

UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN MATEMÁTICA APLICADA

TEMA: MODELO MATEMÁTICO PARA PREDECIR LOS DIVORCIOS EN LA CIUDAD DE LATACUNGA

Trabajo de Titulación previo a la obtención del Grado Académico de
Magíster en Matemática Aplicada
Modalidad de titulación Proyecto de Desarrollo

Autor: Lic. Marco Vinicio Hidalgo Estrella, Mg.

Director: Ing. Marco Enrique Benalcázar Palacios, Ph.D.

Ambato - Ecuador

2022

APROBACIÓN DEL TRABAJO DE TITULACIÓN

A la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial.

El tribunal receptor de la Defensa del Trabajo de Titulación presidido por la Ingeniera Elsa Pilar Urrutia Urrutia Magíster e integrado por los señores: Ingeniero Víctor Santiago Manzano Villafuerte Magíster, Ingeniero Fabián Rodrigo Salazar Escobar PhD, designados por la Unidad Académica de Titulación de Posgrados de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial de la Universidad Técnica de Ambato, para receptor el Trabajo de Titulación con el tema: “Modelo matemático para predecir los divorcios en la ciudad de Latacunga”, elaborado y presentado por el Lic. Marco Vinicio Hidalgo Estrella, para optar por el Grado Académico de Magíster en Matemática Aplicada; una vez escuchada la defensa oral del Trabajo de Titulación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la Universidad Técnica de Ambato.

Ing. Elsa Pilar Urrutia Urrutia, Mg.

Presidente y Miembro del Tribunal de Defensa

Ing. Víctor Santiago Manzano Villafuerte, Mg.

Miembro del Tribunal de Defensa

Ing. Fabián Rodrigo Salazar Escobar, PhD.

Miembro del Tribunal de Defensa

AUTORÍA DEL TRABAJO DE TITULACIÓN

La responsabilidad de las opiniones, comentarios y críticas emitidas en el trabajo de titulación con el tema: “Modelo matemático para predecir los divorcios en la ciudad de Latacunga”, le corresponde exclusivamente a: Lic. Marco Vinicio Hidalgo Estrella Mg, autor bajo la dirección del: Ing. Marco Enrique Benalcázar Palacios PhD, director del trabajo de titulación; y el patrimonio intelectual pertenece a la Universidad Técnica de Ambato.

Lic. Marco Vinicio Hidalgo Estrella, Mg.

AUTOR

Ing. Marco Enrique Benalcázar Palacios, PhD.

DIRECTOR

DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el trabajo de titulación, sirva como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la institución.

Cedo los derechos de mi trabajo de titulación, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad Técnica de Ambato.

Lic. Marco Vinicio Hidalgo Estrella, Mg.

0503210510

ÍNDICE GENERAL

CONTENIDO	Pág.
PORTADA.....	i
APROBACIÓN DEL TRABAJO DE TITULACIÓN.....	ii
AUTORÍA DEL TRABAJO DE TITULACIÓN.....	iii
DERECHOS DEL AUTOR.....	iv
ÍNDICE DE TABLAS.....	viii
ÍNDICE DE FIGURAS.....	ix
AGRADECIMIENTO.....	x
DEDICATORIA.....	xi
RESUMEN EJECUTIVO.....	xii
EXECUTIVE SUMMARY.....	xiv
CAPÍTULO I.....	1
EL PROBLEMA DE INVESTIGACIÓN.....	1
1.1. Introducción.....	1
1.2. Justificación.....	2
1.3. Objetivos.....	3
1.3.1. Objetivo General.....	3
1.3.2. Objetivos Específicos.....	3
CAPÍTULO II.....	4
ANTECEDENTES INVESTIGATIVOS.....	4
2.1. Estado del arte.....	4
2.2. Marco Teórico.....	11
2.2.1. Método de Gottman.....	11
2.2.2. Minería de datos Weka.....	11
2.2.3. Algoritmos de Machine Learning.....	12

2.2.4. Regresión logística binaria.....	13
2.2.5. Estimación y significado de los coeficientes.....	13
2.2.6. Interpretación Odds Ratio.....	14
2.2.7. Validación del modelo.....	15
2.2.8. Matriz de confusión.....	15
2.2.9. Métricas para evaluar el desempeño en la clasificación.....	16
2.2.10. Modelo de clasificación Naive Bayes.....	17
CAPÍTULO III.....	18
MARCO METODOLÓGICO.....	18
3.1. Ubicación.....	18
3.2. Equipos y materiales.....	18
3.3. Tipo de Investigación.....	19
3.4. Prueba de Hipótesis.....	20
3.5. Población o muestra.....	20
3.6. Recolección de la información.....	21
3.7. Procesamiento y análisis estadístico de las variables.....	21
3.8. Análisis descriptivo de datos.....	25
CAPÍTULO IV.....	28
RESULTADOS Y DISCUSIONES.....	28
4.1. Análisis de las características seleccionadas.....	28
4.2. Construcción del modelo de Regresión Logística.....	29
4.3. Verificación de hipótesis.....	30
4.4. Interpretación de los ODDS RATIO (OR).....	30
4.5. Implementación del modelo predictivo Regresión Logística	

para determinar la probabilidad del divorcio.....	32
4.6. Matriz de confusión de Regresión Logística.....	33
4.7. Métricas para evaluar el desempeño de clasificación.....	34
4.8. Implementación del modelo predictivo Naive Bayes para determinar la probabilidad del divorcio.....	35
4.9. Matriz de confusión del modelo Naive Bayes.....	35
4.10. Métricas para evaluar el desempeño de clasificación.....	36
4.11. Análisis de casos Falsos Positivos (FP) y Falsos Negativos (FN).....	38
CAPÍTULO V.....	40
CONCLUSIONES Y RECOMENDACIONES.....	40
5.1. Conclusiones.....	40
5.2. Recomendaciones.....	42
5.3. Bibliografía.....	43
5.4. Anexos.....	46

ÍNDICE DE TABLAS

	Pág.
Tabla 1: Población urbana del cantón Latacunga.....	18
Tabla 2: Primer conjunto de variables de la base de datos.....	22
Tabla 3: Segundo conjunto de variables de la base de datos.....	22
Tabla 4: Tercer conjunto de variables de la base de datos.....	22
Tabla 5: Selección de atributos usando la plataforma Rstudio.....	23
Tabla 6: Características o variables seleccionadas.....	25
Tabla 7: Construcción del modelo de Regresión Logística.....	29
Tabla 8: Interpretación de los ODDS RATIO (OR).....	31
Tabla 9: División del conjunto de datos y testeo para Regresión Logística.....	32
Tabla 10: Valores calculados de la matriz de confusión.....	34
Tabla 11: División del conjunto de datos y testeo para Naive Bayes.....	35
Tabla 12: Valores calculados de la matriz de confusión.....	37
Tabla 13: Resultado de las métricas calculadas de los modelos.....	38

ÍNDICE DE FIGURAS

	Pág.
Figura 1: Plataforma WEKA, selección de atributos usando el evaluador CfsSubsetEval por el método Greedy Stepwise.....	12
Figura 2: Matriz de confusión para una clasificación binaria.....	15
Figura 3: Selección de características más importantes usando el programa WEKA.....	24
Figura 4: Distribución por intervalo de edad en personas divorciadas y no divorciadas.....	25
Figura 5: Resultados obtenidos en las variables independientes encuestadas.....	26
Figura 6: Resultado obtenido en la variable dependiente encuestada.....	27
Figura 7: Matriz de correlación de las variables	28
Figura 8: Matriz de confusión con datos de testeo de Regresión Logística.....	33
Figura 9: Matriz de confusión con datos de testeo de Naive Bayes.....	36
Figura 10: Cuestionario método del Gottman.....	51
Figura 11: Cálculo del Alfa de Cronbach.....	52

AGRADECIMIENTO

En primer lugar, agradecer a Dios sobre todas las cosas, a mis padres por la vida y ser el pilar fundamental para mi desarrollo profesional y humano.

A mi querida esposa Daniela Hidalgo por el gran apoyo recibido durante esta travesía académica y ante todo Dios todo poderoso derrame muchas bendiciones en el trayecto profesional.

Al director de tesis el Dr. Marco Benalcázar Palacios, por compartir experiencias, información y conocimientos en la guía del desarrollo del presente proyecto.

A todos los señores docentes de la maestría mis sinceros agradecimientos en verdad que se cumplió la expectativa.

DEDICATORIA

Dedico este trabajo de manera especial a la gloria de Dios todo poderoso que ha guiado mi camino y que gracias a él llegue a la meta trazada, estoy muy feliz por lo bueno que se ha sido conmigo y bendecirme día a día en todos los ámbitos de mi vida.

A mi familia por el apoyo incondicional recibido por ayudarme alcanzar mis metas y ser mejor profesionalmente cada día.

UNIVERSIDAD TÉCNICA DE AMBATO
FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL
MAESTRÍA EN MATEMÁTICA APLICADA

TEMA:

**“MODELO MATEMÁTICO PARA PREDECIR LOS DIVORCIOS EN LA
CIUDAD DE LATACUNGA”**

AUTOR: Lic. Marco Vinicio Hidalgo Estrella, Mg.

DIRECTOR: Ing. Marco Enrique Benalcázar Palacios, Ph.D.

LÍNEA DE INVESTIGACIÓN:

- Tecnología de la información y sistemas de control

FECHA: 04 de agosto 2022

RESUMEN EJECUTIVO

En la actualidad la predicción temprana de un caso de divorcio es fundamental para salvar el matrimonio. La muestra seleccionada para el presente trabajo fue de 381 personas entre parejas casadas y divorciadas de las parroquias urbanas de la ciudad de Latacunga. Los datos se obtuvieron por medio de la encuesta, a través del instrumento cuestionario que emplea el método del profesor Gottman. Para determinar las variables o características más importantes que pueden llevar a un divorcio se utilizó el software libre Rstudio. En un primer barrido mediante el método de correlación de variables se obtuvieron 15 variables, luego en un segundo barrido usando el software WEKA resultó 5 variables o características. Las variables independientes son: Atr2(tratar de ignorar las diferencias aún en situaciones difíciles), Atr3(retomar las discusiones desde el principio y corregirlas), Atr10 (la mayoría de nuestros objetivos son poco comunes) Atr35(insulto a mi cónyuge en una discusión), Atr43 (me quedo en silencio para calmar el ambiente en una discusión), y como variable dependiente el divorcio. Se emplearon dos algoritmos de clasificación supervisados: Regresión Logística y Naive Bayes, ejecutando las predicciones

con el 70% de datos de entrenamiento y 30% datos para testeo o prueba. Para evaluar el desempeño en la clasificación se usaron las métricas precisión y F1_Score. Las pruebas determinaron que el valor de la puntuación F1-Score del modelo de Regresión Logística es de (98%) con una precisión del (98%) en cambio el valor F1-Score del modelo Naive Bayes es de (96%) y su precisión de (96%). Concluyendo que el modelo de Regresión Logística es el mejor modelo para predecir los divorcios en la ciudad de Latacunga.

Descriptor: Divorcio, cuestionario de Gottman, Weka, Regresión Logística, Naive Bayes, matriz de confusión, valor, FI-Score y precisión.

UNIVERSIDAD TÉCNICA DE AMBATO
FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL
MAESTRÍA EN MATEMÁTICA APLICADA

THEME:

“MATHEMATICAL MODEL TO PREDICT DIVORCES IN THE CITY OF LATACUNGA”

AUTHOR: Lic. Marco Vinicio Hidalgo Estrella, Mg.

DIRECTED BY: Ing. Marco Enrique Benalcázar Palacios, Ph.D.

RESEARCH LINE:

- Information Technology and control systems

DATE: August 04 th, 2022

EXECUTIVE SUMMARY

Currently, the early prediction of a divorce case is essential to save the marriage. The sample selected for this study was 381 people between married and divorced couples from urban parishes in the city of Latacunga. The data was obtained through the survey, through the questionnaire instrument that uses Professor Gottman's method. To determine the most important variables or characteristics that can lead to a divorce, the free software Rstudio was used. In a first sweep using the variable correlation method, 15 variables were obtained, then in a second sweep using the WEKA software, 5 variables or characteristics were obtained. The independent variables are: Atr2(Try to ignore differences even in difficult situations), Atr3(Restart arguments from the beginning and correct them), Atr10 (Most of our goals are unusual) Atr35(I insult my spouse in an argument), Atr43 (I remain silent to calm the atmosphere in an argument), and divorce was the dependent variable. Two supervised classification algorithms were used: Logistic Regression and Naive Bayes, executing the predictions with 70% training data and 30% testing data. To evaluate the performance in the classification, the precision and F1_Score

metrics were used. The tests determined that the value of the F1-Score of the Logistic Regression model is (98%) with an accuracy of (98%), while the F1-Score value of the Naive Bayes model is (96%) and its accuracy of (96%). Concluding that the Logistic Regression model is the best model to predict divorces in the city of Latacunga.

Keywords: Divorce, Gottman questionnaire, Weka, Logistic Regression, Naive Bayes, confusion matrix, value, FI-Score and precision.

CAPÍTULO I

EL PROBLEMA DE INVESTIGACIÓN

1.1. Introducción

En el presente trabajo investigativo se pretende a través de un modelado matemático se pueda predecir un divorcio. Para lo cual se determinó las variables más importantes que dan origen al mismo empleando algoritmos de aprendizaje existentes. Se aplicó la encuesta debidamente fundamentada basada en el Método de Gottman de terapia de pareja que servirá para hacer las predicciones. Se analizaron las variables que de una u otra manera influyen en la disolución conyugal. Mediante un proceso de selección estadística y la puesta en práctica de algunos algoritmos como es Regresión Logística y Naive Bayes, mediante su entrenamiento y testeo se predecirá si el divorcio puede darse o no[1].

