UNIVERSIDAD NACIONAL DE MOQUEGUA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

MODELO DE PREDICCION DE CLIENTES QUE DARAN DE BAJA EL SERVICIO HFC BASADO EN REDES NEURONALES EN LA CIUDAD DE MOQUEGUA

TESIS

PRESENTADO POR:

LUIS ENRIQUE HUERTA LAURA

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO EN SISTEMAS E INFORMÁTICA

MOQUEGUA – PERU

2020

UNAM
UNIVERSIDAD NACIONAL DE MOQUEGUA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Folio **Nº 001**

ACTA DE SUSTENTACIÓN

I. Modalidad:

Tesis : [X]
Trabajo de Suficiencia Profesional : [ ]
Otro: ..... : [ ]

II. Documento de aprobación para la sustentación:

Resolución Comisión Organizadora Nº 462-2021-UNAM

Fecha: 20 de mayo 2021

III. El jurado ha dictaminado la (el) tesis, TITULADO:

"Modelo de predicción de clientes que daran de baja el servicio HFC basado en redes neuronales en la ciudad de Moquegua".

Bachiller: Luis Enrique Huerta Laura

Asesor Principal: Dr. Anibal Fernando Flores García

Co-Asesor: Mgr. Juan Carlos Clares Pesca

IV. Se dictamina:

Como: Aprobado por UNANIMIDAD la sustentación de manera:

Presencial [ ] Virtual [X]

Con nota: Doce (12) por lo que; Sí [X] No [ ]

Opta el Título Profesional de INGENIERO DE SISTEMAS E INFORMÁTICA.

Siendo las 19:08 horas, del 21 de mayo del año 2021 se procede con la firma en conformidad de lo dictaminado por el Jurado.

Presidente
MS. Hugo Ester Tito Chura

M.Sc. Alex Zúñiga Incalla
Primer Miembro

Mgc. Victor Yana Mamani
Segundo Miembro

Luis Enrique Huerta Laura
Bachiller

DEDICATORIA

Comenzar con expresar inmensa gratitud a Dios por la vida y la salud, por sus bondades, el conocimiento que me permite obtener, la protección y bendiciones recibidas hasta el día de hoy.

Expresar seguidamente gratitud y estima que siento por mi familia que estuvieron y siguen estando en los buenos y malos momentos, a la persona que admiro que es mi esposa que día a día contribuye con su ayuda idónea e indispensable para tomar decisiones correctas, por su inspiración, felicidad y actos que endulcen nuestro hogar. A mis hijos que con sus sonrisas imparten alegrías, experiencias y motivos para desear ser mejor persona.

A mi madre por su paciencia y dedicación, por su contribución, sus palabras y consejos, queme disciplinan con amor, a mi padre que antes de partir me dio su ejemplo y sus consejos y hoy desde el cielo se siente orgulloso por los logros realizados.

Agradecer a aquellos docentes que, con su forma única de enseñanza y su manera especialde transmitir el conocimiento, pude desarrollarme de manera más profunda de manera profesional, cultural, y moral.

RESUMEN

La investigación surge al observar el comportamiento de los clientes desde el año 2014 al 2020 en promedio el porcentaje de bajas es de 59.66%, dado que los clientes son uno de losprincipales elementos de toda empresa, es más costoso el que ingrese un nuevo cliente, a diferencia del cliente que permanece, el porcentaje que arroja la investigación es alto para latasa de abandono de clientes. La presente investigación tiene como principal objetivo desarrollar un modelo de predicción de clientes que dará de baja el servicio HFC, basado enRedes Neuronales con la finalidad de predecir clientes que en un determinado tiempo desistirán o darán de baja el servicio HFC, para poder plantear estrategias de retención y tener el porcentaje de bajas a su margen. Se tiene investigaciones modelo, por ello se pretende analizar la data y determinar las variables idóneas para modelar un algoritmo de machine learning, en la actualidad existen múltiples metodologías de investigación, siendo una de las más utilizadas por el amplio estudio en inteligencia artificial, emplean la metodología CRISP DM para desarrollar diversos modelos de predicción teniendo un enfoque enlazado de los objetivos de la metodología con los de la investigación, la investigación es descriptiva cuasi-experimental, con una población de 6793 que representa el 100%, esta data se encuentra particionada al 30% para testear el modelo y el 70% para el entrenamiento del modelo, finalmente este modelo permite estructurar y clasificar la data, para posteriormenteser procesada, aplicándose métricas como precisión, recall, f1-score y support, curva ROC con la finalidad de brindar calidad, objetividad y cuantificabilidad, aplicando una técnica de validación cruzada para analizar los resultados y hacer un ajuste en las derivadas, aplicandodos experimentos en la fase de modelamiento, manipulando la cantidad de neuronas en las capas ocultas, aplicando tres algoritmos de optimización Adam, Sgd y Rmsprop, donde se obtuvo una precisión por Adam del 98%.

ABSTRACT

The research arises from observing the behavior of customers that from 2014 to 2020 on average the percentage of cancellations is 59.66%, given that customers are one of the main elements of any company, the percentage that the research shows is high for the customer churn rate. The main objective of this present is to carry out a model for predicting the loss ofresearch clients, based on neural networks in order to predict clients that in a certain time willgive up or cancel the HFC service, to be able to propose retention strategies and have the percentage of casualties within its margin, for this reason it is intended to analyze the data and determine the ideal variables for modes of a machine learning algorithm, currently there are multiple research methodologies, one of the most used by the extensive study in artificialintelligence , use the CRISP DM methodology to develop various prediction models having an approach linked to the objectives of the methodology with those of the research, finally this model allows to structure and classify the data, to later be processed, application and metric quantifiability. The partition of the data gets 30% was performed to test the model and 70% for the model training, applying a cross-validation technique to analyze the results and make an adjustment in the derivatives, applying two experiments in the modeling phase, manipulating the number of neurons in the hidden layers, applying three optimization algorithms Adam, Sgd and Rmsprop, where a precision by Adam of 98% was obtained.

CONTENIDO

RESUMEN 3
CONTENIDO 5
CAPITULO I: INTRODUCCIÓN 10
1.1. DESCRIPCIÓN Y FORMULACION DEL PROBLEMA 11
1.2. ANTECEDENTES 12
2.1. ESTADO DEL ARTE 18
1.3. FORMULACIÓN DE OBJETIVOS 19
1.3.1. Objetivo General 19
1.3.2. Objetivos Específicos 19
1.4. JUSTIFICACIÓN 19
1.5. HIPÓTESIS 22
1.5.1. Hipótesis General 22
1.5.2. Hipótesis Específicas 22
1.6. FORMULACIÓN DEL PROBLEMA 22
1.6.1. Interrogante General 22
1.6.2. Interrogantes Secundarias 22
CAPITULO II: MARCO TEÓRICO 23
2.2. BASES TEÓRICAS SOBRE EL TEMA DE INVESTIGACIÓN 23
2.2.1. Inteligencia Artificial 23
2.2.2. CRISP DM 23
2.2.2.1. Comprensión del Negocio 24
2.2.2.2. Comprensión de los Datos 24
2.2.2.3. Preparación de los Datos 25
2.2.2.4. Modelado 25
2.2.2.5. Evaluación 26
2.2.2.6. Despliegue 26
2.2.3. Tipos de Churn 26
2.2.3.1. Pasivo 27
2.2.3.2. Activo 27
2.2.4. Clasificación 27
2.2.5. Enfoque Estadístico 28
2.2.6. Tipo de Aprendizaje 28
2.2.6.1. Aprendizaje supervisado 28
2.2.6.2. Aprendizaje no supervisado 29

2.2.6.3. Aprendizaje por refuerzo 30
2.2.7. Imputación de Datos. 31
2.2.8. Redes Neuronales 31
2.2.8.1. Tipos de Redes Neuronales 31
2.2.8.1.2. Perceptrón Multicapa 32
2.2.8.1.3. BackPropagation 34
2.2.8.2. Convergencia 34
2.2.8.3. Función de Activación 35
2.2.8.4. Función Lineal 36
2.2.8.5. Función Tangente Hiperbólica 37
2.2.8.6. Función Escalón 37
2.2.8.7. Función Sigmoidal 38
2.2.9. Algoritmos de Optimización 39
2.2.9.2. Root Mean Square Propagation (RMSprop) 39
2.2.9.3. Estimación de Momento Adaptativo (ADAM) 40
2.2.10. Peso de la red Neuronal 40
2.2.11. Valores de entrada 40
2.2.12. Valores de Salida 41
2.2.13. Bias o Sesgo 41
2.2.14. Regla de Aprendizaje 41
2.4 DEFINICIÓN DE TÉRMINOS 42
CAPITULO III: METODO 44
3.1.1. Tipo 44
3.1.2. Diseño 44
3.2. NIVEL DE INVESTIGACIÓN 45
3.3. POBLACION Y MUESTRA 46
3.3.2. Muestra 47
3.4. Tecnicas e instrumentos para la recolecion de datos 47
3.4.1. Técnicas e Instrumentos 47
3.5. MATRIZ DE CONSISTENCIA 52
3.6. TABLA DE DATOS 54
3.7. OPERACIONALIZACIÓN DE VARIABLES 56
3.7.2. Clasificación de las variables 56
3.7.2.1. Definición de variable dependiente 57
3.7.2.2. Definición de variable independiente 57
3.7.3. Variable Independiente 58
3.7.4. Variable Dependiente 59

3.8. MÉTODOS Y TÉCNICAS PARA LA PRESENTACIÓN Y ANÁLISIS DE DATOS 60
3.8.2. Preparación de Datos 61
3.8.3. Construcción del Modelo 65
3.8.4. Evaluación del Modelo 68
3.8.5. Explotación 70
3.9. VALIDACIÓN Y CONFIABILIDAD DE LOS INSTRUMENTOS 71
3.9.2. Matriz de Clasificación 71
3.9.3. Curva ROC 73
CAPITULO IV: RESULTADOS 74
CAPITULO V: DISCUSION DE RESULTADOS 75
5.1. COMPROBACION DE HIPOTESIS 75
5.2. TRABAJOS FUTUROS 76
CAPITULO IV: CONCLUSIONES 77
CAPITULO VII: RECOMENDACIONES 78
CAPITULO VIII: REFERENCIAS 79
CAPITULO IX: ANEXOS 83
Anexo 01: Flujo de Aprobación para anonimizar datos de Clientes 83
Anexo 02: Código para el análisis de Data 84

INDICE DE ILUSTRACIONES

Ilustración 1: Metodología CRISP DM 24
Ilustración 2: Aprendizaje Supervisado 29
Ilustración 3: Aprendizaje No Supervisado 30
Ilustración 4: Aprendizaje por Refuerzo 30
Ilustración 5: Perceptron Simple 32
Ilustración 6: Perceptron Multicapa 33
Ilustración 7: Arquitectura Feed Forward 34
Ilustración 8: Convergencia 35
Ilustración 9: Funciones de Activación 36
Ilustración 10: Función Lineal 36
Ilustración 11: Función Tangente 37
Ilustración 12: Función Escalón 38
Ilustración 13: Función Sigmoidal 39
Ilustración 14: Diseño Cuasi Experimental 44
Ilustración 15: Validación Cruzada 45
Ilustración 16: Comparación del Conjunto de Datos 63
Ilustración 17: Correlación de las Variables 64
Ilustración 18: Feature Importance 65
Ilustración 19: Arquitectura de la red Neuronal 67
Ilustración 20: Reporte de Métricas 68
Ilustración 21: Modelo de Exactitud 69
Ilustración 22: Modelo de Perdida 70
Ilustración 23: Resultados Predictivos 71
Ilustración 24: Reporte de Métricas de Conjunto de Validación 72
Ilustración 25: Métrica ROC 73
Ilustración 26: Flujo de Aprobación para anonimizar datos de Clientes 83

INDICE DE TABLAS

Ilustración 1: Metodología CRISP DM 24
Ilustración 2: Aprendizaje Supervisado 29
Ilustración 3: Aprendizaje No Supervisado 30
Ilustración 4: Aprendizaje por Refuerzo 30
Ilustración 5: Perceptron Simple 32
Ilustración 6: Perceptron Multicapa 33
Ilustración 7: Arquitectura Feed Forward 34
Ilustración 8: Convergencia 35
Ilustración 9: Funciones de Activación 36
Ilustración 10: Función Lineal 36
Ilustración 11: Función Tangente 37
Ilustración 12: Función Escalón 38
Ilustración 13: Función Sigmoidal 39
Ilustración 14: Diseño Cuasi Experimental 44
Ilustración 15: Validación Cruzada 45
Ilustración 16: Comparación del Conjunto de Datos 63
Ilustración 17: Correlación de las Variables 64
Ilustración 18: Feature Importance 65
Ilustración 19: Arquitectura de la red Neuronal 67
Ilustración 20: Reporte de Métricas 68
Ilustración 21: Modelo de Exactitud 69
Ilustración 22: Modelo de Perdida 70
Ilustración 23: Resultados Predictivos 71
Ilustración 24: Reporte de Métricas de Conjunto de Validación 72
Ilustración 25: Métrica ROC 73
Ilustración 26: Flujo de Aprobación para anonimizar datos de Clientes 83

CAPITULO I: INTRODUCCIÓN

El procesamiento de información en los últimos años ha tenido grandes avances en la investigación, el sueño que las maquinas puedan aprender por sí mismo ha sido un reto durante mucho tiempo, sin embargo, esto cada vez se hace posible debido al avance científico y tecnológico, surgen nuevas técnicas de modelamiento y procesamiento deinformación que nos dan el alcance se hacer dicho sueño (Fielding, 1999).

Todos los algoritmos de aprendizaje deben tener un correcto procesamiento de data para poder analizarlos y almacenarlos, sin embargo, el procesamiento puede ser lento debido a dicho almacenamiento entonces, una opción a este problema es la programación en Python,que tiene soporte GPU para el procesamiento y múltiples librerías como NumPy y SciPy, proporcionan soportes matriciales multidimensionales, los cuales los modelos complejos tienen como objetivo el procesamiento de grandes cantidades de información, la capacidad de reconocer y procesarla ha sido la evolución de la Inteligencia Artificial que como inspiraciónes lograr una inteligencia similar al cerebro humano.(Bengio et al., 2015).

Dicha información comprende de los años 2014 al 2020 del servicio HFC, en sus inicios las bajas representaban un 9% de los clientes que migraban, sin embargo, durante los años aumento significativamente del 39% al 80% de bajas, esta estadística muestra que tal servicio por motivos de oferta y demanda, genera un perdida financiera y estratégica debido que el área de inteligencia comercial no se adecua a la necesidad tecnológica financiera que sigue en constante desarrollo. Durante el año 2018 se extiende el uso de fibra óptica para las personas naturales, este servicio estaba incluido sólo para empresas mineras, sin embargo,generó grandes altas del servicio por ende es vital la clasificación de datos para analizar las características relevantes y poder aplicar un modelo de IA, además, se plantea implementarun modelo de análisis de datos y predicción aplicando la metodología CRISP DM que muestradiversas tareas para desarrollarlo.

Esta investigación comprende un ciclo de vida del modelo-cliente para la evaluación del experimento se validará por CROSS VALIDATION para la partición de la data, de tal manera el resultado será representado porcentualmente frente a los próximos clientes que den baja dicho servicio.

1.1. DESCRIPCIÓN Y FORMULACION DEL PROBLEMA

Empezando el segundo trimestre del año 2005, América Móvil Perú SAC empieza a proporcionar servicios de comunicaciones personales, posteriormente en el mes de octubre del año 2010, América Móvil Perú SAC absorbe a Telmex, y empieza a ofrecerlos servicios de telefonía, internet y televisión digital bajo la marca de Claro ofreciendo también su servicio de GSM, GPRS y EDGE, durante el 2013, AMP decide ingresar al mercado de Moquegua, inicialmente con 14 zonas de cobertura donde se estimó elservicio para un total de 7000 clientes, desde sus inicios la empresa gozo derentabilidad sostenida, dado que al cierre del año 2014 se tenía una taza de clientes que abandonan el servicio de un 9.09%, posteriormente en el 2015 la tasa aumento al 39.67% y durante los 2016 – 2017 la tasa de abandono de clientes en relación al servicio HFC se elevó al 79.08%, en el transcurso del 2018 se construyen nuevas zonas de cobertura, se liberan un total de 11 planos para 6000 clientes aproximadamente, durante ese mismo año la tasa de abandono de clientes fue de 32.32%, sin embargo, para el 2019 y el 2020 la tasa se elevó exponencialmente al 89.19%, conforme evoluciona el producto en el tiempo se tiene un margen de pérdida de clientes que están en paralelo a la suba de clientes.

