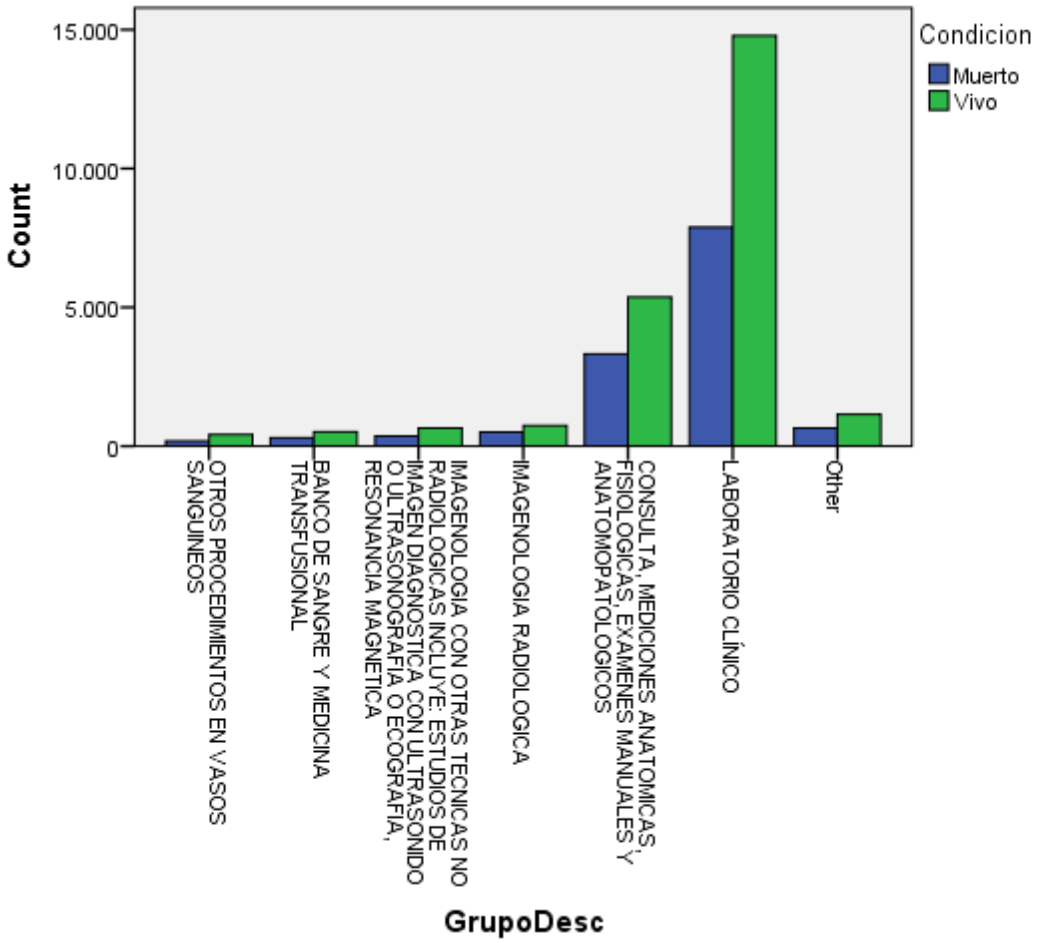
j) Sub Grupo Diagnostico de la Atención - RIPS

Condición Final de los pacientes con trasplante renal realizado en el año 2009 en el transcurso de los primeros 72 meses, según los 10 principales diagnósticos reportados en el Registro Individual de Prestación de Servicios – RIPS



k) Procedimiento CUPS – RIPS

Condición Final de los pacientes con trasplante renal realizado en el año 2009 en el transcurso de los primeros 72 meses, según los 6 principales procedimientos realizados, reportados en el Registro Individual de Prestación de Servicios – RIPS


Condicion
Muerto
Vivo
15.000
10.000
5.000
0
Count
OTROS PROCEDIMIENTOS EN VASOS SANGUINEOS
BANCO DE SANGRE Y MEDICINA TRANSFUSIONAL
IMAGEN DIAGNOSTICA CON ULTRASONIDO O ULTRASONOGRAFIA O ECOGRAFIA, RESONANCIA MAGNETICA
IMAGENOLOGIA CON OTRAS TECNICAS NO RADIOLOGICAS INCLUYE: ESTUDIOS DE
IMAGENOLOGIA RADIOLOGICA
CONSULTA, MEDICIONES ANATOMICAS, FISIOLOGICAS, EXAMENES MANUALES Y ANATOMOPATOLOGICOS
LABORATORIO CLÍNICO
Other
GrupoDesc