El modelo construido debe tener la capacidad de predecir estos escenarios. Por medio de los resultados generados se espera que se logre mitigar el impacto de un posible divorcio en la sociedad, de manera especial para aquellas familias que miran al divorcio como algo negativo. En mira de la revisión literaria se deriva que este tema ha sido enfocado desde un punto de vista más social. Existen ciertos trabajos enfocados al campo de la matemática computacional empleando diferentes técnicas de inteligencia artificial. En la parte de atributos es importante desarrollar un buen proceso de selección para reducir en sí la dimensionalidad del conjunto. Se usaron algunas técnicas para ello, luego se emplea la metodología asociada para abordar el problema[2].

Un ejemplo de estudios realizados es “Un modelo de la probabilidad de divorcio” la razón fundamental de este estudio fue investigar el divorcio que está relativamente poco investigado. La Regresión Logística fue utilizada para predecir el divorcio. Se obtuvo como resultado del 95% de clasificación correcta con el modelo empleado[3]. En los resultados son seis las variables principales que dan lugar al divorcio. Se determinó la precisión del modelo usando las

métricas de evaluación de desempeño como son: exactitud, sensibilidad, F1-Score

1.2. Justificación

En Ecuador en los últimos 10 años la tasa de divorcios se ha incrementado notablemente, el tema a investigarse tiene un gran impacto, tal como se ha podido determinar cada vez es más frecuente el número de parejas que se divorcian, esta realidad no es ajena a la ciudad de Latacunga. El predecir tempranamente un determinado caso de divorcio ayudaría a salvar un matrimonio sin que se genere grave conmoción dentro de su núcleo familiar, sobre todo si existen hijos menores de edad. Para lo cual el presente estudio establece un modelo matemático de predicción en casos de divorcio usando algoritmos de aprendizaje como es Regresión Logística y Naive Bayes. Mediante un análisis estadístico se pretende conocer las principales variables por las cuáles las parejas se han separado o divorciado, no alcanzando una durabilidad esperada por la pareja.[4]

Es conveniente llevar esta investigación usando modelos de aprendizaje supervisado con datos para entrenar y prueba para verificar su precisión. Lo cual sería muy útil para los terapeutas que realizan las terapias de parejas y ayudar a mitigar el impacto de una separación en época temprana tomado mejores decisiones. Los beneficiarios directos son las familias y la sociedad en general, los cónyuges identificarán los aspectos más relevantes en los cuales se necesita centrar para prevenir un divorcio prematuro. El tema es trascendental puesto que se trabajará con una metodología centrada en un análisis estadístico y modelos de aprendizaje utilizando muchos descriptores de entrada para la predicción acertada del divorcio.

Es un proyecto relevante porque en la ciudad de Latacunga no existen trabajos que han empleado esta investigación usando un modelado matemático con algoritmos de aprendizaje, que a la final va ayudar a las familias y sus integrantes a preservar el matrimonio, e identificar aspectos que están relativamente equívocos. Los cónyuges pueden identificar factores negativos que han afectado su relación de pareja evitando una posible separación que pueda dañar a la base de toda sociedad que es la familia. Los resultados que

se determinen en esta investigación servirán para recopilar información para futuros proyectos en beneficio de la sociedad, además los resultados se lo difundirán en una revista buena.

1.3. Objetivos

1.3.1. Objetivo General

Desarrollar un modelo matemático para predecir los divorcios en la ciudad de Latacunga mediante aprendizaje supervisado.

1.3.2. Objetivos Específicos

- Aplicar una encuesta usando el Método de Gottman dirigido a personas divorciadas y en matrimonio en la ciudad de Latacunga.
- Estimar las relaciones que existen entre las variables que dan lugar a predecir el divorcio.
- Desarrollar un modelo matemático con datos disponibles y técnicas de aprendizaje supervisado que se enfoquen en predecir el divorcio.

CAPÍTULO II

ANTECEDENTES INVESTIGATIVOS

2.1. Estado del Arte

En el presente capítulo se va a incluir varias investigaciones realizadas que ayudaron en la construcción de la bibliografía para la realización del presente trabajo de investigación.

Sharma [1] en el trabajo titulado “Predicción de casos de divorcio utilizando algoritmos de aprendizaje automático” los autores realizan la siguiente investigación sobre la predicción del divorcio empleando los algoritmos de aprendizaje automático de clasificación existentes: Perceptrón, Árbol de decisión, Random Forest, Naive Bayes, K-Nearest Neighbor y el clasificador Support Vector Machine. Determinaron cual tiene la mejor precisión al realizar una comparación basándose en el método de Gottman (Cuestionario para realizar las predicciones del divorcio) que cuenta con 54 variables independientes y una sola dependiente. Los autores han logrado una mejor precisión con el modelo de aprendizaje del Perceptrón con una confiabilidad del 98,5%.

Sohail [2] en su obra titulada “Implementación de algoritmo de aprendizaje automático sobre factores que afectan la tasa de divorcio ” utilizan un conjunto de datos que contienen algunos de los principales atributos o características que causan el divorcio, emplean tres algoritmos de aprendizaje automático pero de clasificación diferente: Árbol de decisión, k-vecino más cercano y Naive Bayes. Los investigadores concluyeron que el mejor algoritmo para predecir el divorcio es el clasificador Naive Bayes logrando una precisión del modelo de 92.5%, en cambio KNN es mejor cuando aumentan los datos de entrenamiento.

Para Swenson [3] en su trabajo titulado “Un modelo logit de la probabilidad de divorcio” aplicaron el modelo de regresión logística para predecir la probabilidad del divorcio de los habitantes de Alberta, tomando en cuenta los factores estructurales, procesionales sociales y psicológicos sociales para predecir el divorcio. Los investigadores llegaron a la conclusión que al aplicar la regresión logística son importantes los factores estructurales (ubicación

geográfica, religión), procesionales sociales (participación de la mujer en el trabajo) y psicológicos sociales (actitudes ante el divorcio) se relacionan y desembocan en la probabilidad de divorcio en las parejas.

Castelo, Carrillo & Dávalos [4] en el trabajo de titulación "Factores que inciden en los divorcios prematuros en el Ecuador: un modelo de regresión logística" aplican un modelo de regresión logística que permite predecir si un matrimonio tendrá una duración de por lo menos una década. Mediante la herramienta Rstudio se realizó un análisis de tipo exploratorio y descriptivo con datos del INEC. Los investigadores concluyeron que existen provincias con un alto grado de divorcios las causas son: nivel educativo, número de hijos, algún cónyuge sea ebrio, edad, el modelo obtuvo una capacidad de predicción del 88% se utilizó la curva ROC, como métrica de evaluación.

Para Castrillón [5] en su trabajo de investigación "Predicción del divorcio por medio de técnicas inteligentes" aplican una plataforma de aprendizaje automático y minería de datos Weka, usando el algoritmo de clasificación J48 que es parecido a un algoritmo bayesiano. Se modeló con una base existente de 173 registros existentes y 55 variables siendo 54 independientes y una la dependiente que viene a ser el divorcio. El investigador seleccionó los atributos o variables independientes que tienen influencia moderada sobre la variable dependiente, luego se aplicó una segunda reducción de las variables independientes mediante el proceso estadístico. El investigador estableció cuatro variables más determinantes con una partición mayor al 90%, usando un árbol decisión se mostró los atributos más importantes que pueden ocasionar un divorcio.

Goel,Roshan,Tyagi & Agarwal [6] en su obra titulada "Augur Justice: una técnica de aprendizaje automático supervisado para predecir los resultados de los casos judiciales de divorcio" realizan un análisis comparativo de varios algoritmos de aprendizaje supervisado en el campo de la ley, en función de las características presentes en el conjunto de datos entrenados de los casos judiciales. Los investigadores concluyeron que la capacidad de ciertos algoritmos propuestos para predecir el divorcio era mejor que las técnicas de aprendizaje automático supervisado que se usa regularmente.

En la publicación de Ashrafujjaman [7] titulada “Predicción de divorcio utilizando métodos de aprendizaje automático: perspectiva de Bangladesh ” emplean la Escala de Predicción de Divorcio (DPS) basada en la terapia de pareja de Gottman. Este cuestionario está estructurado por 54 elementos considerados como características o atributos en un modelo de aprendizaje automático. Son 201 individuos encuestados distribuidos de la siguiente forma: 127 (63%) estaban casados y 74 (37%) divorciados. Se usaron los algoritmos existentes como son: Multilayer Perceptrón, Naive Bayes y Random Forest. Los autores realizaron la selección de características o elementos mediante la correlación de variables se determinaron 5 características que son más influyentes en la predicción del divorcio. La tasa más alta fue de 87,14% con el algoritmo Naive Bayes aplicado al conjunto de datos existentes. Luego con las características seleccionadas la predicción más alta fue con el Perceptrón a lo que los investigadores confirman que del DPS puede predecir el divorcio.

Ranjitha [8] en el documento de investigación titulado “Predicción de divorcio mejorada mediante aprendizaje automático: optimización de enjambre de partículas (PSO)” utilizan el algoritmo de aprendizaje denominado (PSO) el mismo que es optimizado para el conjunto de datos disponibles para pronosticar la tasa de divorcio. Los investigadores obtuvieron una precisión de 99,66% con una tasa de error obtenida de 0.0033, con respecto a los datos de entrenamiento tienen 136 instancias y los datos de testeo son de 34 instancias. Los autores manifiestan que estas predicciones podrían orientar a profesionales que de forma particular trabajan en consejería matrimonial y puedan ayudar a las parejas a resolver sus conflictos.

Juárez, Hernández & Chávez [9] en su documento que se titula “Comparación del algoritmos de clasificación utilizando la selección de características ” utilizan algunos algoritmos de clasificación existentes: C4.5, JRip, K-NN, SVM y regresión logística. Las prácticas se realizaron con un conjunto de datos de divorcio existentes que consta de 54 características o atributos con una variable de salida dicotómica, se aplican a 170 individuos y se realiza la selección de los atributos más importantes empleando el evaluador CFS de la plataforma Weka obteniendo al final 4 características. Los investigadores

concluyeron que para obtener el modelo más adecuado es recomendable emplear las características más importantes en lugar de usar todo el conjunto.

En el trabajo de Guillén [10] en su obra titulada “Predicción de disoluciones maritales utilizando redes neuronales de función de base radial”, emplean diversos algoritmos de aprendizaje automático que permitan determinar si un matrimonio que presenta un fase inicial de disolución conyugal temprana se finalizará en un acuerdo amistoso y pacífico, o seguirá el camino legal es decir será llevado a los tribunales. En el presente trabajo se ha utilizado también redes neuronales. Los investigadores llegan a determinar que dicha información es válida para realizar la predicción del divorcio empleando los datos disponibles, además la metodología es capaz de clasificar con una buena precisión una base de datos real aplicada a determinada muestra.

Wang & Chen [11] en su trabajo titulado “Sistema de diálogo automático de derecho matrimonial basado en el árbol de decisión paralelo C4.5” aplican un modelo de predicción de clasificación de consulta legal basado en los algoritmos de clasificación: árbol de decisión paralelo C4.5, algoritmo SVM y Naive Bayes empleando datos disponibles. Los resultados obtenidos son evaluados mediante métricas: precisión, Recall y F1-Score. Los investigadores concluyeron que el árbol de decisión y el modelo de Naive Bayes genera buenos resultados en vista de las métricas de evaluación dando como resultado una mejor consulta legal al usuario.

En la publicación de Flores & Silva [12] titulada “Modelo de Machine Learning para predecir el divorcio de una pareja casada” se basan en un trabajo de Yontem que emplea un cuestionario de 56 preguntas como una variable de salida dicotómica, que ayudan a predecir el divorcio usando 4 modelos de aprendizaje automático: perceptrón, regresión logística, redes neuronales y bosque aleatorio. Los investigadores entrenaron estos modelos de aprendizaje en 5 escenarios aplicando las métricas de evaluación obtuvieron el mejor resultado: precisión, sensibilidad y especificidad es 0,9853, 1,0 y 0,9667 respectivamente dichos valores corresponden al modelo regresión logística y un modelo híbrido. Los autores manifiestan que es muy importante considerar el contexto de la muestra, la cantidad de datos y el país donde se recopilan los datos para su investigación.

Shankhdhar, Tushar & Vardahan [13] en su trabajo titulado “Escala de predicción de divorcio utilizando técnicas de aprendizaje automático improvisadas” utilizan como punto de partida un método de selección de características más importantes aplicando la correlación de variables en un determinado conjunto de datos disponibles. Se realiza la división del conjunto de datos: entrenamiento y testeo, este último sirve para entrenar el modelo que ayuda a predecir un divorcio futuro. Los investigadores concluyeron que al emplear varios algoritmos de aprendizaje automático supervisado y realizando un estudio comparativo de distintos modelos se obtiene buenos resultados en cuanto a su precisión.

De igual forma Mustafa [14] en su documento titulado “Predicción de divorcio utilizando selección de características basada en correlación y redes neuronales artificiales ” aplican una escala de predictores de divorcio (DPS) sobre la base de la terapia de del profesor Gottman, los datos recolectados son: 84 están divorciados y 86 eran parejas casadas. Para determinar las características o atributos más significativos emplean el método de correlación de Pearson. Los algoritmos usados son: árbol de decisión, Perceptrón Neural y C4.5. Los investigadores mediante minería de datos concluyeron que son 5 características las más importantes que pueden conducir al divorcio, la mayor tasa de precisión que se obtuvo fue de 98,82% con ANA.