La relevancia de dicha pérdida durante el transcurso de los años, mantiene que las empresas en sus inicios se preocupan más por el almacenamiento de los datos para analizarlos y saber el margen de ganancia y pérdida, sin embargo, la generación de los reportes no respondían la pregunta de los motivos de perdida y el tiempo del clientecon el servicio, este tipo de negocio que se maneja entre la

oferta y demanda con el tiempo debe mejorar, la tecnología que se aplica en varias empresas encuentra una sostenibilidad económica en la inversión de aplicaciones de modelos predictivos, dadoque es más rentable para la empresa el que el cliente que ingreso permanezca, en referencia al cliente que recién está ingresando, dado que ahí factores que deben considerarse en la pre venta, como gastos de publicidad, gastos de operaciones queincluye pago de locales, pago al personal de ventas, gastos que permiten el funcionamiento y mantenimiento a los sistemas de gestión de datos, posteriormente considerar los gastos luego de realizada la venta como, pago al personal que realiza la atención, en las fases de antes, durante y después, material, herramientas y equipos usados, para ello se debe aplicar una metodología idónea para predecir los clientes latentes a desistir del servicio.

1.2. ANTECEDENTES

Como antecedentes similares a la investigación se tienen los siguientes resúmenes.

(Hu et al., 2020) Aplica árboles de decisión y redes neuronales como objetivo de predecir los clientes como abandono según el tiempo de compra del producto, la clasificación es interpretada en base al tiempo de compra, este intervalo es el resultado de abandono, los cuales según las variables de la siguiente tabla refleja el aumento o disminución del número de compras.

Tabla 1: Características del Cliente

Miembro Tarjeta	Afiliación	Años	Numero de Compras (puntos)	Promedio Compras	Unidad Compras	Valor de Tendencia
1800183	Ordinario	34	0	41.6	0.604	0.112

Fuente: Investigación sobre un modelo de predicción de combinación de abandono declientes basado en decisiones

Como resultado del algoritmo C5.0 y la red neuronal se obtiene el siguiente gráfico, laconfianza de ambos modelos por medio de la ponderación y puntuación ponderada tiene como resultado la probabilidad del abandono del cliente.

Tabla 2: Resultados de Modelos de Predicción

Modelo	Clientes	Abandono Previsto	Unchurn Previsto de Clientes
Árbol de Decisión	Abandono Real	562	107
	No Abandono	68	1994
Red Neuronal	Abandono Real	618	51
	No Abandono	45	1967

Fuente: Investigación sobre un modelo de predicción de combinación de abandono de clientes basado en decisiones

La red neuronal y el árbol de decisión fue aplicada a una muestra de 2681 clientes que comprenden los años 2018 – 2019, la investigación arroja como resultado empírico una probabilidad de 93.47% y 96.42% entre el árbol de decisión y red neuronal respectivamente, esto significa que, la precisión de la red neuronal es alta debido a las múltiples derivadas que cuenta un algoritmo de aprendizaje profundo.

Según la investigación (Group et al., 2013), tiene como objetivo implementar una metodología para predicción de clientes en una compañía de telecomunicaciones, aplicando la metodología KDD, hace referencia al servicio NGN (Next Generation Networks) para la predicción óptima de churn de los clientes, entre los cuales su análisis se basa en 4 factores (Humano, Técnico, Económico, Administrativo). Como experimento cuenta 7 fases que comprenden del 2010 al 2011 en intervalos bimestrales.

Aplica SVM, KNN, Árboles de Decisión, Regresión Logística, utiliza el software WEKAy como medida de evaluación de los modelos, los cuales mediante una validación cruzada obtienen los siguientes resultados con una muestra de 2536 registros.

Tabla 3: Evaluación de Modelos Predictivos

CRITERIOS	J4.8	LADTree	Random Tree	Random forest	Red Neuronal
Accuracy	89.54%	89.76%	68.29%	88.76%	90.03%
Medida F	88.94%	89.31%	62.64%	87.70%	89.23%
AUC	0.928%	0.948	0.672	0.943	0.946
Lift	232.41	231.19	164.47	238.07	238.9
TN	1545	1536	1489	1581	1577
TP	852	867	339	795	933
FN	88	73	601	145	107
FP	192	201	248	156	160

Fuente: Diseño e Implementación de una metodología de predicción de fugade clientes en una compañía de Telecomunicaciones

Como resultado establece un data warehouse que añade la data inicial y la predictiva,posteriormente dicha predicción se transforma en la data inicial y se convierte en unametodología incremental, el resultado de la metodología KDD es investigar la integración del personal con el producto, no obstante está el riesgo del abandono voluntario del cliente, existen variables que representan el riesgo de fuga en la data por la poca información en dicha variable. Como propuesta añade un modelo de segmentación de clientes, tipo de

cliente con relación a fuga o no.

La investigación (Meza, 2018), como necesidad en el sector de telecomunicaciones tiene como objetivo identificar las variables relevantes para predecir la fuga de clientes,compara los modelos Adaboost desbalanceado y Regresión Logística Asimétrica aplicando la metodología SEMMA y CRISP DM, una data desbalanceada es un problema en el modelamiento del algoritmo y los resultados no llegan hacer lo esperado, sin embargo, aplica el algoritmo de sobre muestreo (Oversampling) para generar data desde la distancia de cada registro mostrado en el hiperplano, aplicando un algoritmo de k-vecinos cercanos, como resultado obtiene una muestra de 80300 registros, los cuales aplicando el algoritmo de validación cruzada aplica un 70% de ladata para el entrenamiento y 30% para la validación.

El siguiente recuadro refleja los resultados de las métricas de la data desbalanceada y balanceada.

Tabla 4: Comparación de las medidas de desempeño con los métodos de muestreo

Medidas	**Sin Ajuste**	**Sub Muestreo**	**Sobre Muestreo**	**SMOTE**
Accuracy	0.90785	0.77389	0.77314	0.81341
Tasa de Error	0.09215	0.22611	0.22686	0.18659
Precisión	0.60684	0.28150	0.28096	0.31423
Recall	0.20752	0.82088	0.82213	0.74154
F Measure	0.30927	0.41923	0.41880	0.44141

Fuente: Predicción de fuga de clientes en una empresa de telefonía utilizando el algoritmoadaboost desbalanceado y la regresión logística asimétrica

Estos resultados en cuanto a precisión y mejor desempeño se obtienen el sub muestreo y SMOTE obtienen una precisión moderada. Sin embargo, SMOTE tiene unligero aumento en la clase predictiva, el AUC arroja un 93% de confianza, el siguiente experimento que realiza es aplicando los modelos de predicción con una data balanceada y como resultado se obtiene el siguiente recuadro.

Tabla 5: Medidas comparativas de desempeño de los 3 modelos ajustados

Medidas	Logit Asym	Power Logit	Adaboost Asym
Exactitud	0.91005	0.90643	0.92785
Tasa de Error	0.08995	0.09357	0.07215
Precisión	0.63380	0.59363	0.75406
Recall	0.22547	0.18664	0.40710
F measure	0.33261	0.28399	0.52874

Fuente: Predicción de fuga de clientes en una empresa de telefonía utilizando el algoritmo adaboost desbalanceado y la regresión logística asimétrica

En conclusión, se obtiene como resultado a nivel de desempeño, el algoritmo Adaboost es superior, obtiene una medida AUC del 0.93, la probabilidad de fuga del cliente es mínima.

(Barrueta Meza & Castillo Villarreal, 2018) Aplica algoritmos de regresión logística y árboles de decisión, teniendo en cuenta que durante el 2004 al 2006 tuvo un crecimiento del plan Post Pago del 35%, sin embargo, durante el 2010 y 2012 bajo la tasa de crecimiento teniendo una pérdida del 20% de fuga de clientes, para ese año el 17% de clientes con un plan post pago era de 1.164.074,

ese mismo año considerauna fuga del 5.1%, que fue superior al año 2008 con 3.8%, como primera instancia aborda el tipo de clasificación churn del cliente, como data histórica abarca desde Octubre del 2012 a Marzo del 2013, los cuales el filtro determinó un total de 36.891 fugados y 13.208 no fugados, esto representa el 20% de clientes fugados. Las variables en la investigación se enfocan más en el uso del equipo como el número dellamadas, tiempo, tipo de plan, días de retraso de pago, morosidad, kilobytes usados,cantidad de reclamos y como espacio demográfico, ciudad entre otros. Tiene una estructuración de variables por tiempo, cada variable tiene una representación mensual distinta, hace una categorización de ARPU T, donde T es el mes, T++ representa cada mes del cliente durante el semestre anual.

El primer experimento semestral, se representa económicamente la perdida por la fugade clientes que comprenden el primer semestre.

Tabla 6: Evolución ARPU en 6 meses para Fugados y No Fugados

Tipo Cliente	ARPU T	ARPU T-1	ARPU T-2	ARPU T-3	ARPU T-4	ARPU T-5
No Fugado	17.711,3	18.216,5	18.743,5	18.289,9	27.773,3	32.926,6
Fugado	23.268,9	23.773,4	23.830,3	23.256,4	23.298,0	17.260,9

Fuente: Modelo de predicción de fuga de clientes de telefonía móvil post pago

El segundo experimento es la comparación entre 2 modelos predictivos, árboles de decisión y una regresión logística, el árbol de decisión posee el 18% menos de exactitud que la regresión, los criterios de evaluación son representados en la siguiente tabla.

Tabla 7: *Comparación de Modelos mediante métricas*

Medida	Árbol	Regresión
Lift	1,48	1,21
Accuracy	0,41	0,59
Sensitivy	0,73	0,47
TP	152	98
TN	1545	2360
FP	2417	1602
FN	57	111

Fuente: Modelo de predicción de fuga de clientes de telefonía móvil post pago

La investigación hace una evaluación económica con la siguiente fórmula:

$$Beneficio(B) = N° \; de \; Clientes * ARPU * \frac{1}{\% \; Churn}$$

$$Inversión(Inv) = N° \; de \; Clientes * 1{,}8 * ARPU$$

$$Ingreso \; Neto(IN) = B - Inv$$

Estas fórmulas hacen referencia a la cantidad efectiva que se debe lograr retener paraalcanzar las acciones que la compañía ejecuta.

2.1. ESTADO DEL ARTE

Aplicar un modelo de análisis predictivo (Barrueta Meza & Castillo Villarreal, 2018) en el sector financiero aplicando modelos predictivos, k-means, R-CNR y R-NNet, tiene una precisión del 93.20%, 92.50% y 87.3% respectivamente, la arquitectura de la red neuronal abarca 5 capas ocultas con un límite 100 iteraciones y función de activaciónSoftmax a una muestra de 20,000 registros.

(LEITE, 2018) Desarrolla una comparación de algoritmos aplicados al Churn, como resultado se obtienen datos estadísticos los cuales

recomiendo principalmente los modelos de SVM y RNA, hace una segmentación de clientes para que la empresa pueda generar alertas de los clientes con más alta probabilidad de Churn.

La aplicación de redes neuronales en los diferentes sectores es de gran apoyo en losdiversos problemas, el tipo de complejidad en el análisis de una red neuronal, según (Quiñones Huatangari et al., 2020) aplica una red neuronal para estimar el indica de calidad de agua aplicando el algoritmo de Levenberg-Marquardt con una función de activación no lineal obtuvo un coeficiente de 0.979 en el entrenamiento y 0.940 en la predicción.

1.3. FORMULACIÓN DE OBJETIVOS

1.3.1. Objetivo General

- Desarrollar un modelo de predicción de clientes que dará de baja el servicio HFC.

1.3.2. Objetivos Específicos

- Desarrollar el modelo de red neuronal para el reconocimiento de patrones declientes que dará de baja el servicio HFC.
- Validar el modelo de predicción de clientes, que dará de baja el servicio HFCa través del caso de estudio: América Móvil Perú SAC en la ciudad de Moquegua.

1.4. JUSTIFICACIÓN

En 1950, Alan Turing publica un artículo titulado “Computación e Inteligencia” donde plantea la prueba de Turing, la cual es mostrar el comportamiento de una maquina similar al humano, capaz de tomar decisiones y supere las habilidades. Como modelo inicial entre ML, el 1957 Fran Rosenblatt diseña el modelo de Perceptron que imita elcomportamiento de las neuronas capaz de almacenar y

procesar información, este modelo da paso a la creación del aprendizaje basado en experiencia, posteriormente durante los años 90, ya se empiezan a crear programas capaces de manipular grandes cantidades de datos, IBM presenta nuevas versiones del computador DeePerBlue, como experimento de ganar un juego de ajedrez, el paradigma de este experimento es de aprendizaje profundo.

Grandes empresas que explotan estas técnicas como Google, Amazon, Facebook, emplean como primera opción ML por los amplios concepto y aplicaciones que ha mostrado de manera exponencial en los últimos años.

En cuanto a lo tecnológico existen distintos framework como el de Anaconda, que esuna suite de código abierto, utilizado para diseñar modelos de inteligencia artificial, proporciona un entorno de paquetes e IDE como Jupyter, RStudio entre otros, tiene una amplia variedad de librerías como Pandas, Numpy, Dask, que permite visualizar datos con Matplotlib, este entorno brinda posibilidades de acceder recursos deaprendizaje más avanzado, en la actualidad una API diseñada para resolver problemas de predicción es Keras, que ofrece múltiples variables en su diseño como la hiper parametrización que facilita el encontrar las características mas relevantes del modelo, añadiendo el GPU de TensorFlow como alternativa en la escala de grandes cantidades de datos, esto con la finalidad de desarrollar este tipo de modelos.

(Group et al., 2013) Pretende conocer el churn de año a año, en las empresas Retail siempre existe la fuga de clientes anualmente es el 30%, sin embargo, conocer los motivos del churn para predecirlos y describir el tipo de riesgo potencial que genera ala compañía es el problema general en el mercado. En la industria de telecomunicaciones, las principales razones por el cual existe el churn, es por la disconformidad del cliente a la empresa, por motivos de un mal producto hasta la disposición de su atención,

múltiples factores que no pueden ser analizados por cadacliente y estas empresas donde la fidelización es patrimonio, lo principal es conocer el grado de riesgo de fuga. Conocer el flujo de información para extraer la data es unafase en este proyecto que, como tal la clasificación y desbalanceo de la data son procesos que se deben evaluar y en conclusión, aplican KNN, Naibe Bayes, Arboles de Decisión y Support Vector Machine como experimento de modelos predictivos.

(Galván & Navarro, 2015) Propone como metodología identificar la data, realizar una clasificación de las variables y posteriormente describirlo, en la actualidad los problemas de portabilidad, confidencialidad, imputación de datos tienden hacer el factor de una fuga, según OSIPTEL la portabilidad del cliente debe ser de manera rápida en un periodo de 24 Hrs, esto ocasiona un problema en la predicción de clientes a un menor tiempo, como solución es la creación de un datawarehouse para manteneruna estructuración de la data frente al negocio, adicionando predicciones de portabilidad, esta data debe ser estandarizada para tener información coherente, la propuesta de la investigación es diseñar una arquitectura iniciando por la faseextracción de la data, los componentes que lo involucran como una base de datos deentramiento y evaluación, por último la aplicación de algoritmos de predicción en baseal tiempo aplicando KNN, Naibe Bayres o Redes Neuronales.

Para la presente investigación se conto con los recursos financieros necesarios para que la investigación se lleve a cabo de la forma esperada, así como para que sea exitosa.

Si bien es cierto, la investigación beneficiará en esta oportunidad directamente a la empresa, ya que permitirá identificar de manera oportuna los clientes propensos a darde baja el servicio, también será de beneficio para futuras investigaciones que pueden ser aplicadas al ámbito universitario, escolar, salud, etc.

1.5. HIPÓTESIS

1.5.1. Hipótesis General

Los modelos de redes neuronales afectan positivamente el desarrollo de unmodelo de predicción de clientes.

1.5.2. Hipótesis Específicas

1.5.2.1. Las redes neuronales identifican positivamente el desarrollo de patrones en un modelo de predicción de clientes.

1.5.2.2. La predicción de clientes que dará de baja el servicio HFC se podrá validar através del caso de estudio: América Móvil Perú SAC.

1.6. FORMULACIÓN DEL PROBLEMA

1.6.1. Interrogante General

- ¿Cuál es el impacto en el desarrollo de un modelo de predicción de clientes quedarán de baja el servicio HFC?

1.6.2. Interrogantes Secundarias

- ¿Se podrá identificar al cliente que dará de baja el servicio HFC aplicando redesneuronales?
- ¿Cuál será el patrón de comportamiento del cliente que dará de baja el servicioHFC?

CAPITULO II: MARCO TEÓRICO

2.2. BASES TEÓRICAS SOBRE EL TEMA DE INVESTIGACIÓN

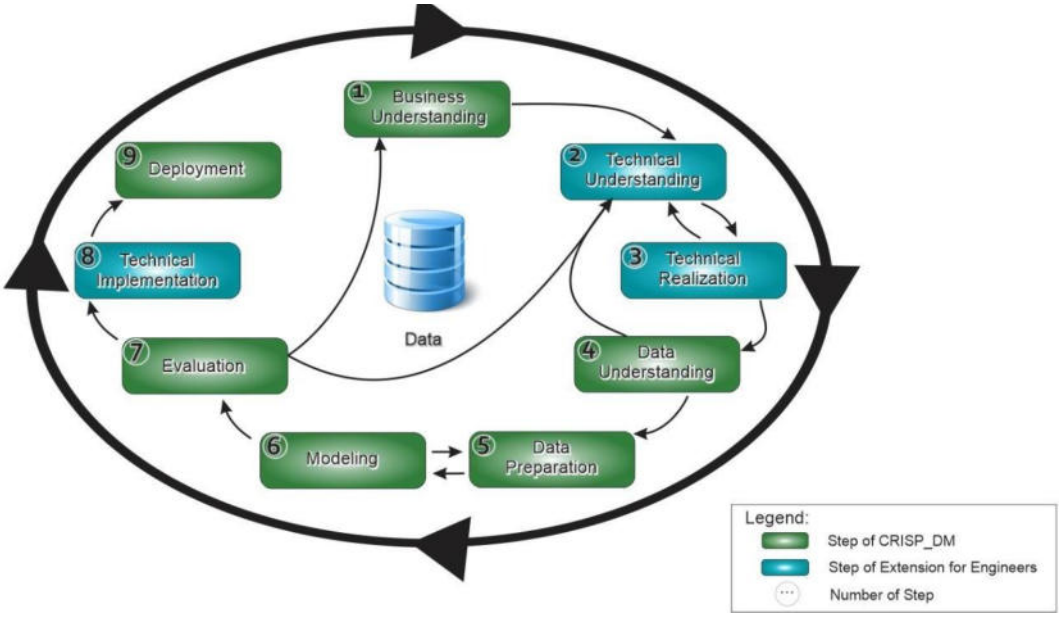
2.2.1. Inteligencia Artificial

Actualmente la Inteligencia Artificial es un área de la ciencia de gran interés por ser un área multidisciplinaria donde se realizan sistemas que tratan de hacer tareas y resolver problemas como lo hace un humano, así mismo se trata de simular de manera artificial las formas del pensamiento y como trabaja el cerebro para tomar decisiones. Aunque en la realidad aún no se ha podido realizar todo lo que las personas sueñan al conocer esta área o al ver lo que se muestra en laciencia ficción es un área que poco a poco va ganando terreno al estar presente en muchas aplicaciones, aparatos, dispositivos que utilizamos de manera cotidiana (Peñafiel & Ávila, 2007).

2.2.2. CRISP DM

Es un modelo de estrategia aplicada a la minería de datos, como metodología engloba fases desglosadas en tareas necesarias para alcanzar el resultado esperado, sin embargo, como proceso es un ciclo de vida del proyecto, la secuencia de estas fases no necesariamente es estrictas, es decir, las fases pueden tener retroalimentación con la fase inicial, esta ventaja permite crear un modelo adaptativo a las necesidades de la empresa, dentro las cuales abarcan las siguientes fases.

***Ilustración 1**: Metodología CRISP DM*



Fuente: Modelo de predicción de fuga de clientes de telefonía móvil post pago

2.2.2.1. Comprensión del Negocio

Esta fase comprende la necesidad empresarial al extraer la información, conocer los objetivos del negocio es un requisito específico para la mineríade datos, esto incluye la selección inicial de herramientas y técnicas que se emplearán para determinar si el proyecto tendrá éxito desde la vista empresarial.

Focaliza los requerimientos y limitaciones del proyecto, también las posibles causas que podrían retrasar la finalización del proyecto y preparación de un plan de contingencia en caso sucedan algún limitante, en resumen, esta fase comprende los beneficios que se obtendrá al finalizar el proyecto en laminería de datos.

2.2.2.2. Comprensión de los Datos

En esta fase tiene como finalidad verificar la data extraída, conocer el tipo de data si es una variable cuantitativa o cualitativa, también evaluar los registros aplicando formulas estadística y conocer si presenta alguna falla de data. La importancia de esta fase en la minería datos permite confirmar si posteriormente se podrá realizar un modelo estadístico, la evaluación delos datos es de vital importancia para identificar problemas en la calidad de datos.

En conclusión, en esta fase se verifica si existe imputación de datos, tipo deerrores en la data determinar las posibles soluciones para la continuidad delproyecto.

2.2.2.3. Preparación de los Datos

Las decisiones se emplean en esta fase, en forma analítica algunas tareascomo verificación de duplicidad de información, data faltante e importanciade cada variable son pasos requeridos para continuar con la fase de modelado, el resultado de esta fase es el informe de limpieza de datos.

Esta fase comprende agrega atributos a las variables más relevantes que se experimentará en la fase de modelado.

2.2.2.4. Modelado

Este proceso es el núcleo de la investigación, acá se aplican técnicas de modelamiento, tipo de partición de la data que comprende el conjunto de entrenamiento, prueba y validación. En forma específica depende del tipo de algoritmo y las variables de esta, también precisa las métricas que evalúan el modelo, estos parámetros verifican el resultado esperado que seconoce en la fase de Comprensión del Negocio.

2.2.2.5. Evaluación

Los resultados obtenidos en la fase anterior se evalúan en esta fase, algunas métricas en la principal, la precisión. Esta fase verifica si la solución del modelo satisface el problema del negocio, también se podría evaluar con otros modelos en tiempo real y saber cuál es más óptimo. El fin de esta fase es implementar el modelo para su mejora.

2.2.2.6. Despliegue

Esta fase comprende la implementación del modelo, generalmente tiene como objetivo profundizar el conocimiento adquirido. La estrategia y continuidad del proyecto son el objetivo en esta fase, así como también identificar mejoras y la aplicación del modelo en diferentes problemas.

2.2.3. Tipos de Churn

En el marco de las telecomunicaciones, el Churn se define como “la acción de cancelar el servicio prestado por la compañía”. En esta cancelación, el cliente decide renunciar a la empresa (voluntaria) o en todo caso la empresa puede expulsarlo por algunas irregularidades o por no cumplir con las obligaciones o pagos acordados (involuntaria). Por lo tanto, se detalla el Churn en función a la decisión del cliente en abandonar la empresa, en el sector de las telecomunicaciones también se define al Churn como aquel término “usado para describir colectivamente el cese de servicios de la suscripción de un cliente, donde el cliente es alguien que se ha unido a la compañía por al menos un periodo de tiempo. Un Churner o fugado es un cliente que ha dejado la compañía”.

2.2.3.1. Pasivo

Es la tasa de Churn de todos aquellos clientes que no se dan de baja de manera directa, a estos se les acaba su periodo de suscripción y no decidenrenovar. Esto ocurre cuando el usuario no está interesado en el producto, olvidándose inclusive de que cuenta con el servicio o producto

2.2.3.2. Activo

Es la tasa de Churn donde los usuarios toman la decisión de terminar suscripción al servicio y buscan los mecanismos formales para darse de baja. (Generalmente suelen llamar por teléfono o lo realizan de forma presencial en un centro de atención al cliente)

Para este estudio, el Churn de clientes son los que voluntariamente decidieron cambiarse de operador y simplemente no renovaron o renunciaron a su servicio, es decir clientes activos que renuncian al servicio.

2.2.4. Clasificación

La clasificación en el marco de las ciencias computacionales se enfrenta a dos tipos de problemas bajo diferentes perspectivas: conocer anticipadamente el número de categorías o grupos en los que se desea hacer una clasificación de datos, de tal forma que hay que definir reglas para clasificar futuras observaciones; o bien establecer determinado número de clases dentro de un grupo de datos obtenidos.

En este caso se utilizó el enfoque de clasificar observaciones futuras en clases o categorías ya conocidas, es decir conociendo las clases como variable principal,clasificar nuevas observaciones a estas categorías.

2.2.5. Enfoque Estadístico

Los métodos estadísticos se caracterizan por estar inmersos en una probabilidadexplícita, la cual proporciona la probabilidad de pertenencia a cada clase en lugarde una clasificación simple. Por otro lado, se asume que las técnicas serán utilizadas por estadísticos lo cual implica cierta intervención humana con la transformación y selección de variables, así como la metodología para la estructuración del problema general.

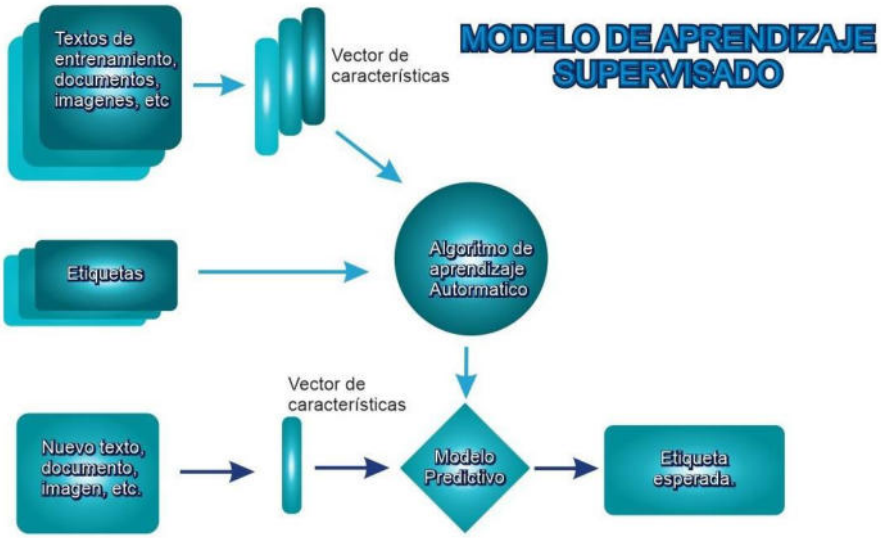
En cuanto al enfoque estadístico en esta investigación se vio reflejada principalmente en la regresión logística asimétrica

2.2.6. Tipo de Aprendizaje

2.2.6.1. Aprendizaje supervisado

Estos tipos de aprendizaje tiene como fase la manipulación de la data, la experimentación del tipo de clasificación que abarca el modelo, la clasificación o categorización de las variables debidamente etiquetadas sonajustes de primera instancia, la exploración de este aprendizaje es conocerel tipo de información que entrará al modelo predictivo e interpretar la salidacon el resultado definido.

Ilustración 2: Aprendizaje Supervisado

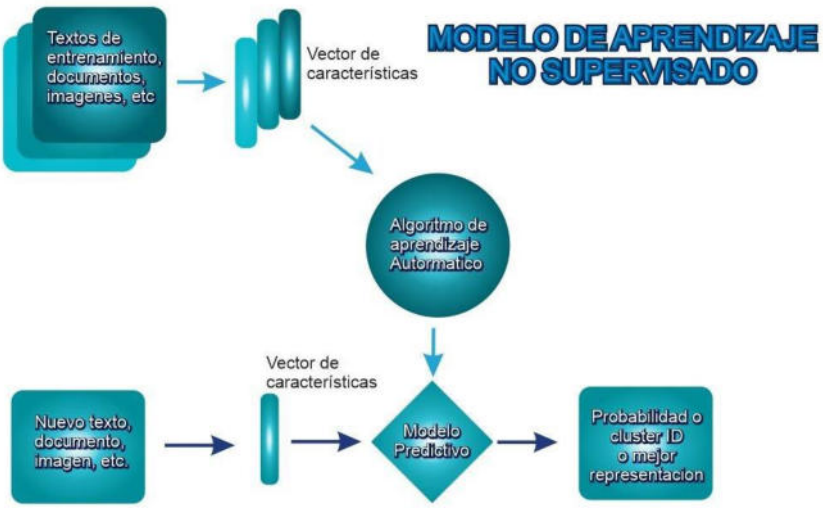


Fuente: Tipos de aprendizaje automático

2.2.6.2. Aprendizaje no supervisado

Estos algoritmos trabajan en forma similar al supervisado con la diferencia que el resultado del modelo no está clasificado, ni definido, esta característica se toma en cuenta para la aplicación de clustering para que la salida se categorice entre las demás salidas similares, aplicando el algoritmo de k-means, este método generalmente es aplicado en geolocalización y procesamiento de imágenes.

Ilustración 3: Aprendizaje No Supervisado

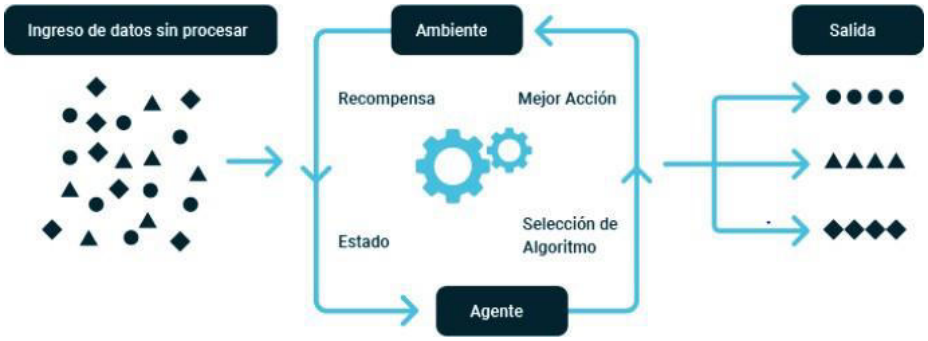


Fuente: Tipos de aprendizaje automático

2.2.6.3. Aprendizaje por refuerzo

Estos modelos de aprendizaje por refuerzo definen funciones enfocadas de maximizar la medida de aprendizaje, el factor clave es el ambiente el cuál un agente se enfrenta para superar el obstáculo, también conocido como "Recompensa – Acción", que generalmente son aplicado a los robots que se enfrentan a múltiples escenarios que deben superar.

Ilustración 4: Aprendizaje por Refuerzo



Fuente: Tipos de aprendizaje automático

2.2.7. Imputación de Datos.

Es la ausencia de información en un espacio muestral, este estudio determina elflujo de información interpretada en los diagramas de caja, también conocido como ruido muestral, actualmente es un problema que se enfrenta al procesar ladata, existen técnicas de imputación de información para llenar este tipo de dato faltante, generalmente suelen eliminar los registros con datos faltantes lo cual generaría perdida de información, por otro lado, existe la ponderación en el hiperplano lo cual hace referencia a la distancia de dicho registro con data faltantea un registro lleno, esta distancia de valores genera un valor aleatorio para compensar la data perdida de información los cuales identifica los valores faltantes como perdida de la muestra, este proceso es denominado imputación dela media o método de sustitución. (Muñoz Rosas & Álvarez Verdejo, 2009)

2.2.8. Redes Neuronales

Tiene como inicio la red de bayer, que se originó en el funcionamiento de las neuronas biológicas, capaz de retener y procesar información donde su estructura consiste en los impulsos nervioso de una neurona a otra, este procesamiento sellevó a una computadora y surge inicial la red de bayes, con el pasar de la tecnología se profundizo dicho modelo, en computación la estimación de neuronas ocultas para el procesamiento es ilimitado, el único obstáculo es el costo computacional, debido a los ajustes en sus derivadas donde el procesamiento enla fase de entrenamiento de la red se evalúan según las métricas que se aplicana todos los algoritmos en machine learning.