Kong & Chai [15] en su obra titulada “¿Es confiable su matrimonio?: análisis de divorcio con algoritmos de aprendizaje automático” para el estudio aplican cuatro algoritmos de aprendizaje automático: Support Vector Machine, Random forest, NGBoost y Regresión Logística con una muestra de 170 datos disponibles empleando un cuestionario de Gottman que contiene 54 preguntas sobre el estado emocional de la pareja. Los investigadores concluyeron que en comparación con SVM y RF, NGBoost y regresión logística tiene un buen rendimiento con una precisión de 0,9828 y 0,9769. Se determinó también que son 5 los factores más importantes que pueden conducir al divorcio en una pareja.

Kong, Kurz & Heinrich [16] en su trabajo titulado "Predicción de divorcios utilizando métodos de aprendizaje automático" aplican una encuesta sugerida por profesor Gottman que está estructurada con 54 preguntas. Para predecir si una pareja actualmente está casada o divorciada usan modelos de aprendizaje estadístico supervisado: K-vecinos más cercanos, análisis discriminante lineal, Regresión Logística y Naive Bayes, además emplean diferentes técnicas y métodos estadísticos que ayudan a determinar el mejor modelo. Los investigadores concluyeron que el modelo de regresión logística obtiene buenos resultados usando las variables o atributos (preguntas de encuesta) 17,18,19,26 y 40 muestran un error de prueba del 0% para que no exista un sobreajuste y una precisión de 0,95.

Joel [17] en su trabajo titulado "El aprendizaje automático descubre los predictores de autoinforme más sólidos de la calidad de la relación en 43 estudios longitudinales de parejas " utilizan dos modelos de aprendizaje automático :Random Forest y Regresión logística, para comprobar si la buena relación entre pareja es predecible se emplean predictores de diferencias individuales : la felicidad, afecto negativo, depresión, la evitación y la ansiedad por el apego. Los investigadores concluyeron que las variables de la relación y las diferencias individuales de las parejas predijeron de dos a cuatro veces más varianza que las variables informadas por la pareja. Esto quiere decir que las diferencias individuales y las experiencias que tiene la pareja tienden a ser determinante en la calidad de la relación.

Heymany & Smith [18] en su trabajo titulado "Los peligros de predecir el divorcio sin validación cruzada " en los estudios que realizan para poder predecir el divorcio los autores sugieren ante un posible divorcio de las parejas es importante tener presente la precisión del modelo. A partir de la selección de variables mediante un proceso estadístico usando el modelo de regresión logística clasificó correctamente al 90% de las parejas, con una sensibilidad del 92% y una especificidad del 89%. Los investigadores concluyeron que los resultados de los estudios que se realizan con modelos de aprendizaje supervisado y no utilicen la validación cruzada se deben interpretar con mucho cuidado, porque posiblemente pueden ser resultados muy buenos pero que en realidad sean malos.

Mohamed & Alkhyeli [19] en su trabajo titulado “Predicción temprana de divorcio en Abu Dhabi ” utilizan datos administrativos disponibles, para predecir el divorcio emplean minería de datos usando la plataforma Weka, mediante técnicas de clasificación como son: regresión logística y árboles de decisión. La conclusión que llegaron los autores con los dos métodos de interpretación es su gran importancia para determinar la variable que tiene más influencia para un determinado cambio de predictores. Los árboles de decisión son empleados para observar cómo los datos se dividen en hojas y nodos, al mismo tiempo como estas variables o características seleccionadas se interrelacionan. El segundo modelo de clasificación que es la regresión logística prueba el modelo de forma estadística mediante la prueba de hipótesis y valor p.

Alkhayyat & Alhaqbani [20] en su trabajo titulado “Una comparación entre la regresión logística binaria y análisis discriminante en la determinación de los factores más importantes que afectan la inestabilidad de la vida marital en El Reino de Arabia Saudita”, comparan dos modelos estadísticos: modelo de regresión logística binaria y el modelo de análisis discriminante usando 38 atributos para determinar los factores más importantes que originan la desequilibrio conyugal afectando la prolongación de la vida matrimonial de las parejas. Los investigadores concluyeron que ambos modelos son buenos en su predicción, los trastornos psicológicos y la duración del matrimonio son los factores más importantes en los modelos: la regresión logística binaria con un valor 92,4% y el análisis de la discriminante 91,3%.

2.2. MARCO TEÓRICO

2.2.1. Método de Gottman

La predicción del divorcio se realiza empleando la Escala de Predictores de Divorcio (DPS) tomando en cuenta la terapia de pareja de Gottman. Es un modelo que explica las posibles causas del divorcio que afronta una pareja basándose en investigaciones empíricas realizadas por el autor. Los principales predictores del divorcio en forma general son los cuatro jinetes del apocalipsis. El profesor Gottman los describe como: “crítica, el desprecio, obstruccionismo y actitud defensiva” presentes en diferentes preguntas[14].

El cuestionario está conformado de 54 atributos las mismas que están de acuerdo con la teoría de parejas de Gottman. Las preguntas son un conjunto de características o variables independientes y la variable dependiente.

2.2.2. Minería de datos Weka

El software WEKA es una poderosa herramienta para el desarrollo del aprendizaje automático, se emplea en Machine Learning (Witten et al. 2016). Este software ayuda al análisis y extracción de datos de manera rápida, segura y confiable. Esta plataforma está conformada de técnicas para: diferentes regresiones, algoritmos de clasificación, agrupamiento de datos y minería de datos que permite la selección de atributos relevantes. En este estudio se elige este software WEKA para la selección de los atributos más importantes y crear el modelo seleccionado. Tiene muchas características integradas con una buena precisión, un punto muy importante es que no requieren de programación, es muy utilizada para la investigación y enseñanza [21].

Selección de atributos

El objetivo de esta herramienta es identificar, mediante un conjunto de datos formado por muchos atributos o variables aquellas que tienen mayor peso o son más importantes para determinar a qué tipo de clase pertenecen [22]. En la siguiente figura para seleccionar los atributos más importantes, el evaluador de atributos es el CfsSubsetEval y el método seleccionado es GreedyStepwise.

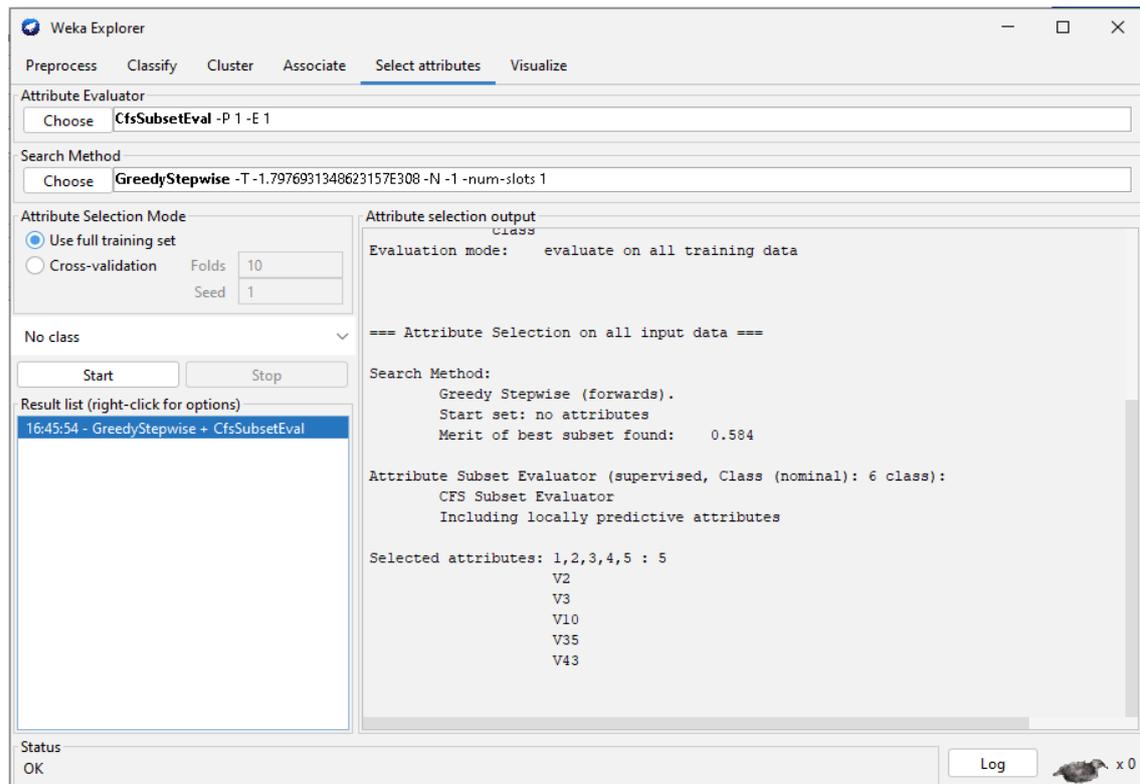


Figura 1: Plataforma WEKA, selección de atributos usando el evaluador CfsSubsetEval por el método Greedy Stepwise.
Elaborado por: Hidalgo, 2022

2.2.3. Algoritmos de Machine Learning

Los algoritmos más importantes y conocidos de Machine Learning se pueden utilizar bien en modelos de aprendizaje supervisado o en modelos de aprendizaje no supervisado. Los primeros utilizan datos conocidos es decir en datos existentes o disponibles. Los datos de entrada se asocian con datos de salida para poder entrenar el sistema. El principal objetivo de estos algoritmos es aprender y ser útiles para realizar predicciones que se ajustan a un conjunto de datos para su estudio[23].

Modelo de Regresión Logística

La regresión logística es empleada principalmente para resolver problemas de clasificación supervisada, además se caracteriza por ser explicativa como predictiva puede ser de tipo binomial y multinomial [24]. La diferencia principal entre la regresión logística multinomial con respecto a la regresión logística binomial es que la variable dependiente no es dicotómica, esta puede tener más de 2 categorías [25].

La regresión logística es conocida como una técnica de análisis multivariante donde la variable dependiente es una variable de tipo dicotómica y las variables o variables independientes pueden ser cualitativas o cuantitativas, la variable dicotómica tiene dos valores y es la más utilizada [26].

2.2.4. Regresión logística binaria

Entre los modelos de elección binaria, que sirven para explicar una variable dependiente binaria (0/1), se encuentra la regresión logística. Se utiliza para predecir la probabilidad estimada $P(Y)$ de que la variable (Y) presente uno de los dos valores posibles ($1 = \text{si}$ $0 = \text{no}$), de acuerdo a la diversidad de valores que adopten el conjunto de variables independientes X_i [27]. La función logística está dada por la siguiente ecuación:

$$P(Y = 1) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}} \quad (1)$$

O bien:

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (2)$$

Donde:

P : Representa la probabilidad de y

Y : Es la variable dependiente

β_0 : Es la ordenada o intercepto en la función

X : Son las variables

β_i : Son los coeficientes de regresión asociado a cada variable independiente X_i

2.2.5. Estimación y significado de los coeficientes

Los diferentes estimadores de los coeficientes para el modelo se calculan mediante el método de la función de máxima verosimilitud, es una compleja técnica de cálculo diferencial, pero existen ordenadores que utilizan programas informáticos que ayudan a resolver este problema. Si los coeficientes de las variables son positivos quiere decir que la variable aumenta la probabilidad del suceso, y al contrario si el coeficiente es negativo disminuye la probabilidad del

suceso, y para establecer qué variables son útiles para explicar la posible probabilidad del suceso, se emplea el contraste de la siguiente hipótesis[26].

Ho: $\beta_i = 0$ (los coeficientes de β_i son igual a cero)

Ha: $\beta_i \neq 0$ (Al menos un coeficiente de β_i es diferente de cero)

En el modelo de Regresión Logística si el valor de p es inferior a 0,05, se procede a rechazar la Ho. Caso contrario, se acepta.

2.2.6. Interpretación Odds Ratio

Para poder interpretar el resultado de los odds ratio del modelo de regresión logística se debe recurrir al concepto de "Odds ratio", que es una de las medidas de asociación entre variables. De esta forma se define como el coeficiente de la probabilidad de presentar una característica y la probabilidad de no presentarla descrita en las siguientes ecuaciones[28].

$$Odds = \frac{p}{(1 - p)} \quad (3)$$

$$\frac{p}{1-p} = \exp(\alpha_0 + \alpha_1 X) \quad (4)$$

$$\frac{p}{1-p} = \exp(\alpha_0) * \exp(\alpha_1 X) \quad (5)$$

Por lo que $\exp(\alpha_1)$ representa el OR es decir el (Odds Ratio) por unidad de incremento de la variable explicativa X.

$$\frac{odds_{x_j+1}}{odds} = \exp(\theta_j (x_j + 1) - \theta_j x_j) = \exp(\theta_j) \quad (6)$$

El exp () hace referencia a un cambio en x_j en determinada unidad incrementa el logaritmo de la razón de probabilidad al multiplicar por un cierto peso[29]. El registro de probabilidades generalmente va desde el $-\infty$ a $+\infty$.

Los coeficientes del modelo se calculan empleado el logaritmo neperiano (ln) en base a los Odds ratio de cada variable independiente (X). De forma viceversa: OR = antilogaritmo (β_i) ; por lo tanto, e (β_i). Es decir, se emplea el exponencial para hallar los coeficientes.

2.2.7. Validación del modelo

Realizando pruebas en modelos de aprendizaje se comprueba la capacidad prediciendo nuevos casos que no se utilizaron para entrenarlo, de esta forma se verifica si el modelo se generaliza. La idea fundamental consiste en dividir el conjunto de datos de entrenamiento en dos partes [30].

1.- Training o entrenamiento (70% de la muestra): sirve para entrenar al modelo de clasificación.