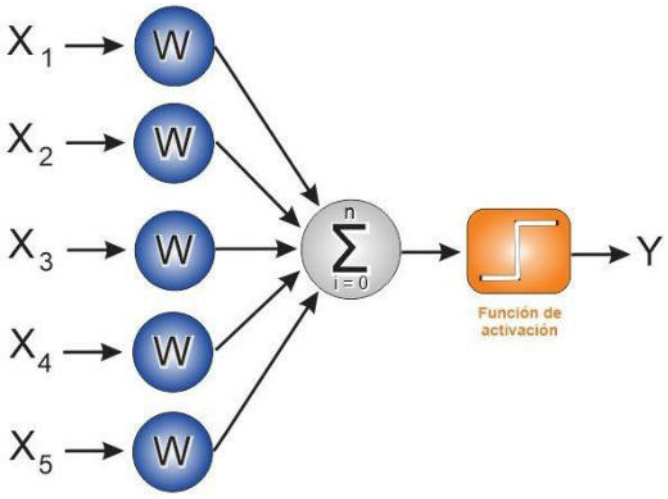
2.2.8.1. Tipos de Redes Neuronales

2.2.8.1.1. Perceptrón

La red perceptrón son redes con una sola capa, es uno de losprimeros modelos que apareció en 1943 y se utilizan generalmente en

tareas relacionadas con lo que se conoce como auto asociación, consta de 2 valores de salida de forma binaria por 1 o 0. Las neuronasde entradas pasan por la función de activación escalón.

Ilustración 5: Perceptron Simple



Fuente: Elaboración Propia

La salida se representa:

$$f(x) = Net_y$$

Cálculo en Y:

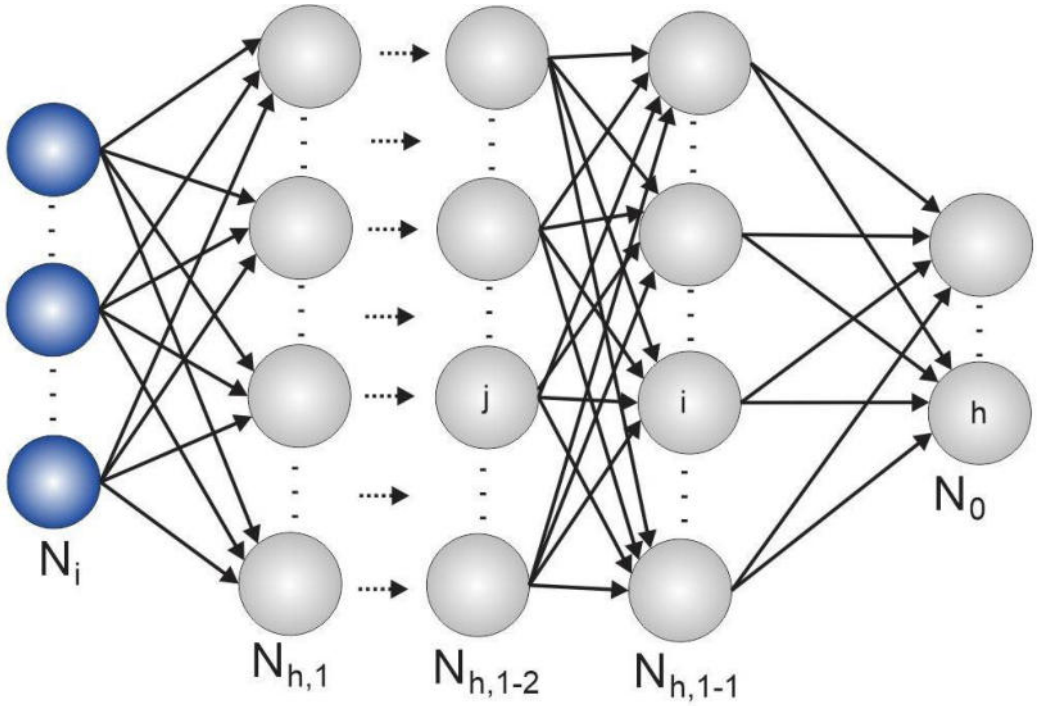
$$Net_y = f\left(\sum_{i=1}^{N} W_i X_i\right)$$

2.2.8.1.2. Perceptrón Multicapa

Este modelo determina el número de capas ocultas, generalmente una red perceptrón, tiene capa de entrada y capa de salida, sin embargo, el perceptrón

multicapa tiene a tener N capas ocultas que están entre la entrada y salida esperada, estas capas determinan la complejidad del modelo, demanda un algo costo computacional según la data, este tipo de red neuronal se interconecta con cada neurona de entrada.

Ilustración 6: Perceptron Multicapa



Fuente: Elaboración Propia

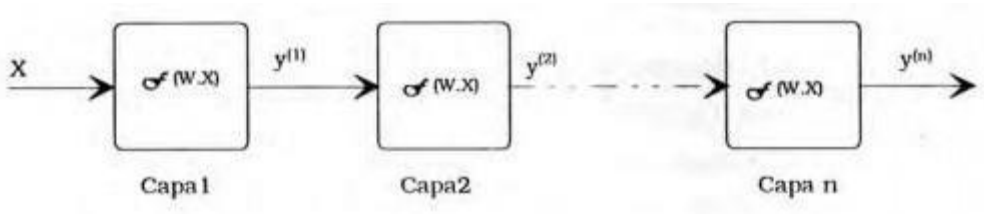
Para el cálculo de la salida empleando la función de activación, el resultado esperado con el resultado de la red tiene un margen de error, la siguiente fórmula tiene como objetivo minimizar dicho error, conocido como el error cuadrático medio.t w

$$E_{SSE} = \frac{1}{2}\sum_{i}\sum_{j}(E_{ij} - S_{ij})^2$$

2.2.8.1.3. BackPropagation

Una red neuronal enfrenta múltiples errores en la configuración de cada variable interna de la red, el más frecuente es el margen de error, una red backpropagation minimiza dicho error aplicando el método del cálculo de la gradiente, el error se propaga hacia atrás, aplicando el método de la cadena se aplica las derivadas desde el cálculo de erroren función de la activación y este en función de la sumatoria de los pesos, este proceso hace una actualización general en toda la red, desde la capa de entrada a la salida y posteriormente de la salida hacia llegar a la entrada, estos procesos tienen N iteraciones que porcada calculo minimiza el error y en un punto X, llega a la convergenciadonde el modelo llega hacer óptimo.

Ilustración 7: Arquitectura Feed Forward



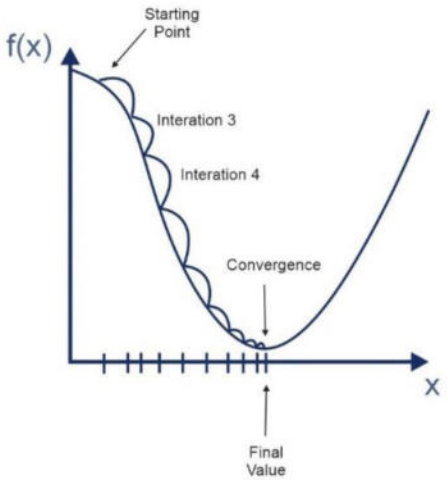
Fuente: Elaboración Propia

2.2.8.2. Convergencia

La selección de las variables de entrada depende, en gran medida, del tipo de procesamiento que se ejecuta, la necesidad de cada variable dentro del modelo y la importancia, dificulta la convergencia en el aprendizaje por variables irrelevantes y el pobre desempeño del

modelo fase ejecución, estos algoritmos de aprendizaje deben alcanzar alto nivel para garantizar ladiminución de error, Si se garantiza la convergencia del error es posible encontrar redes neuronales con un mejor de ajuste y una mejor generalización, mientras que, usando los métodos tradicionales, se puede producir un estancamiento en cierto valor del error de tal forma que aunque se aumente la complejidad del modelo no es posible llegar a errores más bajos.

Ilustración 8: Convergencia

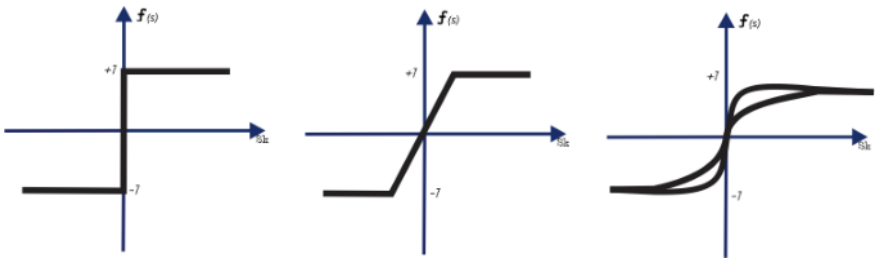


Fuente: Elaboración Propia

2.2.8.3. Función de Activación

Es una función primordial en la elección en la arquitectura de red debido a que tiene influencia en el tipo de problema que puede resolverse con dicha red y también en la convergencia de los algoritmos de aprendizaje, esta función puede ser lineal o no lineal, ambas funciones permiten aprender a clasificar patrones los cuales pueden representarse en el hiperplano.

Ilustración 9: Funciones de Activación



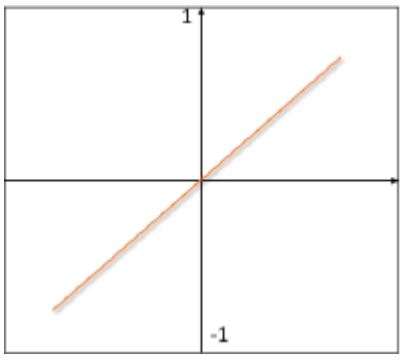
Fuente: Elaboración Propia

2.2.8.4. Función Lineal

Se emplea para estimaciones lineales, donde se definen como 0 o 1, la activación se define como una función lineal de suma de las señales de entrada. La función lineal o mixta corresponde a la función:

$$f(x) = \begin{cases} -1, & x \leq -\frac{1}{a} \\ a * x, & -\frac{1}{a} < x < \frac{1}{a} \\ 1, & x \geq \frac{1}{a} \end{cases}$$

Ilustración 10: Función Lineal



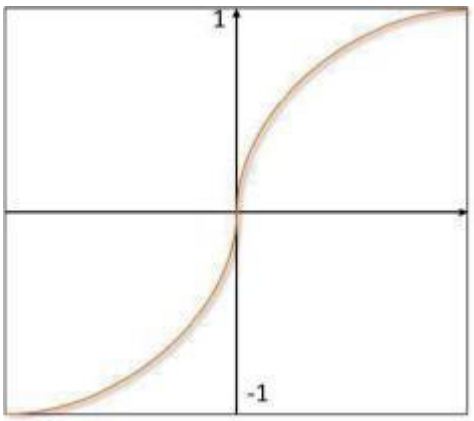
Fuente: Elaboración Propia

2.2.8.5. Función Tangente Hiperbólica

Esta función tiene valores que se aproximan entre un rango de (-1, 1), tieneuna función exponencial, significa que tiende a crecer y se emplea en problemas de estimación,

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

Ilustración 11: Función Tangente



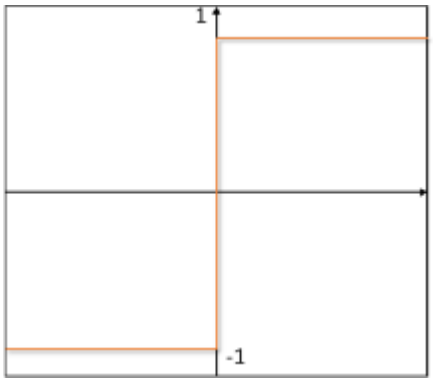
Fuente: Elaboración Propia

2.2.8.6. Función Escalón

Esta función únicamente se utiliza cuando las salidas de la red sonbinarias, cuando la capa de salida es dependiente de la suma de los valores de la capa de entrada. Esta función se usa en problemas de clasificación, la activación es 1 si es menor, y la activación es 0 si es mayor.

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Ilustración 12: Función Escalón



Fuente: Elaboración Propia

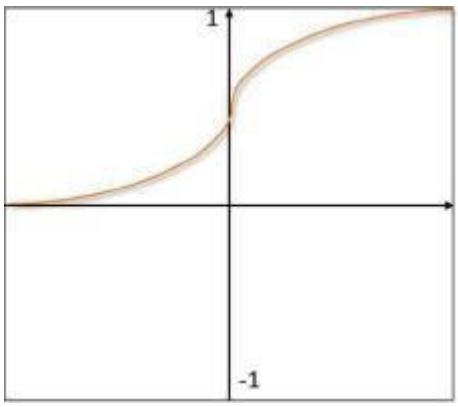
2.2.8.7. Función Sigmoidal

Esta función es utilizada como función de activación de las neuronas que conforman una red neuronal artificial, está definida por un intervalo de posibles valores de entrada, la importancia de lafunción sigmoidal es que su derivada es siempre positiva y cercana

$$f(x) = \frac{1}{1 + e^{-x}}$$

a cero para los valores grandes positivos o negativos, donde laactivación hace un incremento entre 0 a 1.

Ilustración 13: Función Sigmoidal



Fuente: Elaboración Propia

2.2.9. Algoritmos de Optimización

2.2.9.1. Gradiente descendiente

Este algoritmo introduce una variación muy interesante en el concepto de factor de entrenamiento: en vez de considerar un valor uniforme para todos los pesos, se mantiene un factor de entrenamiento específico para cada unode ellos. Sería inviable calcular este valor de forma específica así que, partiendo del factor de entrenamiento inicial, AdaGrad lo escala y adapta para cada dimensión con respecto al gradiente acumulado en cada iteración.

2.2.9.2. Root Mean Square Propagation (RMSprop)

RMSProp es un algoritmo similar. También mantiene un factor de entrenamiento diferente para cada dimensión, pero en este caso el escaladodel factor de entrenamiento se realiza dividiéndolo por la media del declive exponencial del cuadrado de los gradientes

2.2.9.3. Estimación de Momento Adaptativo (ADAM)

El algoritmo Adam combina las bondades de Gradiente descendiente y RMSProp. Se mantiene un factor de entrenamiento por parámetro y además de calcular RMSProp, cada factor de entrenamiento también se ve afectadopor la media del momentum del gradiente. Este algoritmo fue diseñado para acelerar el proceso de optimización.

2.2.10. Peso de la red Neuronal

Normalmente una neurona recibe muchas y múltiples entradas simultáneas. Cadaentrada tiene su propio peso relativo el cual proporciona la importancia de la entrada dentro de la función de agregación de la neurona. Estos pesos realizan la misma función que realizan las fuerzas sinápticas de las neuronas biológicas. En ambos casos, algunas entradas son más importantes que otras de manera que tienen mayor efecto sobre el procesamiento de la neurona al combinarse paraproducir la respuesta neuronal.

Los pesos son coeficientes que pueden adaptarse dentro de la red que determinan la intensidad de la señal de entrada registrada por la neurona artificial. Ellos son la medida de la fuerza de una conexión de entrada. Estas fuerzas pueden ser modificadas en respuesta de los ejemplos de entrenamiento de acuerdo a la topología específica o debido a las reglas de entrenamiento.

2.2.11. Valores de entrada

Cada elemento que contiene datos únicos, y alimentara a la red neuronal, es poreso que la primera capa se le conoce como la capa de entrada.

2.2.12. Valores de Salida

Cada elemento de procesamiento tiene permitido una única salida yi(t) que puedeestar asociada con un número elevado de otras neuronas. Normalmente, la salida es directamente equivalente al valor resultante de la función de activación.

2.2.13. Bias o Sesgo

La aplicación de Bias o Sesgo en las redes neuronales da idea del error en el juicio sobre algo, pero precedida del termino prejuicio, donde indica la posibilidadde existencia que los datos acumulados para la alimentación de algoritmos pueden estar afectos a toda clase de presunciones, prevenciones o preconcepciones, sean estas conscientes o inconscientes, donde resulta en casos de discriminación arbitraria realizada por el algoritmo.

Los sesgos que un modelo puede adquirir en relación con los datos con que se es entrenado suelen clasificarse en, al menos, tres tipos: sesgo de interacción, sesgo latente, sesgo de preselección. El primero consiste en que el propio usuario o programador inadvertidamente introduce un sesgo en el modelo por la maneraen que interactúa con él, el sesgo latente, en cambio, tiene lugar cuando el modelo realiza correlaciones inapropiadas, generalmente al establecer falsos nexos entre puntos de datos, el sesgo de selección, tiene lugar cuando la base de datos no es suficientemente representativa de la diversidad existente en el medio social.

2.2.14. Regla de Aprendizaje

Para las redes neuronales el aprendizaje se define como la modificación del comportamiento de la información, la primera modificación es la información volátil, que se refiere a los datos que se están usando y varían con la dinámica dela computación

de la red, se encuentra almacenada en el estado dinámico de las neuronas, la segunda información es la no volátil que se mantiene para recordar los patrones aprendidos y se encuentra almacenada en los pesos sinápticos.

Biológicamente, se suele aceptar que la información memorizada en el cerebro esta mas relacionada con los valores sinápticos de las conexiones entre las neuronas, que, con ellas mismas, es decir, el conocimiento se encuentra en la sinapsis.

En el caso de las redes neuronales artificiales, el conocimiento se encuentra representado en los pesos de las conexiones entre neuronas. Todo proceso de aprendizaje implica cambios en estas conexiones, es decir, se aprende modificando los valores de los pesos de la red.

Durante el proceso de aprendizaje, los pesos de las conexiones de la red sufren modificaciones por lo tanto se puede afirmar que la red ha aprendido cuando los valores de los pesos permanecen estables ($dw_{ij}/dt = 0$).

2.4 DEFINICIÓN DE TÉRMINOS

Tamayo y Tamayo (2012) sostiene que “las personas que lleguen a una investigación determinada conozcan perfectamente el significado con el cual se va a utilizar el termino o concepto a través de toda la investigación”.

SGA = Sistema de Gestión de Administración, es el nombre del sistema que permite la administración de datos de la empresa.

CV = Cross Validation, Payam Refaeilzadeh, Lei Tang, Huan Lui . (2008). La validación cruzada es un método estadístico de evaluar y comparar algoritmos de aprendizaje dividiendo datos en dos segmentos: uno utilizado para aprender o entrenar un modelo y el otro utilizado para validar el modelo.

ANN = Red Neuronal, Damian Jorge Matich, (2001). Una nueva forma de computación, inspirada en modelos biológicos, así como también un modelo matemático compuesto por un gran número de elementos procesales organizados enniveles.

ML = Machine Learning, Samuel A. (1959). Campo de estudio que dota a los ordenadores de la capacidad de aprender sin ser explícitamente programados

SVM = Support Vector Machine, Vladimir Naumovich Vapnik (1979). Una maquina SVM es un algoritmo de aprendizaje automático supervisado que puede emplearse para fines de clasificación y regresión de dos grupos de datos.

KNN = KNeighbors, E. Fix and J.L. Hodges (1951), Es un método de clasificación supervisada, aprendizaje basado en un conjunto de entrenamiento y prototipos.

LR = Logistic Regression, M. Dolores Fiuza, C. Rodriguez Perez (2000). Es la ley experimental o fórmula matemática que traduce la relación entre variables correlacionadas.

CHURN = Fuga del cliente, Cambridge Advanced Learner´s Dictionary (2013). El número de clientes que deciden dejar de usar el servicio ofrecido por una empresa en detrimento de otra, habitualmente porque ofrece un mejor servicio o precio" y "churn rate" como "el porcentaje de clientes que dejan de utilizar un servicio, calculado en función del total de número de clientes en un determinado periodo de tiempo".

HFC = Hybrid Fiber Coaxial, Henry Bolivar, Hector Cala N. (2005). Una red HFC es una red de telecomunicaciones por cable que combina la fibra óptica y el cable coaxial como soportes de la transmisión de las señales. Se compone de 4 partes claramente diferenciadas, así como las redes telefónicas, cabecera, red troncal, red de distribución y red de acometida de los abonados.

CAPITULO III: METODO

3.1. TIPO Y DISEÑO

3.1.1. Tipo

La siguiente investigación es de tipo descriptivo, Marroquín Peña (2012), también conocida como la investigación estadística, se describen los datos y características de la población o fenómeno en estudio.

3.1.2. Diseño

La presente investigación se realizó bajo el diseño cuasi experimental, Marroquín Peña (2012), Cuasi Experimental con dos grupos no equivalente, conpre test y post test.

Ilustración 14: Diseño Cuasi Experimental

ANTES	VARIABLE	DESPUES
GC_1	-	GC_2
GE_1	X	GE_2

Donde:

G.E. : Grupo Experimental
G.C. : Grupo de Control
GC_1 Y GE_1 Pre Test
GC_2 Y GE_2 Post Test
X : Manipulación de Variable

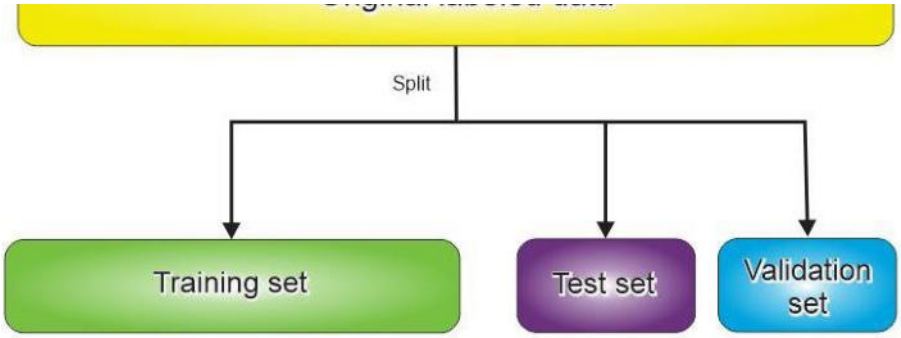
Fuente: Elaboración Propia

Los grupos de observación y testeo se harán mediante la aplicación de una Validación Cruzada, que da cabida a esta investigación debido que de la base de datos original se harán

particiones para entrenar, testear y validar el modeloexperimental, este procesamiento determinará el margen de error del modelo.

Para la red neuronal se utilizarán 2 tipos de data, el primer conjunto será de Testeo dentro del cual existe la partición de entrenamiento donde se usará el 70% (4476) de la data y para la prueba el 30% (1919), sin embargo, el segundoconjunto será sólo de validación y se usara el 100% de datos (398), esto generará confianza y precisión en el modelo de machine Learning.

Ilustración 15: Validación Cruzada



Fuente: Elaboración Propia

3.2. NIVEL DE INVESTIGACIÓN

Este proyecto tiene es de nivel predictivo, Whitney (1970), consiste en prever situaciones futuras, a partir de estudios exhaustivos de la evolución dinámica de loseventos, de su interrelación con el contexto, de las fuerzas volitivas de los actores que intervienen, y del estudio de las probabilidades de que algunos de esos eventos pudieran presentarse.

La investigación pretende realizar un pronóstico de los clientes latentes a desistir elservicio HFC, las múltiples variables serán analizadas para determinar su valor en la base de datos y llegar a obtener las variables más significativas para aplicar elmodelo predictivo.

3.3. POBLACION Y MUESTRA

3.3.1. Población

Arias (2012). La población es un conjunto finito o infinito de elementos con características comunes para las cuales serán extensivas las conclusiones de la investigación.

La población de América Móvil Perú SAC, comprende de los años 2014 hasta el 2020, que es el inicio de la empresa en la región Moquegua.

Se muestra el siguiente gráfico como población de la suba de clientes por año en la región Moquegua.

Tabla 8: Población

POBLACIÓN	
Año	**Clientes**
2014	1265
2015	796
2016	847
2017	596
2018	1612
2019	1279
2020	398
TOTAL	**6793**

Fuente: Sistema de Gestión de Administración – América Móvil Perú SAC.

3.3.2. Muestra

Tamayo y Tamayo (2006), define la muestra como: “el conjunto de operaciones que se realizan para estudiar la distribución de determinados caracteres en totalidad de una población universo, o colectivo partiendo de la observación de una fracción de la población considerada.

La muestra para la presente investigación será el 100% de los clientes registrados dado que los datos disponibles no son muchos y a su vez se tiene acceso a la data, en el periodo comprendido entre los años 2014 al año 2020, a su vez un mayor uso de datos en machin learning indica un mejor índice de precisión en el modelo.

3.4. TECNICAS E INSTRUMENTOS PARA LA RECOLECION DE DATOS

La siguiente investigación usa la técnica de extracción de datos, dado que extrae ladata del Sistema de Gestión de Administración – América Móvil Perú Sac, comprendida entre los años 2014 al 2020 de la ciudad de Moquegua, sin embargo,la metodología CRISP DM hace mención en las fases de desarrollo y validación, elprocesamiento como visión holística en los algoritmos estadísticos tiene como objetivo aplicar modelos estructurados para procesar y adquirir nuevo conocimiento (Huber et al., 2019).

3.4.1. Técnicas e Instrumentos

De acuerdo al criterio de Arias (1999) las técnicas de recolección de datos “son las distintas formas o maneras de obtener la información” y los instrumentos “son los medios materiales para recoger y almacenar la información”.

Machine Learning contempla un conjunto de técnicas para el desarrollo de un modelo predictivo, para esta investigación se aplicará el modelo de Red Neuronal y como herramienta de pruebas se aplicará el software WEKA para modelar el algoritmo.

La ventaja de programación en Python, es el uso de librerías como Panda, Matploptlib, Keras, Numpy que son técnicas para la manipulación de la data.

3.4.2. Recolección de Datos

La importancia de esta fase tiene el límite de la muestra, la data tiene como alcance todo el mercado de telecomunicaciones, sin embargo, por motivos en el sector empresarial en la región de Moquegua, sólo está disponible la data dela ciudad de Moquegua.

El siguiente recuadro refleja un diccionario de la data pre procesada con el objetivo de conocer dichas variables de entradas, así mismo, saber el tipo de resultado que se obtendrá para posteriormente interpretarlo.

Tabla 9: Descripción de la Variable Independiente

VARIABLE	DESCRIPCIÓN	VALORES
Año Instalación	Año de atención donde se da de alta al cliente	Del año 2014 al 2020

Tiempo activo	Indicador que mide el tiempo en semestres, según estipulado por el contrato, desde el momento en que el cliente inicia el servicio hasta su última fecha activa	Comprende los semestres: de 0 a 6 meses de 7 a 12 meses de 13 a 18 meses de 19 a 24 meses de 25 a 30 meses … de 79 a 84 meses
Tipo Cliente	Indicador que define o segmenta al cliente según evaluación de sistema	1 = CLIENTE A 2 = CLIENTE B 3 = CLIENTE C 4 = EMPRESA 5 = GRAN EMPRESA 6 = NEGOCIO 7 = PYME 8 = RESIDENCIAL 9 = SIN DEFINIR
Solicitud de Producto	Se mide la cantidad de servicios del cliente.	1 = 1 Dirección 2 = 2 Direcciones 3 = 3 Direcciones
Portabilidad	Se clasifica el servicio solicitado si proviene deun operador	0 = Alta nueva 1 = Portabilidad
Clasificación Internet	Se clasifica según la velocidad contratada porel cliente	0 = No Tiene 1 = 1000 Kbps a 30 MB 2 = 31 MB a 100 MB 3 = 101 MB a 400 MB
Clasificación Telefonía	Se clasifica según el plan de telefonía contratada por el cliente	0 = No Tiene 1 = 100 min a 199 min 3 = 200 min a 599 min 4 = 600 min a mas

Clasificación Cable	Se clasifica según el plan de cable contratadapor el cliente	0 = No Tiene 1 = Plan Básico 1 decodificador 2 = Plan Básico 2 decodificadores o mas 3 = Plan Avanzado 1 decodificador 4 = Plan avanzado 2 decodificadores o mas 5 = Plan Superior 1 decodificador 6 = Plan Superior 2 decodificadores o mas
Cantidad de Mantenimientos	Indicador que mide el total de mantenimiento hasta la fecha solicitados por el cliente	Cantidad de mantenimientos solicitada por el cliente
Detalle servicio atendido	Permite identificar que parte del servicio tiene afectado el cliente	0 = INTERNET 1 = TELEFONIA 2 = CABLE 3 = BAJA 4 = OTROS 5 = EXTERNO 6 = CONFIGURACION
Código Plano	Indicador que permite saber a que plano o zona pertenece determinado cliente	1 = MQMN001 2 = MQMN002 3 = MQMN003 4 = MQMN004 - Hasta el MQMN029
Detalle de Pago	Indicador que proporciona información sobre el pago del cliente	Retraso de Pagoal día en Pago
Baja	Indicador que permite saber si el cliente dio de baja o no el servicio	0 = Dio de baja1 = Activo

Fuente: Elaboración Propia

La data indica los factores cualitativos, la importancia del cuadro tiene como finalidad conocer el tipo de variable que se obtendrá

en el algoritmo, posteriormente se convertirán en variables cuantitativas para conocer su gradoestadístico.

Tabla 10: Altas y Bajas según SGA

Año	Alta	Baja
2014	1265	58
2015	796	360
2016	847	480
2017	596	453
2018	1612	449
2019	1279	753
2020	6973	2840

Fuente: Sistema de Gestión de Administración

3.5. MATRIZ DE CONSISTENCIA

TITULO: MODELO DE PREDICCION DE CLIENTES QUE DARAN DE BAJA EL SERVICIO HFC BASADO EN REDES NEURONALES EN LA CIUDAD DEMOQUEGUA

Tabla 11: Matriz de Consistencia

<table>
<tr><th>PROBLEMA</th><th>OBJETIVOS</th><th>HIPOTESIS</th><th>VARIABLE</th><th>DISEÑO</th></tr>
<tr>
<td>I. General
¿Cuál es el impacto en el desarrollo de un modelo de predicción de clientes que darán de baja el servicio HFC?

I. Específicos
¿Se podrá identificar al cliente que dará de baja el servicio HFC aplicando redes neuronales?

¿Cuál será el patrón de comportamiento del cliente que dará de baja el servicio HFC?</td>
<td>O. General
Desarrollar un modelo de predicción de clientes que dará de baja el servicio HFC.

O. Específicos
Desarrollar un modelo de red neuronal para el reconocimiento de patrones de clientes que dará de baja el servicio HFC.

Validar el modelo de predicción de clientes que dará de baja el servicio HFC a través del caso de estudio: América Móvil Perú SAC en la ciudad de Moquegua.</td>
<td>H. General
Los modelos de redes neuronales afectan positivamente el desarrollo de un modelo de predicción de clientes.

H. Específicos
Las redes neuronales afectan positivamente el desarrollo de un modelo de predicción de clientes.

La predicción de clientes que dará de baja el servicio HFC se validará mediante el caso de estudio: América Móvil Perú SAC.</td>
<td>V. Independiente
Factores Asociados para la predicción de bajas del servicio HFC

Dimensiones:
- Tiempo
- Cliente
- Producto
- Atención
- Zonal
- Pago
- Deseable

V. Dependiente
Minería de Datos – CRISP DM

Dimensiones:
- Comprensión del Negocio
- Comprensión de Datos
- Preparación de Datos
- Modelado
- Evaluación
- Implantación</td>
<td>Tipo de Investigación
Cuantitativa de tipo Descriptivo

Diseño de Investigación
Cuasi experimental
<table>
<tr><th>Antes</th><th>Variable</th><th>Después</th></tr>
<tr><td>GC1</td><td>-</td><td>GC2</td></tr>
<tr><td>GE1</td><td>X</td><td>GE2</td></tr>
</table>
Población
6973 clientes HFC de América Móvil Perú SAC en Moquegua

Muestra
100% de clientes HFC de la ciudad de Moquegua</td>
</tr>
</table>

				TÉCNICAS E INSTRUMENTOS **Técnica** Redes Neuronales Artificiales **Instrumento** Recolección de Datos

Fuente: Elaboración Propia.

3.6. TABLA DE DATOS

En la siguiente tabla se muestra, las variables de entrada, los datos antes de ser normalizados, y ser convertidos a datos cuantificables.

Tabla 12: Tabla de datos sin normalizar

TIEMPO ACTIVO	AÑO INST.	COD PLANO	TIPO CLIENTE	PORTA	DETALLE SERVICIO	SOLICITUD DE SERVICIO	INTERNET	TELEFONIA	CABLE	CANTIDAD MANTO	DETALLE PAGO	BAJA
78	2014	MQMN008	CLIENTE C	HFC - SISACT INSTALACION PAQUETES TODO CLARO DIGITAL	INTERNET	3 PLAY	HFC InternetFijo 10Mbps	100 minutos multidestino	CLARO HDTV BASICO (1 DECO BASICO HD)	3	AL DIA EN PAGO	ACTIVO
22	2014	MQMN004	RESIDENCIAL	HFC - SISACT INSTALACION PAQUETES TODO CLARO DIGITAL	INTERNET	3 PLAY	Internet 30 Mbps	200 minutos multidestino	CLARO HDTV BASICO (1 DECO BASICO HD)	1	RETRASO DE PAGO	BAJA

Fuente: Elaboración Propia.