2.- Test (30%): Sirve para validar el modelo. La característica fundamental es que esta muestra no debe haber tenido contacto previamente con el funcionamiento del modelo.

2.2.8. Matriz de confusión

Es una herramienta muy utilizada en Machine Learning para evaluar el rendimiento de los diferentes modelos. La matriz de confusión es una tabla compuesta de filas y columnas que representa el número de aciertos y fallas. La matriz ayuda a entender mejor cómo se comporta el modelo, es muy útil para mostrar de forma clara si una clase es confundida con otra[31].

Una matriz de confusión para un clasificador binario está estructurada de la siguiente forma como se presenta en la siguiente figura.

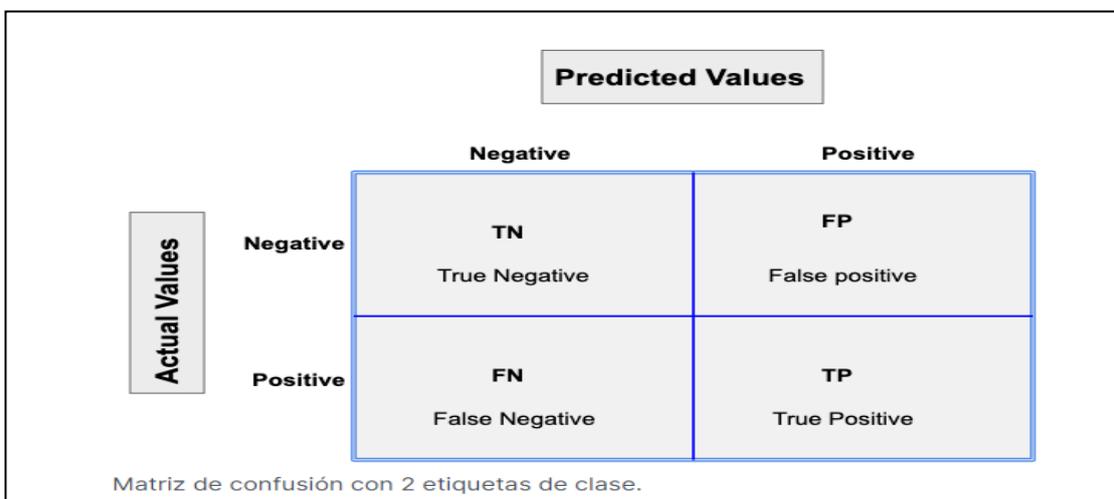


Figura 2: Matriz de confusión para una clasificación binaria.
Elaborado por: Hidalgo, 2022

Donde:

TN: Es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo seleccionado.

TP: Es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo seleccionado.

FN: Es la cantidad de positivos que fueron clasificados incorrectamente como negativos.

FP: Es la cantidad de negativos que fueron clasificados incorrectamente como positivos.

2.2.9. Métricas para evaluar el desempeño en la clasificación

Las métricas de evaluación que son las más usadas para medir la efectividad para un clasificador binario son las siguientes[32].

Exactitud

Llamado también como Accuracy o precisión general es la proporción de predicciones que son correctas. Se calcula sumando los valores de la diagonal y dividido por el total de casos [31]:

$$Accuracy = \frac{(VP + VN)}{(VP + FN + VN + FP)} \quad (6)$$

Precisión

Corresponde al porcentaje de ejemplos que son etiquetados correctamente como positivos, respecto a todas las muestras que son etiquetadas como tal.

$$Precisión = \frac{VP}{VP + FP} \quad (7)$$

Sensibilidad

Conocida en el idioma inglés como Recall es el porcentaje de ejemplos positivos que son correctamente clasificados.

$$Sensibilidad (Recall) = \frac{VP}{VP + FN} \quad (8)$$

Puntuación F1-Score

Se considera como una medida de la precisión y robustez del modelo. Combina la precisión y la sensibilidad, se usa para datos desbalanceados.

$$F1_{score} = \frac{2 * (precisión * recall)}{precisión + recall} = \frac{2VP}{2VP + VP + FN} \quad (18)$$

2.2.10. Modelo de clasificación Naive Bayes

Es uno de los algoritmos de aprendizaje automático más simples y utilizados para realizar predicciones. Dada su facilidad de utilización se ha convertido en una especie de estándar con la cual se compara las bondades de los diferentes métodos. Su denominación proviene de la hipótesis ingenua sobre la que se construye, las variables predictoras son de forma condicional independientes dada la variable a clasificar en la etiqueta del modelo[33].

Es un buen clasificador que emplea la probabilidad para hacer las predicciones. Se usa el Teorema de Bayes para calcular la probabilidad posterior de cada clase. Thomas Bayes en 1763 expresa la probabilidad condicional de un evento aleatorio A dado, se expresa en la siguiente formula [34]:

$$p(A|B) = \frac{P(A) * P(B|A)}{P(B)} \quad (9)$$

Donde:

$p(A)$ representa la probabilidad a priori de que el suceso A sea verdadero.

$p(A|B)$ es la probabilidad de que el suceso A sea verdadero tras considerar B.

$P(B|A)$ indica la probabilidad de que el suceso B sea verdadero, teniendo presente que A lo es.

$P(B)$ representa la probabilidad a priori que el suceso B sea verdadero. Actúa de coeficiente en la fracción.

Entonces un clasificador Naive Bayes o conocido también con el nombre de Bayesiano ingenuo es un clasificador de tipo probabilístico basado en el teorema de Bayes y ciertas hipótesis simplificadoras adicionales.

CAPÍTULO III

MARCO METODOLÓGICO

3.1. Ubicación

El cantón de Latacunga es conocido también como San Vicente Mártir, es la capital de la Provincia de Cotopaxi. Actualmente es la urbe más poblada y grande de la provincia. Se encuentra ubicada en el centro-norte de Zona interandina de Ecuador con una altitud de 1750 msnm con un clima frío que bordea en promedio los 12°C[35].

En el año 2010 el Instituto Nacional de Estadística y Censos la ciudad de Latacunga tenía una población de 174.235 habitantes de los cuales 63842 habitantes de la zona urbana son considerados para el estudio. Tiene 5 parroquias urbanas: La Matriz, Eloy Alfaro (San Felipe), Ignacio Flores (La Laguna), Juan Montalvo (San Sebastián) y San Buenaventura[35], como se presenta en la siguiente tabla.

Tabla 1: Población urbana del cantón Latacunga.

Parroquias Urbanas	habitantes
La Matriz	22565
Eloy Alfaro	13526
Ignacio Flores	15325
Juan Montalvo	10552
San Buenaventura	1874

Elaborado por: Hidalgo, 2022

Fuente: Censo de Población y Vivienda 2010, INEC

3.2. Equipos y materiales

Entre las herramientas y programas utilizados en el presente trabajo son:

En Google Forms se genera la encuesta para obtener las repuestas que dan lugar a la construcción de la data.

Mediante el uso de Excel se realiza la codificación de resultados obtenidos para posteriormente realizar la tabulación.

Para la selección de las características más importantes del modelo se emplea el software WEKA.

Se emplea el software Rstudio en la versión más actual para realizar el análisis estadístico de las variables. Se realiza la construcción del modelo de Regresión Logística y el modelo de Naive Bayes para realizar las predicciones del divorcio. Se compara la precisión y la puntuación F1-Score de los dos modelos para determinar cuál es mejor prediciendo.

Los equipos empleados son:

Laptop Hp Core i3-1005G1 de décima generación con un almacenamiento de 1TB HDD. 4GB RAM.

Teléfono celular Samsung Galaxy A30 de 3 GB de memoria RAM y 32GB de almacenamiento interno.

Impresora Epson L355

3.3. Tipo de Investigación

El presente trabajo de investigación se realiza con un análisis de tipo exploratorio, descriptivo, correlacional.

Investigación exploratoria

Este tipo de investigación exploratoria se emplea normalmente cuando se pretende analizar una tema o problema de investigación a desarrollarse y ha sido escasamente estudiado o abordado, en el abordaje de la revisión literaria existe poco material relacionado al tema a investigarse en la localidad[36].

La investigación en el presente trabajo es de tipo exploratoria porque existen pocos estudios acerca de la predicción de divorcios en la ciudad de Latacunga además servirá para estudios a futuro relacionado al mismo tema estudiado.

Investigación descriptiva

Este tipo de investigación descriptiva tiene como finalidad explicar las propiedades o características y perfiles de las personas, grupos, comunidades,

o cualquier fenómeno que se necesite analizar. Pretende medir o recolectar la información de forma independiente o conjunta de las variables propuestas en el tema de estudio[36].

La investigación realizada es de tipo descriptiva porque ayudó a conocer las variables o atributos de las personas casadas y divorciadas para luego someter a un modelo de clasificación y predecir.

Investigación Correlacional

La investigación de tipo correlacional pretende evaluar el grado de asociación entre dos o más variables, es decir busca establecer la relación que exista generalmente entre dos o más atributos en una determinada muestra[36].

La investigación correlacional ayudó a buscar las relaciones existentes entre las variables o atributos de estudio que dan origen al divorcio. Definida la correlación de variables se realiza la modelación matemática con los dos modelos para predecir los divorcios en la ciudad de Latacunga.

3.4. Prueba de Hipótesis

Hipótesis nula: El modelo construido no es significativo.

Hipótesis alternativa: El modelo construido si es significativo.

3.5. Población o muestra

La población de estudio es la ciudad de Latacunga que está conformada por 63842 habitantes de la localidad mayores de edad que responderán las preguntas del cuestionario.

La muestra es tomada de las 5 parroquias urbanas de la ciudad de Latacunga. La población seleccionada para la presente investigación son parejas casadas y parejas divorciadas.

Para el cálculo del tamaño de la muestra se desarrolló siguiente fórmula estadística.

$$n = \frac{Z^2 N * p * q}{[E^2 - (N - 1)] + [Z^2 * p * q]}$$

Donde:

n = Tamaño de la muestra.

N = Tamaño de la población.

Z = Nivel de confiabilidad del 95%.

p = Probabilidad de éxito 0,5%

q = Probabilidad de fracaso 0,5%

E = Error de estimación al 5%.

En el último censo efectuado a la población en el año 2010 la ciudad de Latacunga tiene un total de 63842 habitantes en la zona urbana cuyo grupo de edad es de los 15 años en adelante. Se presenta el Desarrollo de la ecuación:

$$n = \frac{(1.96)^2 * (63842) * 0.5 * 0.5}{[0.05^2 (63842 - 1)] + [1.96^2 * 0.5 * 0.5]}$$

$$n = \frac{61313,86}{160.56}$$

$$n = 381$$

La muestra se tomó de las 5 parroquias urbanas de la ciudad de Latacunga y son 381 personas en matrimonio y divorciadas.

3.6. Recolección de la información

Para recoger la información se aplicó la técnica de la encuesta por medio de Google Forms (Formularios de Google) . En forma física se aplicó como instrumento el cuestionario que tiene 54 preguntas con opción múltiple. Las variables relacionadas están en una escala de 0 a 3. Donde 0 representa el peor estado y 3 es el mejor estado posible, las respuestas están en base a una escala de 3 puntos (0=Nunca,1=Pocas veces,2=Frecuentemente,3=Siempre).

3.7. Procesamiento y análisis estadístico de las variables

El procesamiento de la información se realiza una vez recolectada los datos a continuación se describe el extracto de datos de las encuestas aplicas en las siguientes tablas.

Tabla 2: Primer conjunto de variables de la base de datos.

Atr1	Atr2	Atr3	Atr4	Atr5	Atr6	Atr7	Atr8	Atr9	Atr10	Atr11	Atr12	Atr13	Atr14	Atr15	Atr16	Atr17	Atr18
3	2	3	2	3	0	0	2	2	2	2	3	3	3	3	3	3	3
2	3	2	3	2	0	1	3	2	3	3	3	3	3	3	3	2	3
3	2	3	2	3	0	0	2	3	2	2	3	3	3	3	2	3	3
3	3	3	3	3	0	0	3	3	2	2	3	3	3	3	2	3	3
2	2	2	3	2	1	0	2	2	3	2	2	3	2	2	3	2	3

Elaborado por: Hidalgo, 2022

Fuente: Programa Rstudio, 2022

En la tabla 2 se puede observar las variables independientes de entrada que va desde la pregunta Atr1 a la pregunta Atr18 tomando el resultado de cinco encuestados.

Tabla 3: Segundo conjunto de variables de la base de datos.

Atr19	Atr20	Atr21	Atr22	Atr23	Atr24	Atr25	Atr26	Atr27	Atr28	Atr29	Atr30	Atr31	Atr32	Atr33	Atr34	Atr35	Atr36
3	3	2	3	2	3	0	0	2	2	2	2	3	3	3	3	3	3
3	2	3	2	3	2	0	1	3	2	3	3	3	3	3	3	3	2
3	3	2	3	2	3	0	0	2	3	2	2	3	3	3	3	2	3
3	3	3	3	3	3	0	0	3	3	2	2	3	3	3	3	2	3
3	2	2	2	3	2	1	0	2	2	3	2	2	3	2	2	3	2

Elaborado por: Hidalgo, 2022

Fuente: Programa Rstudio, 2022

En la tabla 3 se puede observar las variables independientes de entrada que va desde la pregunta Atr19 a la pregunta Atr36 con cinco encuestados.

Tabla 4: Tercer conjunto de variables de la base de datos.