En la siguiente tabla se muestra un ejemplo de los datos de las 13 variables, ya normalizados.

Tabla 13: Tabla de datos normalizada

TIEMPO ACTIVO	AÑO INST	COD PLANO	TIPO CLIENTE	PORTA	DETALLE SERVICIO	SOLICITUD DE SERVICIO	INTERNET	TELEFONIA	CABLE	CANTIDAD MANTO	DETALLE PAGO	BAJA
13	2014	8	3	1	0	3	1	1	1	3	2	1
4	2014	4	8	1	0	3	1	2	1	1	1	0

Fuente: Elaboración Propia.

3.7. OPERACIONALIZACIÓN DE VARIABLES

3.7.1. Definición de las variables

Hernández, Fernández y Baptista (2010). Señalan que una variable es una propiedad que puede variar y cuya variación es susceptible de medirse u observarse. De manera que entendemos como cualesquiera característica, propiedad o cualidad que presenta un fenómeno que varía, en efecto puede sermedido o evaluado.

3.7.2. Clasificación de las variables

Otra perspectiva que es relevante e importante es de Ander (1982) agrupa teniendo en cuenta a su naturaleza y características en: cualitativas y cuantitativas; continuas y descontinuas; dependientes e independientes; exploratorias o externas y generales; intermedias y empíricas.

Tabla 14: Vista de Variables

VARIABLE	INDICADOR	ESCALA
Dependiente: Modelo de predicción	Accuracy Precision Sensitivity Recall	[0 – 1]
Independiente: Deserción de clientes	Baja	[0 – 1]

Fuente: Elaboración Propia

3.7.2.1. Definición de variable dependiente

Actúa como efecto de una causa que ejerce coerción. Cabe precisar que éstas designan las variables a explicar, los efectos o resultados respecto a los cuales hay que buscar un motivo o razón de ser. De manera podemos denominar variable efecto o condicionada, es aquella que es afectada por la presencia o acción de la variable independiente en los resultados. Son las que el estudioso observa o mide, el propósito de esta observación es determinar si la variable independiente ha generado o no los cambios anunciados en las hipótesis.

3.7.2.2. Definición de variable independiente

Tiene el dominio causal, es la variable que se presume que es la que causa o afecta a la otra en los resultados. Que según Pino (2010) variable independiente es aquella que el experimentador modifica a voluntad para averiguar si sus modificaciones provocan o no cambios en las otras variables. Recuerde que la variable dependiente es la que toma valores diferentes en función de las modificaciones que sufre la variable independiente. En consecuencia, la variable independiente ejerce influencia o causan efecto en otras variables llamadas dependientes, y son las que permiten explicar a éstas. Además, son aquellas que dentro de la relación causal que propone una hipótesis, se determinan como causas. Estas variables en un experimento, son manipuladas por el experimentador: la finalidad de este control directo es ver si genera cambios en la otra variable relacionada.

3.7.3. Variable Independiente

Tabla 15: Operacionalización de la Variable Independiente

VARIABLE	DEFINICION		DIMENSION	INDICADOR	ESCALA
	CONCEPTUAL	OPERACIONAL			
DESERCION DE CLIENTES	Es la que permite identificar que clientes dan de baja el servicio HFC o continúan con el servicio HFC.	Son aquellas condiciones que determinan si el cliente dará de baja el servicio a solicitud propia o por retraso en sus pagos, o continuará con el servicio contratado.	Tiempo	Año Instalación	Del año 2014 al 2020
				Tiempo activo	Rango del 1 al 14
			Cliente	Tipo Cliente	Rango del 1 al 9
			Producto	Solicitud de Producto	Rango del 1 al 3
				Portabilidad	Rango del 0 al 1
				Clasificación Internet	Rango de 0 a 3
				Clasificación Telefonía	Rango de 0 a 3
				Clasificación Cable	Rango de 0 a 6
			Atención	Cantidad de Mantenimientos	Rango del 0 al 53
				Detalle servicio Atendido	Rango del 1 al 7
			Zonal	Código Plano	Rango del 1 al 29
			Pago	Detalle de Pago	Rango del 1 al 2
			Deseable	Baja	Rango del 0 al 1

Fuente: Elaboración Propia

3.7.4. Variable Dependiente

Tabla 16: Operacionalización de la Variable Dependiente

VARIABLE	DEFINICION		DIMENSION		INDICADOR	ESCALA
	CONCEPTUAL	OPERACIONAL				
MODELOS DE PREDICCION	Es el modelo que agrupa una variedad de técnicas estadísticas, aprendizaje automático y minería de datos.	El análisis de los efectos predictivos a través del porcentaje de exactitud, permitirá influir, determinar o informar en los procesos de la organización en la que participe.	Comprensión del Negocio	- Objetivos del Negocio - Objetivos del DM - Plan de Proyecto - Valorización de la Situación	- Score F1 - Accuracy - Sensitivity - Precision	Rango del 0 al 1
			Comprensión delos Datos	- Recolección - Descripción - Exploración - Calidad		
			Preparación delos Datos	- Selección - Limpieza - Estructuración - Integración		
			Modelado	- Técnica - Construcción - Evaluación		
			Evaluación	- Revisión		
			Implantación	- Plan de Implantación - Monitoreo - Revisión		

Fuente: Elaboración Propia

3.8. MÉTODOS Y TÉCNICAS PARA LA PRESENTACIÓN Y ANÁLISIS DE DATOS

3.8.1. Análisis de Datos

Este análisis tiene como tareas específicas, inicialmente en la extracción de datos para el procesamiento, es paso para interpretar la información y conocerel tipo de análisis general de la investigación, en esta fase tiene impacto en el rendimiento del algoritmo de clasificación, algunas técnicas de extracción y análisis del pre procesamiento fueron usadas en **Weka, Knime, Scikit-Learnm,R** como fin, este paso es reducir el procesamiento y evaluar las variables relevantes (Bilalli et al., 2019).

(Hu et al., 2020), Como modelo predictivo aplicando redes neuronales y árbolesde decisión a un conjunto de datos, basa su información en la adquisición de compras, el intervalo de compra – cliente es determinado por el tiempo, este periodo refleja si el cliente está próximo a realizar una siguiente compra.

Se realizo la investigación, en esta parte con la finalidad de entender los procesos internos y las variables que tendrían un mayor impacto en el resultadode la investigación, inicialmente esta se vino enfocando en base al tipo de mantenimiento que los clientes solicitaban durante todo su periodo de servicio activo, posteriormente se cambió el enfoque de la data a la cantidad de mantenimiento que el cliente había solicitado el cliente durante el tiempo de servicio, como fase de interpretación de data, se obtiene la siguiente tabla la cual hace mención al tipo de rango en el que será evaluado en la faseprocesamiento, este conjunto de datos lo convierte a una variable cuantitativa.

Tabla 17: Valores de los Datos

DIMENSION	INDICADOR	RANGO
Tiempo	Año Instalación	Del año 2014 al 2020
	Tiempo activo	Rango del 1 al 14
Cliente	Tipo Cliente	Rango del 1 al 9
Producto	Solicitud de Producto	Rango del 1 al 3
	Portabilidad	Rango del 0 al 1
	Clasificación Internet	Rango de 0 a 3
	Clasificación Telefonía	Rango de 0 a 3
	Clasificación Cable	Rango de 0 a 6
Atención	Cantidad de Mantenimientos	Rango del 0 al 53
	Detalle servicio Atendido	Rango del 1 al 7
Zonal	Código Plano	Rango del 1 al 29
Pago	Detalle de Pago	Rango del 1 al 2
Deseable	Baja	Rango del 1 al 2

Fuente: Elaboración Propia

3.8.2. Preparación de Datos

Esta fase como tarea de realizar limpieza o categorización de variables puede ser realizada con la función **StandardScaler**, este procesamiento minimiza la carga en el procesamiento del algoritmo, tal forma también analiza si el las variables de entrada están dentro de la media de los cuartiles y muestra la dispersión del conjunto de datos entre la media y la desviación estándar, la estadística del ruido y pertenencia de la variable a la muestra es analizada porlos cuartiles.

La interpretación estadística se aplica con las siguientes fórmulas de la media y desviación estándar, estos procesos se representan como cuartiles en el diagrama de cajas.

$$z = \frac{x - \mu}{\frac{\sigma}{N}}$$

$$\mu = \frac{1}{N} \sum_{i=1}^{N} (X_i)$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2}$$

El siguiente recuadro muestra la dispersión y tendencia central del conjunto dedatos.

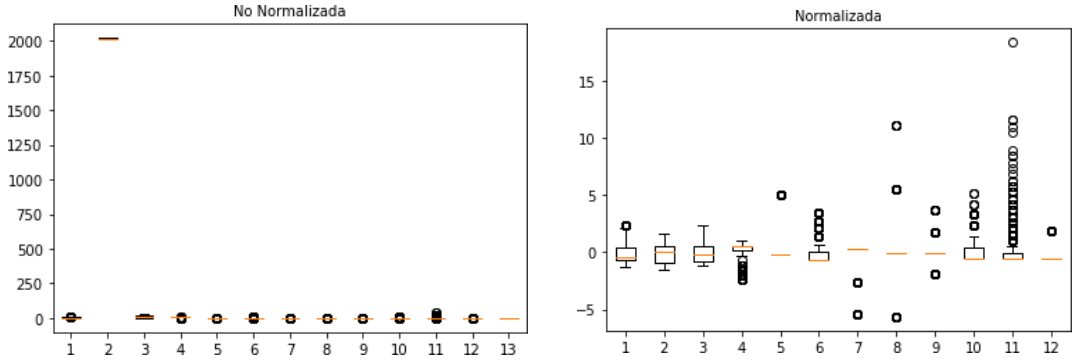
Tabla 18: Descripción Estadística

	count	mean	std	min	25%	50%	70%	max
TIEMPO ACTIVO	6793.0	4.4828	3.6331	0.0	2.0	3.0	6.0	13.0
FECHA INSTALACION	6793.0	2016.8719	1.9419	2014.0	2015.0	2017.0	2018.0	2020.0
CODIGO PLANO	6793.0	10.6683	7.9961	1.0	4.0	9.0	15.0	29.0
TIPO CLIENTE	6793.0	6.6509	2.3463	1.0	7.0	8.0	8.0	9.0
PORTABILIDAD	6793.0	1.0359	0.1861	1.0	1.0	1.0	1.0	2.0
DETALLE SERVICIO	6793.0	0.9284	1.5157	0.0	0.0	0.0	1.0	6.0
TIPO SOLICITUD SERVICIO	6793.0	2.8987	0.3466	1.0	3.0	3.0	3.0	3.0
INTERNET	6793.0	1.0089	0.1842	0.0	1.0	1.0	1.0	3.0
TELEFONIA	6793.0	1.0413	0.5335	0.0	1.0	1.0	1.0	3.0
TIPO CABLE	6793.0	0.5795	1.0464	0.0	0.0	0.0	1.0	6.0
CANTIDAD MANTENIMIENTO	6793.0	1.0089	1.9397	0.0	0.0	0.0	1.0	39.0
COSTO SERVICIO	6793.0	1.2197	0.4141	1.0	1.0	1.0	1.0	2.0
Class	6793.0	0.5817	0.4933	0.0	0.0	1.0	1.0	1.0

Fuente: Elaboración Propia

Para la representación del cuadro estadístico se aplica el diagrama de cajas o cuartiles, que refleja los datos que no deberían considerarse en la muestra, el diagrama observa el tipo de procesamiento para una data general y la misma procesada para la normalización de la muestra.

Ilustración 16: Comparación del Conjunto de Datos



Fuente: Elaboración Propia

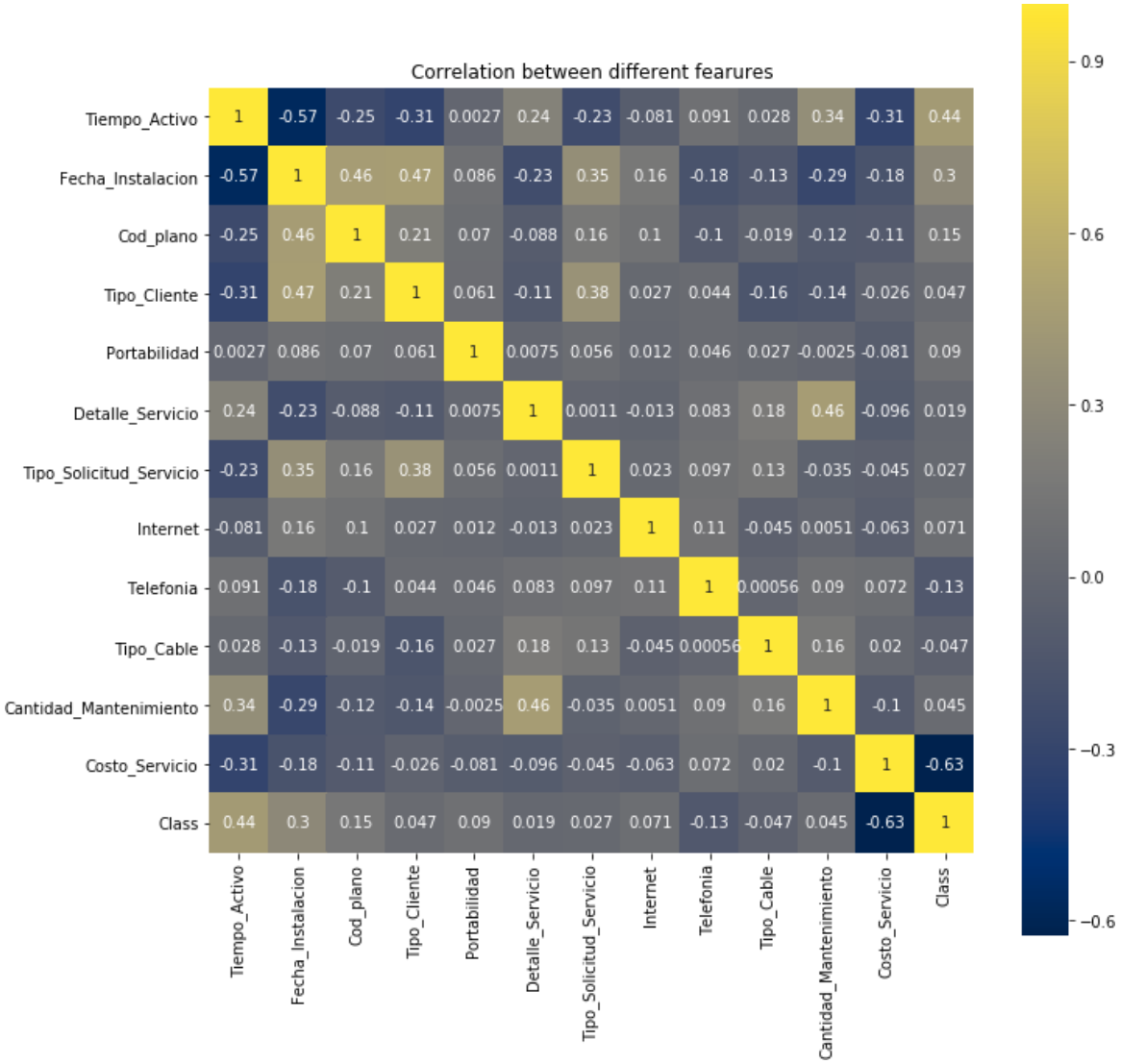
En esta fase de preparación de datos, también se observa el tipo de variable que ingresará a la data, la complejidad del modelo predictivo es en base a las variables, la cantidad de variables no siempre determina que el modelo sea óptimo, sin embargo, saber porcentualmente la importancia de la variable en elconjunto de datos nos ayuda a tomar decisiones sobre la dependencia de las variables, en **Scikit-Learn** la aplicación de **Importance Feature** da solución a este problema aplicando la correlación entre las variables.

En esta fase de preparación de datos, también se observa el tipo de variable que ingresará a la data, la complejidad del modelo predictivo es en base a las variables, la cantidad de variables no siempre determina que el modelo sea óptimo, sin embargo, saber porcentualmente la importancia de la variable en elconjunto de datos nos ayuda a tomar decisiones sobre la dependencia de las variables, en Scikit-Learn la aplicación de Importance Feature da solución a este problema aplicando la correlación entre las variables.

Se obtiene que la correlación de variables mas resaltantes son la fecha de instalación con el código de planos, así como la fecha de

instalación con el tipode cliente, la cantidad de mantenimientos que el cliente ah solicitado en el tiempo que ah estado activo el servicio, cantidad de mantenimientos con eldetalle de servicio brindado en la atención.

Ilustración 17: Correlación de las Variables

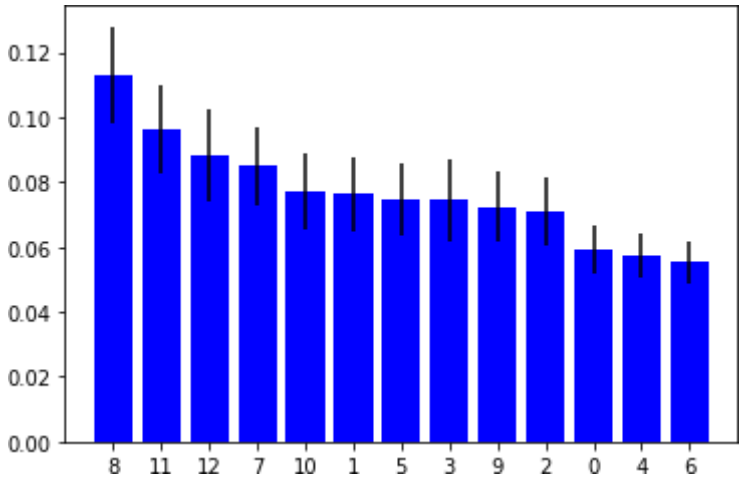


Fuente: Elaboración Propia

Como resultado de la correlación del conjunto de datos se construye un modelo de la importancia de variables, su

construcción se basa en el número de features, tipo de clase a predecir, cantidad de step y aplicando una validación cruzada, este pequeño modelo de Importance Feature es aplicado en clasificador de árboles de decisión, como resultado se obtiene la siguiente ilustración.

Ilustración 18: Feature Importance



Fuente: Elaboración Propia

Su interpretación representa el porcentaje de relevancia de todas las variablespara la construcción del modelo, como resultado se obtiene que las variables 8 (internet), 11 (cantidad de mantenimientos), 2 (fecha de instalación), 7(tipo de solicitud de servicio), 10 (cable) y 1(tiempo activo) son las más relevantes, sin embargo, el punto de convergencia del total de la muestra se refleja en la variable 6 (detalle de servicio), como resultado estas variables son óptimas parapasar a la fase de modelamiento del modelo predictivo.

3.8.3. Construcción del Modelo

Esta fase, se diseña la arquitectura de la red neuronal, sabiendo que como datos de entrada se tiene 12 variables, que representa

toda la base de datos, los pesos se dan por el clasificador kernel_initializer que de forma automática yaleatoria da valores entre -0.5 a +0.5, el tipo de conexiones y tamaño de neuronas en las capas ocultas, como primer experimento del tamaño de la ANN, se pone a prueba 1 capa oculta con los optimizadores Adam, SGD y Rmsprop,teniendo las funciones de activación tangencial y sigmoidal respectivamente, como resultado del experimento se formuló el siguiente recuadro.

Tabla 19: Experimento 1 - Capa 1

CAPA 0			
Número de Neuronas	Optimizador		
	ADAM	SGD	RMSPROP
3	0.98	0.32	0.99
6	0.93	0.61	0.99
9	0.95	0.85	0.98

Fuente: Elaboración Propia

Del primer experimento se puede deducir que en la capa oculta número 1, el tamaño de la red neuronal (3, 6 y 9) tienen un alto porcentaje de precisión siendo este el optimizador Rmsprop el más alto con 3 neuronas de entrada, con un porcentaje de 0.99.

Durante el experimento 2 se agregó otra capa oculta para conocer el grado de precisión entre los 3 optimizadores, como resultado se obtuvo el siguiente recuadro.

Tabla 20: Experimento 2 - Capa 1

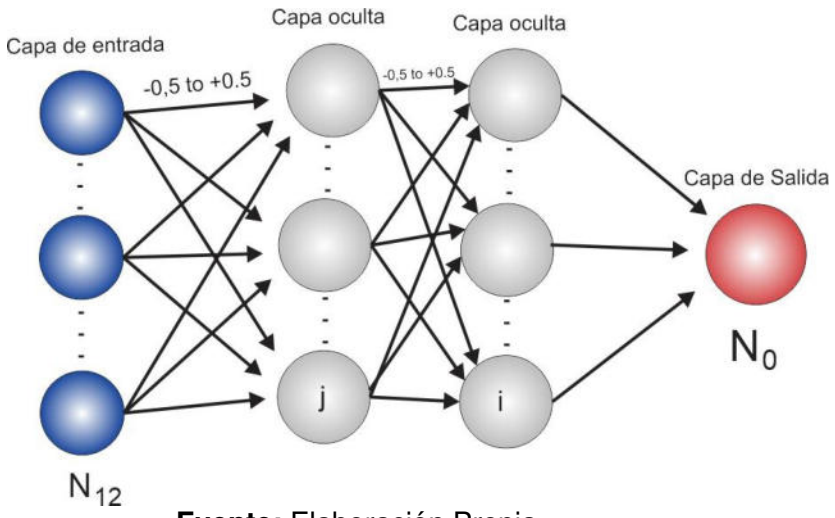
CAPA 1			
Número de Neuronas	Optimizador		
	ADAM	SGD	RMSPROP
3	0.95	0.86	0.93
6	0.96	0.88	0.95
9	0.95	0.92	0.94

Fuente: Elaboración Propia

Se deduce que, del siguiente experimento pese a incrementar otra capa ocultaen el procesamiento, los resultados arrojan con mejor precisión una optimización Adam con 6 neuronas en la capa oculta 2, sin embargo, según el primer experimento la optimización Rmsprop, sigue teniendo resultados favorables.

De estos experimentos, se concluye que en ambos algoritmos de optimización aportan grandes resultados y no existe problemas en la convergencia.

Ilustración 19: Arquitectura de la red Neuronal



Fuente: Elaboración Propia

3.8.4. Evaluación del Modelo

Para la evaluación del modelo, se necesita particionar la data en 3 o 2 conjuntos,conjunto de entrenamiento, prueba y validación, estas etiquetas miden laprecisión del modelo de ML, estos datos sufren un proceso shuffle para no repetir la data, en keras se aplica una validación cruzada aplicando la siguiente línea de código.

$$X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30, random_state = 0)$$

Esta evaluación particiona el conjunto de datos en 30% para realizar las pruebas, como resultado de este proceso en el modelo se obtienen las siguientes métricas.

Ilustración 20: Reporte de Métricas

Reporte de Metricas:

	precision	recall	f1-score	support
0	1.00	0.97	0.99	866
1	0.98	1.00	0.99	1053
avg / total	0.99	0.99	0.99	1919

Fuente: Elaboración Propia

Parte de las métricas es la evaluación del grado de exactitud y precisión del modelo, esta evaluación es paralela del conjunto de entrenamiento y prueba, como resultado se obtienen el siguiente

Ilustración 21: Modelo de Exactitud

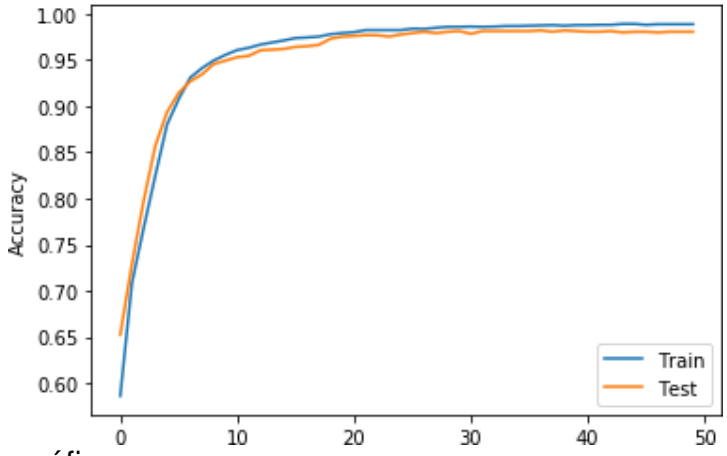
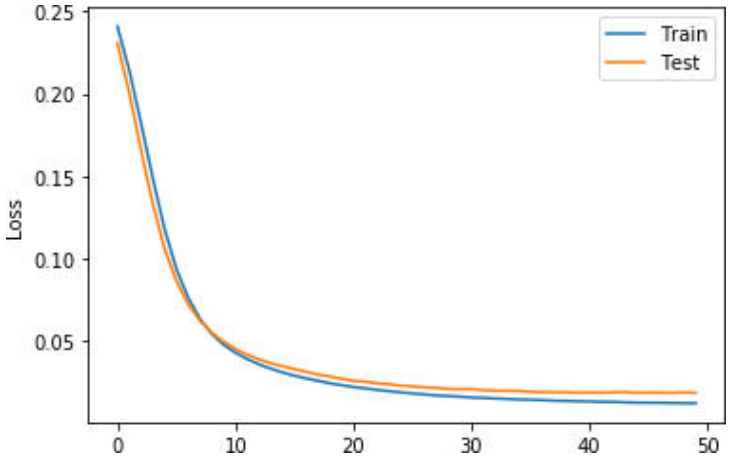


gráfico.

Fuente: Elaboración Propia

En paralelo al modelo de exactitud es ver el grado de perdida de la red neuronalpara verificar problemas de overfiting o underfiting, estos problemas de sobre ajuste de evalúan con la convergencia. El resultado del experimento 1 y 2, no mostraron problemas de ajuste, sin embargo, puede surgir con la implementación del modelo, como resultado se obtuvo la siguiente gráfica.

Ilustración 22: Modelo de Perdida



Fuente: Elaboración Propia

Este resultado demuestra que dentro del punto 0 – 10 se da la convergencia, teniendo como resultado la disminución del error.

3.8.5. Explotación

Esta fase es la implementación de la ANN, en la fase de experimentación del modelo, se observa el particionamiento del conjunto de datos en 3 grupos, el grupo de entrenamiento y prueba mostraron resultados favorables, sin embargo, la implementación del modelo se da con el conjunto de validación el cuál es el grupo aparte que no entro al modelo de ML, este conjunto perteneceal año 2020, del cual se pretende predecir los posibles clientes en fuga, la explotación del modelo tuvo los siguientes resultados, en la siguiente ilustraciónse obtiene 98% para los resultados predictivos

Ilustración 23: Resultados Predictivos

0	0.00	0.00	0.00	5
1	0.99	1.00	0.99	393
avg / total	0.98	0.99	0.98	398

Fuente: Elaboración Propia

3.9. VALIDACIÓN Y CONFIABILIDAD DE LOS INSTRUMENTOS

3.9.1. Validación cruzada

Se emplea para estimar el error en el conjunto de datos de test del modelo y evaluar la capacidad predictiva en el conjunto de validación, conocido como resampling, que son empleados para la validación de la data, en general la forma de evaluar esta validación es aplicando el error cuadrático medio. Este tipo de validación, estima el test error rate dependiendo del tipo de procesamiento, otro método usado es por k-fold, es un proceso iterativo que consiste en dividir los datos de forma aleatoria en k grupos de aproximadamente el mismo tamaño. k-1 grupos se emplean para entrenar el modelo y uno de los grupos se emplea como test, este proceso se repite k veces utilizando un grupo distinto como test en cada iteración. El proceso genera k estimaciones del test error cuyo promedio se emplea como estimación final.

3.9.2. Matriz de Clasificación

Es un método que clasifica los valores de la predicción, está técnica evalúa los valores reales y predictivos del modelo, las fórmulas que emplean el modelo son las siguientes:

La precisión:

$$Precisión_i = \frac{N_{ii}}{\sum_{k=1}^{n} N_{ki}}$$

Recall:

$$Recall_i = \frac{N_{ii}}{\sum_{k=1}^{n} N_{ik}}$$

F-Score:

$$F - Score_i = \frac{2\,x\,Precision_i * Recall_i}{Precision_i + Recall_i}$$

Ilustración 24: Reporte de Métricas de Conjunto de Validación

Reporte de Metricas:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	5
1	0.99	1.00	0.99	393
avg / total	0.98	0.99	0.98	398

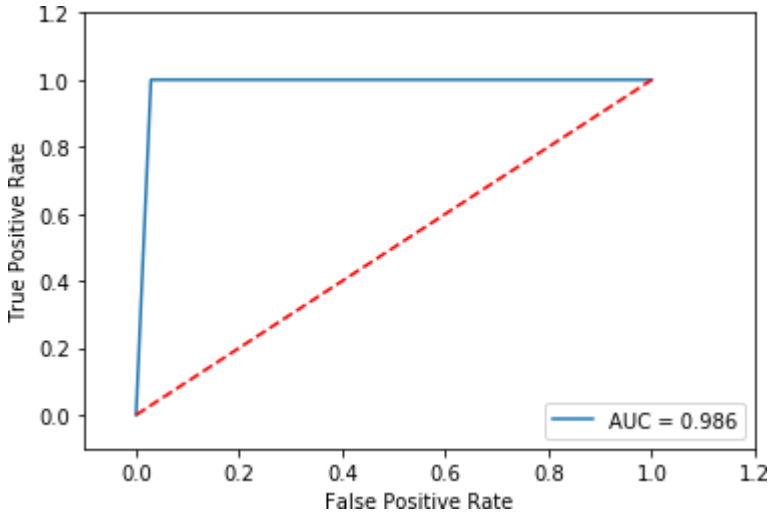
Fuente: Elaboración Propia

Este resultado es la precisión del conjunto de validación que no participó comoconjunto de entrenamiento y prueba en el modelo de la ANN.

3.9.3. Curva ROC

Esta técnica de evaluación dibuja la una curva ROC de predicción iniciando enla esquina superior izquierda en la coordenada (0,1), lo cual representa el 100%de sensibilidad según la escala de los fasos negativos y falsos positivos, el cuálpara llegar hacer un modelo óptimo debería sobrepasar esta curva (50%), como resultado del modelo ANN se obtuvo el 98% de precisión en la curva ROC.

Ilustración 25: Métrica ROC



Fuente: Elaboración Propia

CAPITULO IV: RESULTADOS

El experimento de partición por validación cruzada, en la fase del modelamiento arrojo como resultado de 96%, sin embargo se optó por aplicar 2 experimentos en el fase de modelamiento, el primer experimento fue de 1 capa oculta con 3, 6 y 9 entradas en la capa, aplicando algoritmos de optimización Adam, Sgd y Rmsprop, estos resultados fueron favorables, puesto que el conjunto de datos estaba semi balanceado lo cual evita problemas de balanceo, sin embargo, en el segundo experimento, se añadió otra capa oculta, teniendo un total de 4 capas, 1 de entrada,2 ocultas y 1 de salida, los resultados mostrando que una optimización por Adam obtiene una precisión del 98% que comparado a los demás modelos de optimización también muestran resultados favorables.

Del experimento se tiene como resultado que los factores determinantes para la baja del cliente están asociados con el tipo de producto donde resalta la variable internet y cable, el factor atención con las variables cantidad de mantenimientos y el tipo de solicitud de servicio, el factor tiempo con las variables fecha de instalación y tiempo activo del servicio.

Del experimento se obtiene que la correlación de variables más altas para la baja del servicio por parte del cliente para este modelo son la variable fecha de instalación con el tiempo activo, código de plano y tipo de cliente, la variable tipo de cliente con tipo de solicitud de servicio, la variable cantidad de mantenimientos con tiempo activo, detalle de servicio.

CAPITULO V: DISCUSION DE RESULTADOS

5.1. COMPROBACION DE HIPOTESIS

- El desarrollo de los modelos de ML permite que un sistema por sí mismo pueda automatizarse, descubrir patrones, relaciones de datos y genere conocimientos, esta competencia entre los algoritmos se determinan en nuestra vida en diferentes sectores, banca financiera, servicio de salud, mercadotécnica, industria entre otros,cada modelo da solución a un problema determinado. En la investigación el modelo de ML permitió alcanzar el objetivo trazado que se enfocó desde la necesidad de laempresa.