Atr37	Atr38	Atr39	Atr40	Atr41	Atr42	Atr43	Atr44	Atr45	Atr46	Atr47	Atr48	Atr49	Atr50	Atr51	Atr52	Atr53	Atr54	p(1=N,0=S)
3	3	3	3	3	3	3	3	3	3	3	2	3	2	3	0	3	2	1
3	3	3	3	3	3	3	2	3	3	2	3	2	3	2	0	2	3	1
3	3	3	3	3	3	2	3	3	3	3	2	3	2	3	0	3	2	1
3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	0	3	3	1
3	3	2	3	2	2	3	2	3	3	2	2	2	3	2	1	2	2	1

Elaborado por: Hidalgo, 2022

Fuente: Programa Rstudio, 2022

En la tabla 4 se puede observar las variables independientes de entrada que va desde la pregunta Atr37 a la pregunta Atr54 tomando el resultado de cinco encuestados. Además, se distingue la última columna donde muestra la variable de salida que es; divorciado y no divorciado. Hay 114 personas divorciadas y 267 personas no divorciadas.

Para la validación del cuestionario se calcula el índice del Alfa de Cronbach que es igual a 0.99. Esto indica que los atributos del instrumento que es el cuestionario de Gottman están correlacionados. Para su cálculo se utilizó el programa Rstudio (Ver Anexo1). Posteriormente se analiza la correlación presentada en la siguiente tabla.

Tabla 5: Selección de atributos usando la plataforma RStudio

Atributos	Correlación	Atributos	Correlación	Atributos	Correlación
Atr1	0,8	Atr19	0,9	Atr37	0,8
Atr2	0,7*	Atr20	0,8	Atr38	0,9
Atr3	0,7*	Atr21	0,7*	Atr39	0,8
Atr4	0,6*	Atr22	0,7*	Atr40	0,8
Atr5	0,8	Atr23	0,6*	Atr41	0,8
Atr6	0,3	Atr24	0,8	Atr42	0,8
Atr7	0,3	Atr25	0,3	Atr43	0,7*
Atr8	0,8	Atr26	0,3	Atr44	0,9
Atr9	0,9	Atr27	0,8	Atr45	0,8
Atr10	0,7*	Atr28	0,9	Atr46	0,9
Atr11	0,9	Atr29	0,7*	Atr47	0,8
Atr12	0,8	Atr30	0,9	Atr48	0,7*
Atr13	0,8	Atr31	0,8	Atr49	0,7*
Atr14	0,8	Atr32	0,8	Atr50	0,6*
Atr15	0,8	Atr33	0,8	Atr51	0,8
Atr16	0,7*	Atr34	0,8	Atr52	0,8
Atr17	0,9	Atr35	0,7*	Atr53	0,8
Atr18	0,8	Atr36	0,9	Atr54	0,7*

Elaborado por: Hidalgo, 2022
Fuente: Programa Rstudio, 2022

En la tabla 5 mediante el programa Rstudio se determinó la correlación entre las variables independientes respecto a la variable dependiente. Se muestra 15 variables seleccionadas (variables señaladas con un *), las cuales presentan una correlación mayor que 0,3 y menor que 0,8. Las variables independientes seleccionadas corresponden a los atributos: Atr2, Atr3, Atr4, Atr10, Atr16, Atr21, Atr22, Atr23, Atr29, Atr35, Atr43, Atr48, Atr49, Atr50, Atr54.

Con base en la plataforma Weka se realiza el proceso de selección de características de los atributos más importantes. Por medio del evaluador de atributos CfsSubsetEval usando el método de Greedy Stepwise como se muestra en la figura 3.

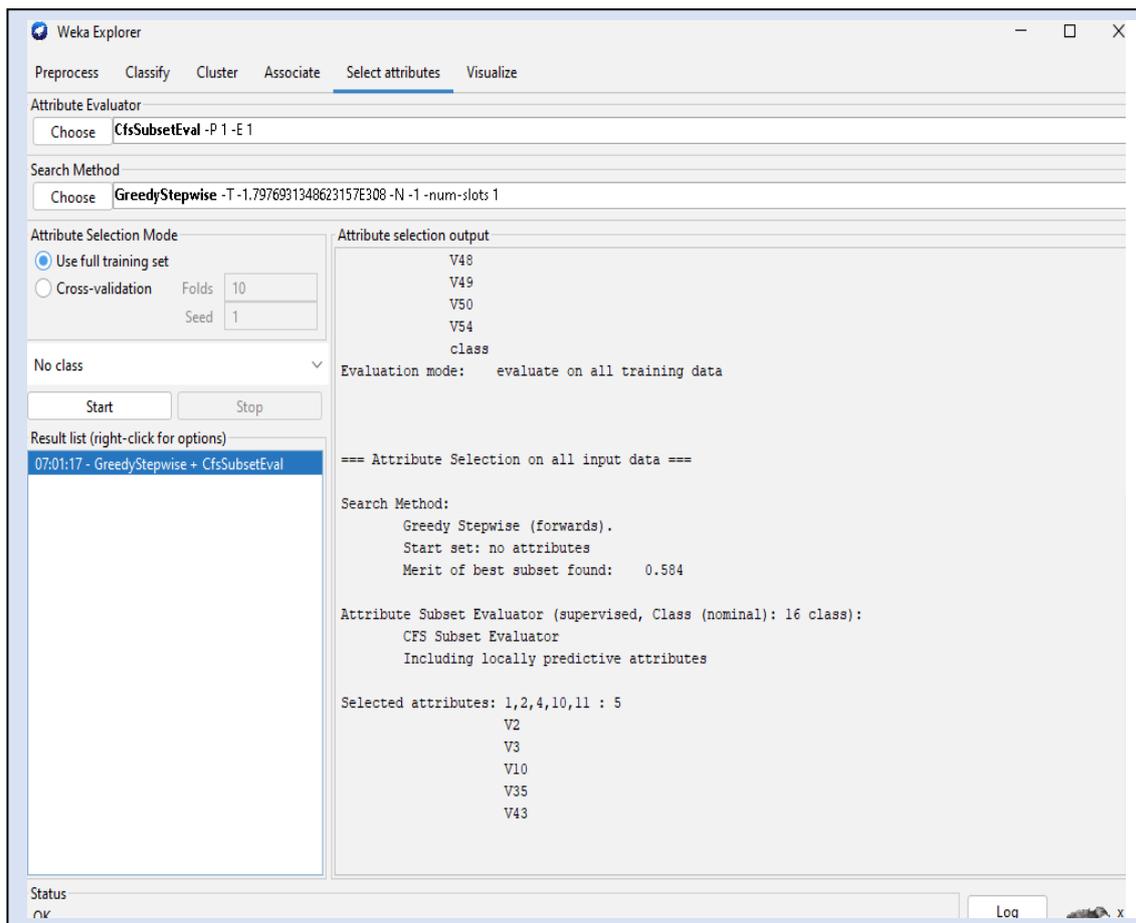


Figura 3: Selección de características más importantes usando el programa WEKA

Elaborado por: Hidalgo, 2022

Generándose cinco atributos entre ellos tenemos: Atr2, Atr3, Atr10, Atr35, Atr43. Estos atributos o características seleccionadas se utilizan para la

construcción del modelo empleando Regresión Logística presentado en la siguiente tabla.

Tabla 6: Características o variables seleccionadas

Variable independiente	Variable dependiente
Atr2: Tratar de ignorar las diferencias aún en situaciones difíciles	Divorcio
Atr3: Retomar las discusiones desde el principio y corregirlas	
Atr10: La mayoría de nuestros objetivos son poco comunes	
Atr35: Insulto a mi cónyuge en una discusión	
Atr43: Me quedo en silencio para calmar el ambiente en una discusión	

Elaborado por: Hidalgo, 2022

Fuente: Plataforma Weka, 2022

3.8. Análisis descriptivo de datos

De acuerdo a los resultados de la encuesta realizada a las personas divorciadas y no divorciadas en las parroquias urbanas, se describe los resultados alcanzados tomando como referencia el intervalo de edad.

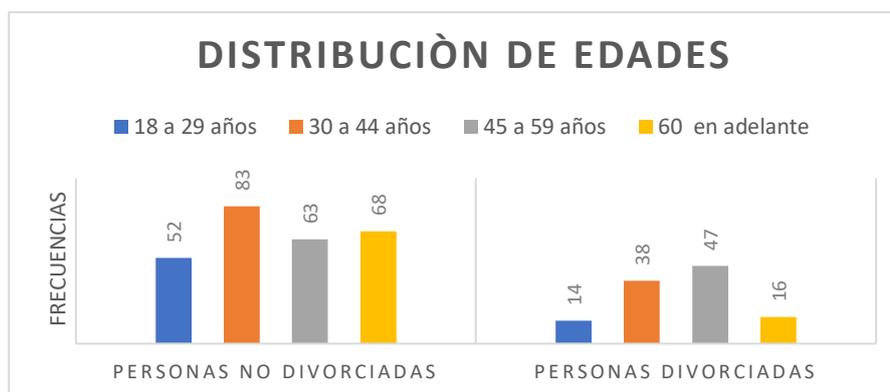


Figura 4: Distribución por intervalo de edad en personas divorciadas y no divorciadas.

Elaborado por: Hidalgo, 2022

En la figura 4 se ubica las respuestas obtenidas por las personas encuestadas. El 19,5% respondieron de 18 a 29 años, 31,2% de 30 a 44 años, 23,7% de 45 a 59 años y 25,6% de 60 años en adelante. Esto equivale a un total de 52, 83, 63

y 68 personas respectivamente. En resumen, más de la mitad de las parejas no divorciadas están en las edades de 18 a 29 años y de 30 a 44 años.

El 12,2% respondieron de 18 a 29 años, 33,0% de 30 a 44 años, 40,9% de 45 a 59 años y 13,9% de 60 años en adelante. Esto equivale a un total de 14, 38, 47 y 16 personas respectivamente. En resumen, más de la mitad de las parejas divorciadas están en las edades de 45 a 59 años y de 60 años en adelante.

En base a la encuesta aplicada se procede a describir los resultados que se obtuvieron en los gráficos con respecto a los atributos seleccionadas y que guardan relación con el divorcio estas son: Atr2, Atr3, Atr10, Atr35, Atr43.

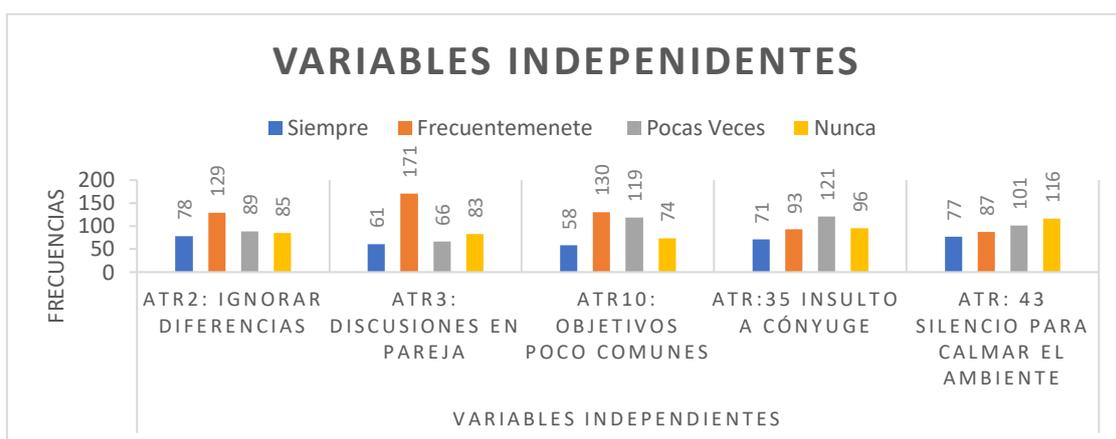


Figura 5: Resultados obtenidos en las variables independientes encuestadas. Elaborado por: Hidalgo, 2022

La figura 5 muestra las respuestas que dieron los usuarios a las siguientes variables: Atr2; Atr3; Atr10; Atr35; Atr45.

Para ello un 22,3% respondieron nunca, 23,3% pocas veces, 33,9% frecuentemente y 20,5% siempre. Esto equivale a un total de 85, 89, 129 y 78 personas respectivamente. En resumen, más de la mitad de las parejas encuestadas consideran que sí es posible tratar de ignorar las diferencias en situaciones difíciles.

El 21,8% respondieron nunca, 17,3% pocas veces, 44,9% frecuentemente y 16% siempre. Esto equivale a un total de 83, 66, 171 y 61 personas respectivamente. En resumen, más de la mitad de las parejas encuestadas consideran que sí es posible retomar las discusiones desde el principio y corregirlas.

El 19,5% respondieron que nunca, 31,2% pocas veces, 34,1 frecuentemente y 15,2% siempre. Esto equivale a un total de 74, 119, 130 y 58 personas respectivamente. En resumen, más de la mitad de las parejas encuestadas consideran que la mayoría de los objetivos no son poco comunes.

El 25,2% respondieron que nunca, 31,8% pocas veces, 24,4% frecuentemente y 18,6% siempre. Esto equivale a un total de 96, 121, 93 y 71 personas respectivamente. En resumen, más de la mitad de las parejas indican que no insultan a su cónyuge en una discusión.

El 30,5% respondieron que nunca, 26,5 pocas veces, 22,8 frecuentemente y 20,2% siempre. Esto equivale a un total de encuestados que representa 116, 101, 87 y 77 personas respectivamente. En resumen, más de la mitad de parejas indican que no se quedan en silencio para calmar el ambiente en una discusión.

Variable Dependiente: Divorcio

La figura 6 muestran el número de parejas encuestadas clasificadas en divorciados y no divorciados.

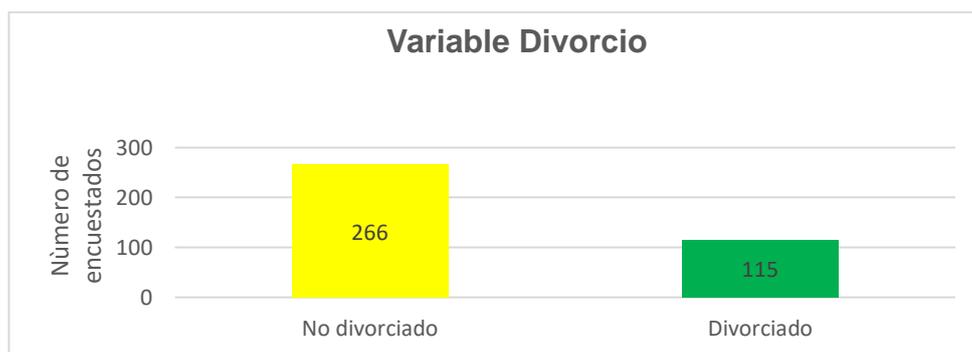


Figura 6: Resultado obtenido en la variable dependiente encuestada.

Elaborado por: Hidalgo, 2022

Para ello el 30.1% respondieron que son divorciados y el 69,9% contestaron que no son divorciados. Esto equivale a un total de 266 y 115 personas respectivamente. En resumen, más de la mitad son personas encuestadas no son divorciadas.

CAPÍTULO IV

RESULTADOS Y DISCUSIONES

En este capítulo se encuentran los resultados que se obtuvo luego de aplicar los procesos estadísticos.

4.1. Análisis de las características seleccionadas

Luego de haber seleccionado las variables independientes usando el programa Weka. Es necesario conocer qué relación existen entre las variables entonces, según la dirección es negativa o positiva que va desde -1 a +1. Si este valor es inferior al 25% se considera débil la correlación. En un rango del 25% y 75% se dice que tiene una relación moderada y si es mayor al 80% es fuerte la relación que existe entre las variables. Para el presente trabajo se visualiza la siguiente figura con la Matriz de correlación de las variables.

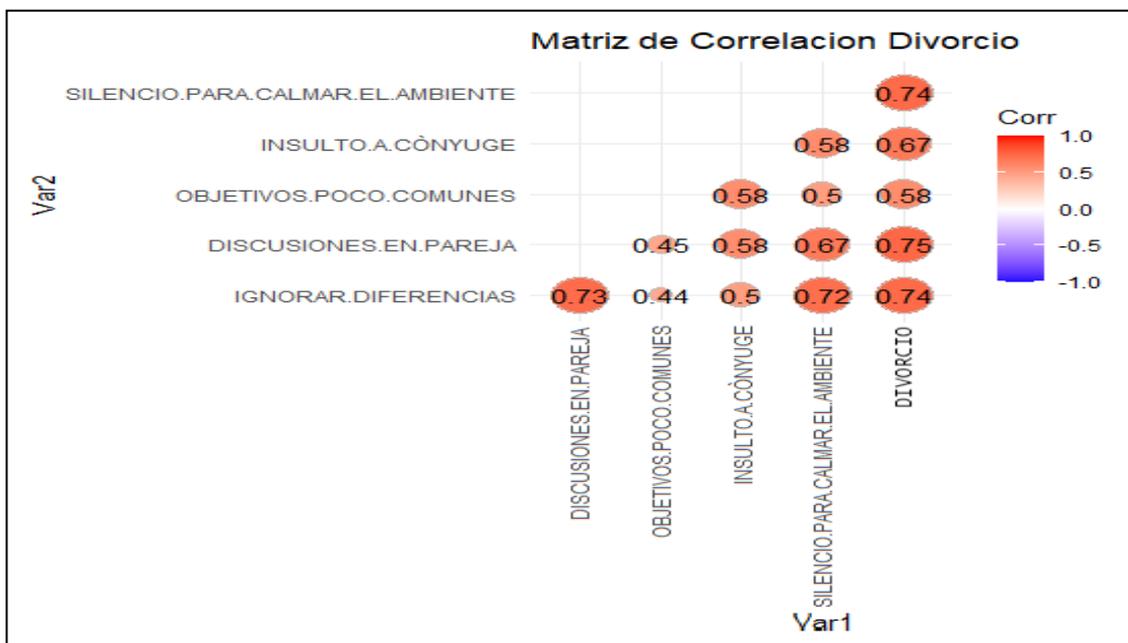


Figura 7: Matriz de correlación de las variables
Elaborado por: Hidalgo, 2022

De acuerdo a la figura 7 elaborada en el programa Rstudio con los datos de la encuesta aplicada. Se puede visualizar que el divorcio se encuentra relacionado un 74% con el Atr43(Me quedo en silencio para calmar el ambiente en una discusión), un 67% con el Atr35(Insulto a mi cónyuge en una discusión), se observa un 58% con el Atr10(La mayoría de nuestros objetivos son poco comunes), además un 75% con el Atr3(Retomar las discusiones desde el

principio y corregirlas) y por último un 74% con el Atr2(Tratar de ignorar las diferencias aún en situaciones difíciles). Se puede observar que las variables descritas en el gráfico están relacionadas moderadamente y de forma positiva con la variable dependiente divorcio.

4.2. Construcción del modelo de Regresión Logística

Para determinar si las variables seleccionadas son adecuadas se utiliza el programa RStudio descrito en la siguiente tabla:

Tabla 7: Construcción del modelo de Regresión Logística

MODELO DE REGRESIÓN LOGÍSTICA				
VARIABLES SIGNIFICATIVAS				
Min	1Q	Median	3Q	Max
-2,0179	-0,0540	0,0124	0,0742	3,1959
Coefficientes				
	Estimación	Error estándar	Estadístico (Z)	Probabilidad de Z
Intercepto	-6.5297	1.1094	-5.886	3.97e-09***
Atr2: IGNORAR DIFERENCIAS	1.4289	0.6269	2.279	0,022659*
Atr3: DISCUSIONES EN PAREJA	1.0616	0.4717	2.250	0.024417*
Atr10: OBJETIVOS POCO COMUNES	1.1379	0.5219	2.180	0,029223*
Atr35: INSULTO A CÓNYUGE	1.5079	0.4582	3.291	0.000998***
Atr43: SILENCIO PARA CALMAR EL AMBIENTE	2.1515	0.5794	3.713	0.000204***
Nivel de Significancia: 0 ‘***’ 0,001‘***’ 0,01‘**’ 0,05‘.’ 0.1’’ 1				
Parámetro de dispersión para la familia binomial tomada como 1				
Desviación nula:	466,659 en 380 grados de libertad			
Desviación residual:	60,071 en 375 grados de libertad			
AIC: 72,071				

Elaborado por: Hidalgo, 2022

Fuente: Programa Rstudio, 2022

En la tabla 7 se puede visualizar, en la primera columna los cinco atributos seleccionados más relevantes para crear el modelo de regresión logística. En la segunda columna están los coeficientes del modelo que acompañarán a las variables predictoras. En la tercera columna muestra el error estándar de cada atributo, y sirven para verificar en qué medida los coeficientes son distintos de cero en la población. En la cuarta columna se puede observar que valores del estadístico (Z) son mayores que 1,96. Esto corresponde al valor crítico que se usa para la distribución normal. Se obtiene dividiendo cada coeficiente para su error estándar

4.3. Verificación de hipótesis

Para verificar la hipótesis planteada en el presente trabajo se utiliza un nivel de significancia del 0,05 y un contraste de hipótesis.

Hipótesis nula (Ho): El modelo matemático no es significativo (todos los coeficientes β_i son cero).

Hipótesis alternativa (Ha): El modelo matemático es significativo (al menos un coeficiente β_i es diferente de 0).

Si el valor de la probabilidad de $|Z|$ es mayor que o igual a α se acepta la hipótesis nula.

Si el valor de la probabilidad de $|Z|$ es menor que α se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

Observando en la tabla 7 los valores de la quinta columna se ubican los valores de la probabilidad de $|Z|$ que son inferiores con respecto al nivel preestablecido de significancia 0,05. Por lo tanto, se procede a aceptar la hipótesis alternativa. En conclusión, el modelo de Regresión Logística es significativo porque existe al menos un coeficiente β_i es diferente de 0. Por lo tanto, estas variables o atributos sirven para explicar la variable dependiente.

4.4. Interpretación de los ODDS RATIO (OR)

Los coeficientes β del modelo de regresión logística se calculan usando el logaritmo natural con los valores de la tercera columna de los Odds ratio de

cada atributo (X). Para calcular de forma viceversa se halla su exponencial: OR = antilogaritmo (β_i) ; por lo tanto, $e (\beta_i)$.

Tabla 8: Interpretación de los ODDS RATIO (OR)

Características o variables seleccionadas	Coefficientes de β_i	Odds ratio
Intercepto	-6,52	$1,47 \times 10^{-3}$
Atr2: IGNORAR DIFERENCIAS	1,42	4,17
Atr3: DISCUSIONES EN PAREJA	1,06	2,89
Atr10: OBJETIVOS POCO COMUNES	1,13	3,12
Atr35: INSULTO A CÒNYUGE	1,50	4,51
Atr43: SILENCIO PARA CALMAR EL AMBIENTE	2,15	8,59

Elaborado por: Hidalgo, 2022
Fuente: Programa Rstudio, 2022

De acuerdo a la tabla 8 se puede apreciar los resultados de la columna Odds ratio. Estos valores son positivos en cada variable independiente. Al aumentar alguna variable o atributo hay la probabilidad que aumente la variable respuesta.

En la variable Atr2, cuando una pareja trata de ignorar diferencias aún en situaciones difíciles presenta un valor de razón odd ratio de 4,17. Por lo tanto, la probabilidad de divorcio aumenta cuatro veces más cuando se ignoran las diferencias. Con respecto a la variable Atr3, cuando una pareja trata de retomar las discusiones desde el principio y corregirlas tiene un valor odd ratio de 2.89. Por lo tanto, la probabilidad de divorciarse aumenta dos veces más cuando se presenta discusiones entre los cónyuges. Mientras que para la variable Atr10, cuando la pareja tiene objetivos poco comunes tiene un resultado de odd ratio 3,12. Por lo tanto, la probabilidad de divorcio aumenta tres veces más cuando la pareja tiene objetivos poco comunes. Por otro lado, la variable Atr35, cuando se insulta a un cónyuge en una discusión tiene un valor de odd ratio 4.51. Por lo tanto, la probabilidad que tiene la pareja de divorciarse aumenta cuatro veces más cuando se insulta a un cónyuge. Por último, la variable Atr43 cuando en la pareja un cónyuge se queda en silencio para calmar el ambiente

en una discusión tiene un resultado de odd ratio 8.59. Por lo tanto, se puede afirmar que la probabilidad de divorcio en una pareja aumenta ocho veces más cuando un cónyuge se queda en silencio en una discusión.

Con la verificación de los 5 atributos o variables independientes seleccionadas. Se observa que son adecuadas y sirven para la construcción del respectivo modelo de regresión logística. El planteamiento de la siguiente ecuación ayuda a predecir los divorcios en la ciudad de Latacunga y tiene la siguiente estructura:

$$P(y_i = 1/x_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5)}}$$

Donde:

P: Probabilidad de ocurrencia según las variables.

Y_i : variable dependiente.

X_i : Atributos o variables independientes.

B: Son los coeficientes del modelo.

Al reemplazar los coeficientes que se calculó en la tabla 10, resulta la siguiente ecuación del modelo que ayudará a predecir de divorcios en la ciudad de Latacunga.

$$P(y_i = 1/x_i) = \frac{1}{1 + e^{-(-6.5297 + 1.4289X_1 + 1.0616X_2 + 1.1379X_3 + 1.5079 X_4 + 2.1515X_5)}}$$

4.5. Implementación del modelo predictivo Regresión Logística para determinar la probabilidad del divorcio.

Tabla 9: División del conjunto de datos y testeo para regresión logística

Nombre	Entrenamiento y Prueba		
	Total (datos)	Datos de entrenamiento (70%)	Datos de testeo (30%)
Data (Divorcio)	381	267	114

Elaborado por: Hidalgo, 2022

Fuente: Programa Rstudio, 2022

En la tabla 9 se toma el archivo de la base de datos (divorcio). Se realiza la división en dos grupos: El 70% para entrenamiento existen 267 casos, y el 30% para realizar el testeo está estructurado por 114 casos.

4.6. Matriz de confusión de Regresión Logística

Para la realización de este modelo predictivo se emplea el 30% de los datos que son considerados para la prueba del modelo. Se determinó que el 98% de los datos de testeo son clasificados de forma correcta. La respectiva matriz de confusión con datos de prueba se muestra en la siguiente figura.

		Predicciones	
		Divorciado (0)	No divorciado (1)
Valor verdadero	Divorciado (0)	29	1
	No divorciado (1)	1	83

Figura 8: Matriz de confusión con datos de testeo de Regresión Logística
Elaborado por: Hidalgo, 2022

En la figura 8 se puede observar la matriz de confusión con los datos de testeo, se analiza los siguientes resultados: Tenemos 114 personas. De esas 114, 30 son personas divorciadas y 84 no son divorciadas.

En diagonal principal se observa 29 verdaderos negativos (VN) que representa la cantidad de divorciados que fueron clasificados correctamente como divorciados. Son 83 verdaderos positivos (VP) que corresponde a la cantidad de no divorciados que son clasificados correctamente como no divorciados.

En la otra diagonal se observa 1 falso positivo (FP) que representa el valor que la predicción ha clasificado como divorciado, pero que realmente es no divorciado. Además, existe 1 falso negativo (FN) que corresponde al valor que la predicción ha clasificado como no divorciados, pero que en realidad son divorciados.

4.7. Métricas para evaluar el desempeño de clasificación

Se realiza los siguientes cálculos: valores de exactitud (Accuracy), precisión, sensibilidad (Recall), y puntuación F1-score. Se calcula en base a los datos que tiene la matriz de confusión de los datos de testeo. Los valores determinados se representan en la siguiente tabla.