- La aplicación de redes neuronales tuvo gran alcance en el proyecto, alcanzo la capacidad de aprender tareas determinadas en la fase de modelamiento para entrenar, evaluar y validar la data, como resultado se obtuvo una precisión del 98%de precisión.

- Del modelo de ML, permite realizar nuevas tendencias de investigación aplicando múltiples algoritmos de clasificación aplicado al sector de telecomunicaciones, en referencia a la empresa América Móvil Perú SAC.

5.2. TRABAJOS FUTUROS

- Ir progresivamente con otras ciudades continuas a la ciudad de Moquegua, levantarinformación para que la base de datos crezca y el resultado de la predicción sea más óptimo.

- Utilizar la información que el modelo proporcione para determinar que clientes son más probables que dejen el servicio.

- Generar un modelo semejante al realizado en este proyecto, para la tecnología FTTH, presente en las ciudades de Ilo, Mollendo, Camaná, Juliaca y Quillabamba, teniendo en cuenta las variables que pueden modificarse debido al tipo de tecnología que es DTH para el servicio de cable.

- Generar un sistema que permita realizar seguimiento al cliente desde el momento que contrata el servicio, con la finalidad de detectar las deficiencias en el servicio que contribuyen a que el cliente decida dar de baja el servicio.

- Presentar el proyecto a gerencia de inteligencia comercial, gerente de ventas ClaroHogar, Gerente de atención al cliente, con la finalidad de que el modelo sea aprobado para que al ser implementado pueda contribuir a disminuir el flujo de clientes que dan de baja el servicio en América Móvil Perú.

CAPITULO IV: CONCLUSIONES

- Se concluye que los modelos de redes neuronales, afectan positivamente en el desarrollo de un modelo de predicción de cliente.

- Se identifica los patrones en el modelo de predicción de clientes, más resaltantes y demayor impacto que contribuyen a la baja del servicio HFC por parte del cliente.

- Se validó el modelo de ML con la data del Sistema de Gestión de Administración de laEmpresa América Móvil Perú SAC.

- Se desarrollo el modelo predictivo aplicando técnicas de machine learning, como resultado se obtuvo la predicción de los clientes, este modelo fue validado con las métricas de ML.

- Se implemento la estructura del modelo de redes neuronales, en la etapa de procesamiento de data y aplicando feature importance se obtuvo como resultado el cuadro estadístico de las variables más importantes, estas variables se determinan patrones de fuga.

- Se concluye que el modelo es adecuado y optimo, dado que se tiene un 98% de precisión, el cual es alto siendo este validado por las métricas de ML.

CAPITULO VII: RECOMENDACIONES

- Es importante al momento de realizar modelos predictivos conocer, entender e identificar los procesos para lograr filtrar variables que serán usados en el modelo, mediante un aprendizaje supervisado, conocer los valores de clase predictiva y uniformizado aplicando técnicas de muestreo para que de esta manera el modelo tengauna precisión alta y pueda generar una probabilidad de clasificación apropiada.

- Se recomienda que la base de datos este en continuo crecimiento para que el modelo pueda seguir aprendiendo y de esta manera se pueda experimentar con distintas capasocultas, distintas capas de activación.

- Tener presente que en este tipo de proyectos se encuentra la base de confidencialidad de datos del cliente, por lo que se tiene que tener cuidado para no ir en contra de las leyes promulgadas en el País.

- Desarrollar un sistema que pueda extraer directamente los datos del sistema que usa América Móvil Perú SAC, a su vez pueda limpiarlos, normalizarlos, clasificarlos y estospuedan ser usados directamente para detectar el abandono de clientes.

- Recomendar que las futuras investigaciones, puedan asemejarse a las variables propuestas debido al porcentaje de exactitud obtenido.

CAPITULO VIII: REFERENCIAS

Barrueta Meza, R. A., & Castillo Villarreal, E. J. P. (2018). *Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos*. http://hdl.handle.net/10757/626023

Bengio, Y., Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). *Continuous control with deep reinforcement learning* (Issue November). https://doi.org/10.1561/2200000006

Bilalli, B., Abelló, A., Aluja-Banet, T., & Wrembel, R. (2019). PRESISTANT: Learning based assistant for data pre-processing. *Data and Knowledge Engineering*, *123*(July), 101727. https://doi.org/10.1016/j.datak.2019.101727

Fielding, A. H. (1999). An introduction to machine learning methods. In *Machine Learning Methods for Ecological Applications*. https://doi.org/10.1007/978-1-4615-5289-5_1

Galván, A. R. M., & Navarro, K. R. C. (2015). Arquitectura de Big Data para la Predicciónde la Portabilidad Numérica en Empresas de Telefonía Móvil. *CEUR Workshop Proceedings*, *1478*, 106–116.

Group, I. E., Trends, E., Barrientos, F. S. R., Nemirovsky, M., Ojeda Villanueva, Á., &Bunny, D. B. (2013). Ontologies in the Context of Cloud Computing and Big DataOntologies and Conceptual Models for. *Sistemas de Informacion*, *XXVII*, 159.

Hu, X., Yang, Y., Chen, L., & Zhu, S. (2020). *Árbol y red neuronal*. 129–132.

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications - A holistic

extension to the CRISP-DMmodel. *Procedia CIRP*, *79*, 403–408. https://doi.org/10.1016/j.procir.2019.02.106

LEITE, K. et al. (2018). Brazilian Journal of Development. *J. of Develop*, *6*(1), 4303–4308.

Meza, A. richard. (2018). *Prediccion de fuga de clientres en una empresa de telefonía utilizando el algoritmo adaboost desbalanceado y la regresión logística asimétrica*.

http://repositorio.lamolina.edu.pe/bitstream/handle/UNALM/3245/meza-rodriguez- aldo-richard.pdf?sequence=1&isAllowed=y

Muñoz Rosas, F. J., & Álvarez Verdejo, E. (2009). Métodos de imputación para el tratamiento de datos faltantes: Aplicación mediante R/Splus. *Revista de Metodos Cuantitativos Para La Economia y La Empresa*, *7*(7), 3–30.

Besse, Juan (1999). Revista Bibliográfica de Geografía y Ciencias Sociales. UniversidadAutónoma de Barcelona.

Peñafiel, C., & Ávila, R. (2007). Inteligencia Artificial. *Inteligencia Artificial*, *2*(6), 1–33.https://doi.org/10.13140/2.1.3720.0960

Carrasco Díaz, Sergio (2006). *Metodología de la investigación científica*. Editorial SanMarcos, Lima.

Quiñones Huatangari, L., Ochoa Toledo, L., Kemper Valverde, N., Gamarra Torres, O., Bazán Correa, J., & Delgado Soto, J. (2020). Red neuronal artificial para estimar un índice de calidad de agua. *Enfoque UTE*, *11*(2), 109–120. https://doi.org/10.29019/enfoque.v11n2.633

Hernández, R. Fernández, C. y Baptista, L. (2010). Metodología de la Investigación. México: Mc Graw Hill, p. 143

De La Lama García, Alfredo (2005). Estrategias para elaborar investigaciones científicas.Ed. Trillas, México DF

Ander, Ezequiel. (1982) *Técnicas de investigación social*. Buenos Aires: EditorialHumanitas.

Pick, A. y López, A. (2002) *Cómo investigar en ciencia sociales.* México: Trillas.

Pino, Raúl. (2010). Metodología de la Investigación. Lima: Editorial San Marco, p. 134. Payam Refaeilzadeh, Lei Tang, Huan Lui . (2008). k-fold Cross-Validation, Arizona State University. USA

Damian Jorge Matich, (2001). Redes Neuronales: Conceptos básicos y aplicaciones. Universidad Tecnológica Nacional – Grupo de Investigación Aplicada. Rosario,Argentina.

Samuel A. (1959). Programming computers to learn experience, Universidad de Stanford. Juan J. Suarez Garcia (2017), Maquina de soporte vectorial SVM, escrito para

Numerentur.org

E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951) Author(s): B.

W. Silverman and M. C. Jones Source: International Statistical Review / RevueInternationale de Statistique, Vol. 57, No. 3 (Dec., 1989), pp. 233-238

M. Dolores Fiuza, C. Rodriguez Perez (2000). La regresión logística una herramienta versátil, unidad de investigación epidemiológica, las Palmas de Gran Canaria, IslasCanarias, España.

Henry Bolivar, Hector Cala N. (2005). Tecnología de redes de banda ancha redes HFC.Facultad de ingeniería eléctrica y electrónica, Cartagena.

Tamayo Tamayo, Mario (1998). El proyecto de investigación. Bogota

Tamayo y Tamayo, M (2006). Tecnicas de investigación (2da Edicion). Mexico: EditorialMc Graw Hill

Arias, F. (2012). El proyecto de investigación. Introduccion a la metologia científica. (6taedición). Caracas : Editorial Episteme.

CAPITULO IX: ANEXOS

Anexo 01: Flujo de Aprobación para anonimizar datos de Clientes

Ilustración 26: Flujo de Aprobación para anonimizar datos de Clientes



Fuente: Sistema de Requerimiento de Información

Anexo 02: Código para el análisis de Data

```
import pandas as pd
import numpy as np
import warnings
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

pd.set_option('display+.expand_frame_repr', False)
warnings.filterwarnings("ignore")

Path =  pd.read_csv('Analisis_Data_HFC.csv', delimiter=',')
print( Path.describe().T)

X = Path.iloc[:,0:12]
y = Path.iloc[:,-1]

#Path.head()
```

```
data = [ Path['Tiempo_Activo'], Path['Fecha_Instalacion'], Path['Cod_plano'], Path['Tipo_Cliente']
        Path['Portabilidad'], Path['Detalle_Servicio'], Path['Tipo_Solicitud_Servicio'], Path['Int
        Path['Telefonia'],Path['Tipo_Cable'],Path['Cantidad_Mantenimiento'],Path['Costo_Servicio']

plt.figure()
plt.title('No Normalizada', fontsize = 10)
plt.boxplot(data)

plt.show()
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=4)
```

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(X_train)

X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)

plt.figure()
plt.title('Normalizada', fontsize = 10)
plt.boxplot(X_train)
plt.show()
```

```
from sklearn.preprocessing import StandardScaler
import tflearn.data_utils as du
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.externals.six import StringIO
from IPython.display import Image
from sklearn.tree import export_graphviz
import pydotplus
```

```
data = pd.read_csv('Analisis_Data_HFC.csv')
data.info()
```

```
import seaborn as sns
correlation = data.corr()
plt.figure(figsize=(11,11))
sns.heatmap(correlation, vmax=1, square=True,annot=True,cmap='cividis')

plt.title('Correlation between different fearures')

['Tiempo_Activo', 'Fecha_Instalacion', 'Cod_plano', 'Tipo_Cliente',
 'Portabilidad', 'Detalle_Servicio', 'Tipo_Solicitud_Servicio', 'Internet',
 'Telefonia','Tipo_Cable','Cantidad_Mantenimiento','Costo_Servicio','Class']
```

```
X = data.iloc[:,0:12].values
y = data.iloc[:,-1].values
columns = ['Tiempo_Activo', 'Fecha_Instalacion', 'Cod_plano', 'Tipo_Cliente'
 'Portabilidad', 'Detalle_Servicio', 'Tipo_Solicitud_Servicio', 'Internet',
 'Telefonia','Tipo_Cable','Cantidad_Mantenimiento','Costo_Servicio','Class']
X_train , X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30,
```

```
import numpy as np
import matplotlib.pyplot as plt

from sklearn.datasets import make_classification
from sklearn.ensemble import ExtraTreesClassifier

# Build a classification task using 3 informative features
X, y = make_classification(n_samples=10000,
                           n_features=13,
                           n_informative=13,
                           n_redundant=0,
                           n_repeated=0,
                           n_classes=2,
                           random_state=0,
                           shuffle=False)

# Build a forest and compute the feature importances
forest = ExtraTreesClassifier(n_estimators=250,
                              random_state=0)

forest.fit(X, y)
importances = forest.feature_importances_
std = np.std([tree.feature_importances_ for tree in forest.estimators_],
             axis=0)
indices = np.argsort(importances)[::-1]

# Print the feature ranking
print("Feature ranking:")

for f in range(X.shape[1]):
    print("%d. feature %d (%f)" % (f + 1, indices[f], importances[indices[f]

# Plot the feature importances of the forest
plt.figure()
plt.title("Feature importances")
plt.bar(range(X.shape[1]), importances[indices],
       color="b", yerr=std[indices], align="center")
plt.xticks(range(X.shape[1]), indices)
plt.xlim([-1, X.shape[1]])
plt.show()
```

Anexo 03: Código del Modelo de la Red Neuronal

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import keras

from keras.models import Sequential
from keras.layers import Dense
from keras import initializers
from keras.utils.vis_utils import plot_model
from keras.optimizers import Adam
from matplotlib import pyplot
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from keras.wrappers.scikit_learn import KerasClassifier


#Data
dataset = pd.read_csv('Data_Train_HFC.csv')
X = dataset.iloc[:, 0:12].values
y = dataset.iloc[:, -1].values

#Validación Cruzada
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.30, random_state=0)

#Procesamiento
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)

def Model():
    classifier = Sequential()
    classifier.add(Dense(units = 12, activation = 'tanh', input_dim = 12))
    classifier.add(Dense(units = 3, activation = 'tanh'))
    classifier.add(Dense(units = 3, activation = 'tanh'))
    classifier.add(Dense(units = 1, activation = 'sigmoid'))
    classifier.compile(optimizer = 'sgd' , loss = 'mean_squared_error',  metrics = ['acc','mse'])
    return classifier

from keras.wrappers.scikit_learn import KerasClassifier
classifier = Model()
history = classifier.fit(X_train, y_train, validation_split=0.30, batch_size = 10, epochs = 50)

print('\n', history.history.keys())

y_pred = classifier.predict(X_test)
y_pred = (y_pred > 0.5)
```

```
from sklearn.metrics import classification_report, confusion_matrix
cm = confusion_matrix(y_test, y_pred)

print('Matriz de Confusión: \n', cm,'\n')
print('Reporte de Metricas: \n',classification_report(y_test,y_pred),'\n')
```

```
plt.plot(history.history['acc'])
plt.plot(history.history['val_acc'])
plt.title('Model accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(['Train', 'Test'], loc='lower right')
plt.show()
```

```
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('Model loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Train', 'Test'], loc='upper right')
plt.show()
```

```
from sklearn.metrics import roc_curve, auc
false_pos_rate, true_pos_rate, thresholds = roc_curve(y_test,y_pred)
roc_auc = auc(false_pos_rate, true_pos_rate,)
plt.plot(false_pos_rate, true_pos_rate,  label='AUC = %0.3f'% roc_auc)

plt.legend(loc='lower right')
plt.plot([0,1],[0,1],'r--')
plt.xlim([-0.1,1.2])
plt.ylim([-0.1,1.2])
plt.ylabel('True Positive Rate')
plt.xlabel('False Positive Rate')
plt.show()
```

```
from sklearn.preprocessing import StandardScaler
dataset = pd.read_csv('Data_Validation_HFC.csv')
X_valid = dataset.iloc[:, 0:12].values
y_valid = dataset.iloc[:, -1].values

c = sc.transform(X_valid)
new_prediction = classifier.predict([c])
new_prediction_class = (new_prediction>0.5)
prediction = (new_prediction>0.5)
#print(new_prediction)
```

```
header = ['Tiempo_Activo', 'Fecha_Instalacion', 'Cod_plano', 'Tipo_Cliente',
 'Portabilidad', 'Detalle_Servicio', 'Tipo_Solicitud_Servicio', 'Internet',
 'Telefonia','Tipo_Cable','Cantidad_Mantenimiento','Costo_Servicio','Class']
df = pd.DataFrame(dataset, columns=header)
df["Prediction"] = new_prediction
df.to_csv("DATA_WIFI(PREDICTION).csv", sep=",", index=False, float_format="%.6f")
```

```
cm = confusion_matrix(y_valid, prediction)

print('Matriz de Confusión: \n', cm,'\n')
print('Reporte de Metricas: \n',classification_report(y_valid,prediction),'\n')
```

```
X_valid = dataset.iloc[:, 0:12].values
y_valid = dataset.iloc[:, -1].values

fig = plt.figure(figsize=(15, 5))
plt.plot( new_prediction, 'b-', label='Prediction')
plt.plot( y_valid, 'g-', label='Data')
plt.legend(loc='upper right')
plt.xlabel('FUGA')
plt.title('Original Dataset vs Churn')
```