Tabla 10: Valores calculados de la matriz de confusión

Métricas	Valores
Exactitud (Accuracy)	0,98
Precisión	0,98
Sensibilidad (Recall)	0,98
Puntuación (F1-Score)	0,98

Elaborado por: Hidalgo, 2022

Fuente: Programa Rstudio, 2022

En la tabla 10 se puede observar las diferentes métricas calculadas que sirven para verificar el rendimiento del modelo.

La Exactitud evalúa los casos que el modelo acertó. Se calculó dividiendo el total de los aciertos (VP+VN) sobre el número de casos. En nuestro caso sería 112 sobre 114. Su valor es de 0,98 que representa la tasa de aciertos del modelo. Es decir, de 100 casos 98 son clasificados correctamente en forma general.

La precisión se calculó dividiendo la cantidad de VP sobre la suma de los (VP+FP). En nuestro modelo sería 83 sobre (83 + 1). Eso nos da un valor de 0,98. Esto representa la cantidad de personas que son clasificados realmente como no divorciados. De 100 casos 98 son clasificados correctamente como no divorciados. Es un valor alto y representa que existe pocos falsos positivos.

La sensibilidad se calculó dividiendo los VP sobre la suma de los (VP+FN). En nuestro caso sería 83 sobre (83 + 1). Da un valor de 0,98. Es un valor alto por que existen pocos falsos negativos. Muestra el porcentaje de personas no

divorciadas que fueron correctamente identificadas. De 100 casos 98 son identificados.

La puntuación F1-Score se utiliza cuando los datos se encuentran desbalanceados. Se calculó dividiendo la multiplicación de la precisión por la sensibilidad sobre la suma de la precisión más la sensibilidad. El resultado final se multiplica por dos. En nuestro modelo sería 0,98 por 0,98 dividido para 0,98 + 0,98. Da un valor de 0,49 y se multiplica por dos. Su resultado da un valor de 0,98. De 100 casos, 98 casos son predichos correctamente por el modelo de Regresión Logística.

4.8. Implementación del modelo predictivo Naive Bayes para determinar la probabilidad del divorcio.

Tabla 11: División del conjunto de datos y testeo para Naive Bayes

Nombre	Entrenamiento y Prueba		
	Total (datos)	Datos de entrenamiento (70%)	Datos de testeo (30%)
Data (Divorcio)	381	267	114

Elaborado por: Hidalgo, 2022
Fuente: Programa Rstudio, 2022

En la tabla 11 se observa el archivo de la base de datos (divorcio). Se realiza la división de la data en dos grupos: El 70% para entrenamiento corresponde a 267 casos, y el 30% para realizar el testeo está compuesto por 114 casos según la descripción.

4.9. Matriz de confusión del modelo Naive Bayes

Para la realización de este modelo predictivo se considera el 30% de datos que son considerados para la prueba del modelo. Se determinó que el 96% de los datos de testeo son clasificados de forma correcta. La respectiva matriz de confusión con datos de prueba se muestra en la siguiente figura que describe los resultados obtenidos

		Predicciones	
		Divorciado (0)	No divorciado (1)
Valor verdadero	Divorciado (0)	27	3
	No divorciado (1)	1	83

Figura 9: Matriz de confusión con datos de testeo de Naive Bayes
Elaborado por: Hidalgo, 2022

En la figura 9 se observa la matriz de confusión del modelo Naive Bayes con los datos de testeo, se describe los siguientes resultados: Existen 114 personas. De esas 114, 30 corresponde a personas divorciadas y 84 a personas no son divorciadas.

En la primera diagonal se puede observar 27 verdaderos negativos (VN) que representa la cantidad de divorciados que son clasificados correctamente como divorciados. Existen 83 verdaderos positivos (VP) que corresponde a la cantidad de no divorciados que fueron clasificados correctamente como no divorciados.

En la otra diagonal se observa 1 falso positivo (FP) que representa el valor que la predicción ha clasificado como divorciado, pero que realmente es no divorciado. Además, existe 3 falsos negativos (FN) que corresponde al valor que la predicción ha clasificado como no divorciados, pero que en realidad son divorciados.

4.10. Métricas para evaluar el desempeño de clasificación

Se realiza los siguientes cálculos: valores de exactitud (Accuracy), precisión, sensibilidad (Recall), y puntuación F1-score en base a los datos que tiene la matriz de confusión de los datos de testeo, estos valores calculados se representan en la siguiente tabla 7.

Tabla 12: Valores calculados de la matriz de confusión

Métricas	Valores
Exactitud (Accuracy)	0,96
Precisión	0,96
Sensibilidad (Recall)	0,98
Puntuación(F1-Score)	0,96

Elaborado por: Hidalgo, 2022

Fuente: Programa Rstudio, 2022

En la tabla 12 se observan las diferentes métricas calculadas que sirven para verificar el rendimiento del modelo.

La Exactitud (Accuracy) evalúa los casos que el modelo acertó. Se calculó dividiendo el total de los aciertos (VP+VN) sobre el número de casos. En nuestro caso sería 110 sobre 114. Su valor es de 0,96 que representa la tasa de aciertos del modelo. Es decir, de 100 casos 96 son clasificados correctamente en forma general.

La precisión se calculó dividiendo la cantidad de VP sobre la suma de los (VP+FP). En nuestro modelo sería 83 sobre (83 + 1). Eso nos da un valor de 0,98. Esto representa la cantidad de personas que son clasificados realmente como no divorciados. De 100 casos 98 son clasificados correctamente como no divorciados. Es un valor alto y representa que existe pocos falsos positivos.

La sensibilidad se calculó dividiendo los VP sobre la suma de los (VP+FN). En nuestro caso sería 83 sobre (83 + 3). Da un valor de 0,96. Muestra el porcentaje de personas no divorciadas que fueron correctamente identificadas. De 100 casos 98 son identificados.

La puntuación F1-Score se emplea cuando los datos se encuentran desbalanceados. Se calculó dividiendo la multiplicación de la precisión por la sensibilidad sobre la suma de la precisión más la sensibilidad. El resultado final se multiplica por dos. En nuestro modelo sería 0,98 por 0,96 dividido para 0,98 +0,96. Da un valor de 0,48 multiplicado por dos. Su resultado un valor de 0,96. De 100 casos 96 casos son predichos correctamente.

Se construye la siguiente tabla con las dos métricas más importantes para evaluar el modelo.

Tabla 13: Resultado de las métricas calculadas de los modelos

Modelos	Precisión	F1-score
Regresión Logística	0,98	0,98
Naive Bayes	0,96	0,96

Elaborado por: Hidalgo, 2022

Fuente: Programa Rstudio, 2022

En la tabla 13 la precisión indica el porcentaje de predicciones positivas correctas que realiza el modelo. Esto representa la cantidad de personas que realmente son clasificadas como no divorciadas. La puntuación F1-Score combina la precisión y la sensibilidad. Se utiliza cuando existe un desbalanceo de datos.

El modelo de regresión logística presenta una precisión de 0,98. Es decir de cada 100 casos el modelo clasifica 98 casos correctamente y 2 se equivocó. La puntuación F1-Score tiene un resultado de 0,98. De 100 casos 98 son correctos.

El modelo de Naive Bayes tiene una precisión de 0,96. Indica de 100 casos el modelo clasifica correctamente 96 casos y 4 se equivocó. La puntuación F1-Score muestra un resultado de 0,96. De cada 100 casos el modelo clasifica 96 de forma correcta.

Por lo tanto, la puntuación de F1-Score de la regresión logística es mayor que el modelo de Naive Bayes. Por lo tanto, el mejor modelo es de regresión logística por tener mejor capacidad de predicción y ayudará a predecir los divorcios en la ciudad de Latacunga.

4.11. Análisis de casos Falsos Positivos (FP) y Falsos Negativos (FN)

FP son los falsos positivos: Corresponde al número de personas no divorciadas en los que la prueba predijo como personas divorciadas. Arroja un diagnóstico incorrecto en la realidad. El principal motivo de un FP es que la persona no ha respondido correctamente o no fue sincera a su realidad. Muchas parejas son casadas por fuera, pero por dentro pueden experimentar

situaciones difíciles y las razones están arraigadas en su vida interior. Entre los principales motivos tenemos: Existen familias que tienen como mandato que el matrimonio es para toda la vida, y muchas parejas en la actualidad siguen juntas por temor a defraudar a sus familias. Otro factor es el miedo a quedarse solos, en muchos casos prefieren continuar con su vida aburrida y monótona antes de estar solos especialmente por su edad. El continuar juntos porque tienen hijos en común, sin darse cuenta que los hijos no serán personas alegres si viven en ambientes hostiles y desagradables conllevando a una infancia dolorosa. Además, existen personas que dependen económicamente de su pareja y ni siquiera se permiten imaginarse separados porque son amas de casa o temen insertarse a un mercado laboral. Por otro lado, existen personas que tienen miedo de arrepentirse a futuro luego de haber tomado esa decisión y prefieren no hacerlo.

Esto conlleva a tener parejas infelices aun siendo parejas casadas en matrimonio generándose falsos positivos, estos son los principales motivos que mucha gente a su alrededor no termina de entender. Pocos recurren a la ayuda de un profesional para mejorar su relación de pareja o la toma de decisiones acertadas y otros quedan aferrados a esa vida con la esperanza que en un futuro la situación pueda cambiar a mejor.

FN son los falsos negativos: Corresponde al número de personas divorciadas en los que al aplicar la prueba predijo como personas no divorciadas. Generando un diagnóstico incorrecto en la realidad. Su principal razón para que se genere un FN es que la persona no respondió con sinceridad el test propuesto. Muchas parejas en la actualidad son divorciadas, pero no han superado el proceso del divorcio ni a su ex pareja. Algunas personas divorciadas piensan que todavía están casadas ante los ojos de Dios, aduciendo la cuestión religiosa y en otros casos existen personas que están separadas de su pareja y no han iniciado el proceso de divorcio. En el peor de los casos siguen viéndose con su ex pareja lo que genera contextos inequívocos en el subconsciente de la persona, desentonando la realidad de las cosas, no respetándose como persona dejando atrás su dignidad como valor individual de cada ser humano.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

En base a la elaboración del trabajo se obtiene las siguientes conclusiones:

- En este trabajo se ha presentado un modelo para predecir los divorcios, donde el área de estudio es la ciudad de Latacunga. En esta ciudad se tomó una muestra de 381 personas formadas por parejas casadas y parejas divorciadas. Se aplicó la encuesta sugerida por el profesor Gottman que se utiliza para predecir el divorcio. Más de la mitad de las parejas no divorciadas están en el rango de 18 a 44 años, y más de la mitad de las parejas divorciadas en un rango de 45 años en adelante.
- Para el proceso de selección de características más importantes para la construcción del modelo se utilizó la plataforma Weka. Por medio del evaluador de atributos CfsSubsetEval, empleando el método de GreedyStepwise, se generó cinco atributos o características que pueden causar el divorcio: Atr2 (Ignorar las diferencias aún en situaciones difíciles), Atr3(Retomar las discusiones desde el principio y corregirlas), Atr10 (La mayoría de nuestros objetivos son comunes), Atr35: (Insulto a mi cónyuge en una discusión) y Atr43: Me quedo en silencio para calmar el ambiente.
- Para el proceso de modelación se utilizó dos algoritmos de clasificación diferentes: Regresión Logística y Naive Bayes, usando el 70% de datos de entrenamiento y el 30% como datos de testeo. Para verificar el desempeño de clasificación de los modelos se aplicó las siguientes métricas de evaluación: Exactitud, precisión, sensibilidad y puntuación F1-Score. Los falsos positivos se originan porque existen familias que tienen como mandato que el matrimonio es para toda la vida. Otro factor es el miedo a quedarse solos, el continuar juntos porque tienen hijos en común. Además, existen personas que dependen económicamente de su pareja. Los falsos negativos se dan porque muchas parejas en la

actualidad son divorciadas, pero no han superado el proceso del divorcio ni a su ex pareja.

- El valor de la puntuación F1-Score del modelo de Regresión Logística es de 0,98 con una precisión del 0,98, en cambio el valor F1-Score del modelo Naive Bayes es de 0,96 y su precisión de 0,96. El modelo de Regresión Logística es el mejor modelo, porque tiene mayor capacidad de predicción para estimar los divorcios en la ciudad de Latacunga.

5.2. RECOMENDACIONES

- Aplicar otros modelos predictivos en futuras investigaciones: Perceptrón, Árbol de decisión, Máquina de vectores soporte, que permita evaluar y probar el rendimiento de los modelos.
- Utilizar el algoritmo de clasificación Regresión Logística dirigido a la zona rural, para evidenciar si existen otras variables que pueden conllevar al divorcio en una pareja.
- Medir la precisión en diferentes modelos de aprendizaje, pero con diferentes entrenamientos y divisiones de prueba y establecer la mejor métrica de evaluación.
- Usar modelos de aprendizaje no supervisado para verificar si existen otras variables diferentes que ayuden a predecir el divorcio en la ciudad de Latacunga. Además, sería importante realizarlo por rango de edades para conocer diferentes causas.

5.3. Bibliografía

- [1] A. Sharma, A. S. Chudhey, y M. Singh, «Divorce case prediction using Machine learning algorithms», en *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, mar. 2021, pp. 214-219. doi: 10.1109/ICAIS50930.2021.9395860.
- [2] S. Sohail, S. Aziz, F. Tahir, S. Haqqi, y A. Hussain, «Implementation of machine learning algorithm on factors effecting divorce rate», en *2018 International Conference on Engineering and Emerging Technologies (ICEET)*, feb. 2018, pp. 1-5. doi: 10.1109/ICEET1.2018.8338618.
- [3] D. Swenson, «A Logit Model of the Probability of Divorce», *J. Divorce Remarriage*, vol. 25, n.º 1-2, pp. 173-194, feb. 1996, doi: 10.1300/J087v25n01_12.
- [4] M. J. C. Cabay, M. E. C. Pataron, y M. A. D. Castelo, «Factores que inciden en los divorcios prematuros en el Ecuador: un modelo de regresión logística», *Polo Conoc. Rev. Científico - Prof.*, vol. 6, n.º 3, pp. 1692-1706, 2021.
- [5] O. Castrillón, «Predicción del divorcio por medio de técnicas inteligentes», *Inf. Tecnológica*, vol. 32, pp. 111-120, oct. 2021, doi: 10.4067/S0718-07642021000500111.
- [6] S. Goel, S. Roshan, R. Tyagi, y S. Agarwal, «Augur Justice: A Supervised Machine Learning Technique To Predict Outcomes Of Divorce Court Cases», en *2019 Fifth International Conference on Image Information Processing (ICIIP)*, nov. 2019, pp. 280-285. doi: 10.1109/ICIIP47207.2019.8985764.
- [7] Md. Ashrafujjaman Tutul, Md. M. Hasan, S. K. Mondol, M. Hossain, y A. Al Marouf, «Divorce Prediction using Machine Learning Methods-Bangladesh Perspective», en *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, jul. 2021, pp. 1-5. doi: 10.1109/ICCCNT51525.2021.9579949.
- [8] P. Ranjitha y A. Prabhu, «Improved Divorce Prediction Using Machine learning- Particle Swarm Optimization (PSO)», en *2020 International Conference for Emerging Technology (INCET)*, jun. 2020, pp. 1-5. doi: 10.1109/INCET49848.2020.9154081.
- [9] A. Juárez-López, J. Hernández-Torruco, B. Hernández-Ocaña, y O. Chávez-Bosquez, «Comparison of classification algorithms using feature selection», en *2021 Mexican International Conference on Computer Science (ENC)*, ago. 2021, pp. 1-6. doi: 10.1109/ENC53357.2021.9534831.
- [10] A. Guillén *et al.*, «Predicting marital dissolutions using radial basis function neural networks», en *The 2010 International Joint Conference on Neural Networks (IJCNN)*, jul. 2010, pp. 1-4. doi: 10.1109/IJCNN.2010.5596691.
- [11] C. Wang, D. Chen, Y. Hu, Y. Ceng, J. Chen, y H. Li, «Automatic Dialogue System of Marriage Law Based on the Parallel C4.5 Decision Tree», *IEEE Access*, vol. 8, pp. 36061-36069, 2020, doi: 10.1109/ACCESS.2020.2972586.
- [12] N. Flores y S. Silva, «Machine learning model to predict the divorce of a married couple», *3C Tecnol. Innov. Apl. Pyme*, pp. 83-95, may 2021, doi: 10.17993/3ctecno.2021.specialissue7.83-95.

- [13] A. Shankhdhar, T. Gupta, y Y. V. Gautam, «Divorce Prediction Scale Using Improvised Machine Learning Techniques», en *Inventive Systems and Control*, Singapore, 2021, pp. 777-788. doi: 10.1007/978-981-16-1395-1_57.
- [14] Mustafa, «(PDF) DIVORCE PREDICTION USING CORRELATION BASED FEATURE SELECTION AND ARTIFICIAL NEURAL NETWORKS □». https://www.researchgate.net/publication/334170931_DIVORCE_PREDICTION_USING_CORRELATION_BASED_FEATURE_SELECTION_AND_ARTIFICIAL_NEURAL_NETWORKS (accedido 13 de junio de 2022).
- [15] Kong - Chai, «¿Es su matrimonio confiable? | Actas de la 6.ª Conferencia Internacional sobre Computación e Inteligencia Artificial de 2020». <https://dl.acm.org/doi/abs/10.1145/3404555.3404559> (accedido 13 de junio de 2022).
- [16] J. Kong, K. Kurz, y S. Heinrich, «Predicting Divorces Using Machine Learning Methods», p. 15.
- [17] S. Joel, «Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies | PNAS». <https://www.pnas.org/doi/10.1073/pnas.1917036117> (accedido 13 de junio de 2022).
- [18] R. E. Heyman y A. M. Smith Slep, «The Hazards of Predicting Divorce Without Crossvalidation», *J. Marriage Fam.*, vol. 63, n.º 2, pp. 473-479, may 2001, doi: 10.1111/j.1741-3737.2001.00473.x.
- [19] R. I. A. Mohamed y M. K. Alkhyeli, «EARLY DIVORCE PREDICTION IN ABU DHABI», p. 16.
- [20] S. L. AlKhayyat y A. M. Alhaqbani, «A comparison between binary logistic regression and discriminant analysis in determining the most important factors affecting the instability of marital life in the Kingdom of Saudi Arabia», *Appl. Math. Sci.*, vol. 15, n.º 15, pp. 725-736, 2021, doi: 10.12988/ams.2021.916585.
- [21] D. T. González, M. C. P. Arana, y E. V. Torres, *Cartagena en movimiento*. Universidad Tecnológica de Bolívar, 2021.
- [22] «MANUAL WEKA.pdf». Accedido: 30 de junio de 2022. [En línea]. Disponible en: <https://knowledgesociety.usal.es/sites/default/files/MANUAL%20WEKA.pdf>
- [23] P. J. C. Parrilla, J. Á. G. Requena, y S. M. González, *Cumplimiento cooperativo y reducción de la conflictividad: hacia un nuevo modelo de relación entre la Administración tributaria y los contribuyentes*. Aranzadi / Civitas, 2021.
- [24] *Introducción a la Minería de Datos*. Editora E-papers.
- [25] C. G. González y A. B. Felpeto, *Tratamiento de datos*. Ediciones Díaz de Santos, 2006.
- [26] R. Á. Cáceres, *Estadística multivariante y no paramétrica con SPSS: aplicación a las ciencias de la salud*. Ediciones Díaz de Santos, 1994.
- [27] B. García y J. Jesús, «HERRAMIENTAS INFORMÁTICAS DE PRODUCTIVIDAD APLICADAS A LOS MÉTODOS CUANTITATIVOS: MODELO DE REGRESIÓN LOGÍSTICA.», p. 20.
- [28] M. Rodríguez Díaz, Á. León, A. Henao, y M. E. Mora Díaz, *Introducción al análisis estadístico multivariado aplicado: Experiencia y casos en el Caribe colombiano*. Universidad del Norte, 2015.

- [29] Mustafa Murat, «Logistic Regression / Odds / Odds Ratio / Risk», *Mustafa Murat ARAT*, 5 de septiembre de 2019. <https://mmuratarat.github.io//2019-09-05/odds-ratio-logistic-regression> (accedido 3 de agosto de 2022).
- [30] «modelo_predictivo_exito_academico_algoritmos.pdf». Accedido: 30 de junio de 2022. [En línea]. Disponible en: https://repositoriotec.tec.ac.cr/bitstream/handle/2238/9389/modelo_predictivo_exito_academico_algoritmos.pdf?sequence=1&isAllowed=y
- [31] J. Torres, *Python Deep Learning: Introducción práctica con Keras y TensorFlow 2*. Marcombo, 2020.
- [32] Y.-N. González-Meneses, B.-E. Pedroza-Méndez, F. López, C. Pérez-Corona, y J.-F. Ramírez-Cruz, «Implementación del clasificador Naive Bayes para la Acentuación Automática de Palabras Ambiguas del Español.», p. 15.
- [33] C. de V. de Informática, *Sistemas expertos probabilísticos*. Univ de Castilla La Mancha, 1998.
- [34] A. Rodríguez Serrano, Y. Abreu Bartomeo, y L. Dominguez Cruz, «Clasificación supervisada de documentos mediante el algoritmo Naive Bayes», ene. 2015, Accedido: 5 de julio de 2022. [En línea]. Disponible en: <https://repositorio.uci.cu/jspui/handle/ident/8587>
- [35] «dc-latacunga.pdf». Accedido: 6 de julio de 2022. [En línea]. Disponible en: <https://colectivosciudadanoseducacion.files.wordpress.com/2017/02/dc-latacunga.pdf>
- [36] R. Hernandez Sampieri, C. Fernandez Collado, y P. Baptista Lucio, *Metodología de la investigación*. México: McGraw Hill Interamericana, 2014.

5.4. ANEXOS

Anexo 1

TEST DE GOTTMAN PARA PAREJAS CASADAS Y DIVORCIADAS

Encuesta para identificar las características que definen a una persona casada o divorciada en la zona urbana de la ciudad de Latacunga.

INSTRUCCIONES:

Contestar a todas las preguntas o variables considerando las siguientes opciones de respuesta en base a su experiencia vivida.

Marcar con una X la respuesta que considere:

I. ASPECTO GENERALES

1. Genero

Masculino	
Femenino	

2. EDAD

De 18 a 29 años	
De 30 a 44 años	
De 45 a 59 años	
De 60 en adelante	

II. ASPECTOS ESPECÍFICOS

VARIABLES O ATRIBUTOS

1. Si alguien se disculpa, cuando la discusión empeora, la discusión termina.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

2. Tartar de ignorar las diferencias aun en situaciones difíciles.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

3. Retomar las discusiones desde el principio y corregirlas.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

4. Cuando discuto con mi cónyuge, funcionara contactarlo.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

5. El tiempo que paso con mi cónyuge es especial para nosotros.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

6. No pasamos tiempo en casa como socios.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

7. Somos dos extraños que comparten el mismo ambiente en casa más que en familia.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

8. Disfruto los festivos con mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

9. Disfruto viajar con mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

10. La mayoría de nuestros objetivos son poco comunes.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

11. Pienso que un día en el futuro, cuando mire hacia atrás, veré que cada uno ha estado en armonía.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

12. Tenemos valores similares referentes a la libertad.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

13. Tenemos formas similares de entretenernos.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

14. La mayoría de los objetivos para las personas son los mismos.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

15. Nuestros sueños con mi cónyuge son similares y armoniosos.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

16. Somos compatibles sobre lo que debería ser el amor.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

17. Compartimos los mismos puntos de vista sobre cómo ser felices en nuestra vida conyugal.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

18. Tenemos ideas similares sobre cómo debería ser el matrimonio.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

19. Tenemos ideas similares sobre cómo deben ser los roles en el matrimonio.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

20. Tenemos valores similares de confianza.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

21. Se exactamente que le gusta a mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

22. Se cómo mi cónyuge quiere que lo cuiden cuando está enfermo.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

23. Conozco la comida favorita de mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

24. Puedo decir que tipo de estrés enfrenta mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

25. Conozco el mundo interior de mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

26. Conozco las ansiedades básicas de mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

27. Conozco las fuentes actuales de estrés de mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

28. Conozco las esperanzas y deseos de mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

29. Conozco muy bien a mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

30. Conozco los amigos de mi cónyuge y sus relaciones sociales.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

31. Me siento agresivo cuando discuto con mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

32. Cuando discuto usualmente uso expresiones como: usted siempre, usted nunca.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

33. Uso frases negativas acerca de la personalidad de mi cónyuge en una discusión.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

34. Use expresiones ofensivas en una discusión.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

35. Insulto a mi cónyuge en una discusión.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

36. Puedo ser humillado cuando discutimos.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

37. Las discusiones con mi cónyuge no son calmadas.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

38. Odio la forma en que mi cónyuge abre un tema.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

39. Las discusiones a menudo ocurren de repente.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

40. Comenzamos una discusión antes de saber que está pasando.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

41. Cuando hablo con mi conyugue sobre algo, de repente rompo mi calma.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

42. Cuando discuto con mi conyugue solo salgo y no digo una sola palabra.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

43. Me quedo en silencio para calmar el ambiente en una discusión.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

44. Pienso que es bueno que me vaya de casa un rato.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

45. Prefiero quedarme en silencio que discutir con mi conyugue.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

46. Incluso si tengo razón permanezco en silencio para no lastimar a mi cónyuge.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

47. Cuando hablo con mi cónyuge, me quedo en silencio porque tengo miedo de no poder controlar mi ira.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

48. Me siento bien en nuestras discusiones.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

49. No tengo nada que ver con lo que me acusan.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

50. No soy culpable de lo que me acusan.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

51. No soy el que se equivoca con los problemas en casa.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

52. No dudaría en decirle a mi cónyuge sobre su insuficiencia.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

53. Cuando discuto le recuerdo a mi cónyuge su insuficiencia.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

54. No tengo miedo de decirle a mi cónyuge sobre su incompetencia.

0=Nunca		2=Frecuentemente	
1=Pocas veces		3=Siempre	

Figura 10: Cuestionario método de Gottman

Fuente: Elaborado por el autor

ANEXO 2

```
# Paquete de R para examinar el "Alfa de Cronbach"
install.packages("psych")
library(psych)

# Base de datos
install.packages("tidyverse")
library(tidyverse)

View(Prediccion_Divorcio_R)
psych::alpha(Prediccion_Divorcio_R)

Reliability analysis
Call: psych::alpha(x = Prediccion_Divorcio_R)

  raw_alpha std.alpha G6(smc) average_r S/N   ase mean  sd median_r
    0.99     0.99    0.98     0.67 113 0.00045 1.5 0.93    0.76

  95% confidence boundaries
        lower alpha upper
Feldt   0.99  0.99  0.99
Duhachek 0.99  0.99  0.99
```

Figura 11: Calculo del Alfa de Cronbach

Fuente: Elaborado por el autor utilizando el programa Rstudio