



UNIVERSIDAD DE GUAYAQUIL

TRABAJO DE TITULACIÓN QUE SE PRESENTA COMO
REQUISITO PARA OPTAR POR EL GRADO DE MAGISTER EN
INGENIERIA EN SOFTWARE

DISEÑO DE UN MODELO PREDICTIVO COMO APOYO DE
TOMA DE DECISIONES EN LA RELACIÓN ENTRE SOÑADOR
Y VOLUNTARIO DE LA FUNDACIÓN “AMBULANCIA DESEO”

AUTORA: CINDY DENNYS CUZCO OCHOA.

DIRECTOR DE TRABAJO DE TITULACION:

ING. ROBERTO CRESPO, MSIG

GUAYAQUIL, JUNIO 2022



**FACULTAD: CIENCIAS MATEMATICAS Y FISICAS
MAESTRIA INGENIERIA EN SOFTWARE**

Título del Trabajo de Titulación: DISEÑO DE UN MODELO PREDICTIVO COMO APOYO DE TOMA DE DECISIONES EN LA RELACIÓN ENTRE SOÑADOR Y VOLUNTARIO DE LA FUNDACIÓN “AMBULANCIA DESEO”

Autora: Cindy Dennys Cuzco Ochoa

Tutor: Ing. Roberto Crespo, Msig

RESUMEN

La importancia de brindar apoyo a personas con enfermedades catastróficas es notaría cuando se conoce a profundidad de sus necesidades. Existiendo fundaciones como "Ambulancia Deseo" quien conoce el valor de brindar emociones positivas, generándolas, mediante el cumplimiento de sueños vivenciales a personas con enfermedades catastróficas, es ahí donde se logra comprender el valor significativo de ayudar a las fundaciones. Entre los múltiples desafíos que enfrenta la fundación “Ambulancia Deseo” se destaca la necesidad de relacionar adecuadamente el soñador y voluntario a la hora de cumplir un sueño. Así, este proyecto tiene como objetivo, en primer lugar, distinguir que factores influyen en la relación entre soñador y voluntario para el cumplimiento de un sueño. En segundo lugar, analizar las diferentes metodologías que son utilizadas en minería de datos, se dispuso que la metodología CRISP-DM proporciona las fases necesarias para el desarrollo de este trabajo. En las diferentes fases que brinda la metodología se analizaron los datos y paralelamente los métodos de machine Learning a utilizar, definiendo arboles de decisiones como el algoritmo seleccionado para el desarrollo de la propuesta por su simplicidad y eficacia.

Palabras clave: CATASTRÓFICA, POSITIVAS, FUNDACIONES, ALGORITMO, MINERÍA.



**FACULTAD: CIENCIAS MATEMATICAS Y FISICAS
MAESTRIA INGENIERIA EN SOFTWARE**

Title Of Degree Work Carried Out: DISEÑO DE UN MODELO PREDICTIVO COMO APOYO DE TOMA DE DECISIONES EN LA RELACIÓN ENTRE SOÑADOR Y VOLUNTARIO DE LA FUNDACIÓN “AMBULANCIA DESEO”

Author: Cindy Dennys Cuzco Ochoa

Advisor: Ing. Roberto Crespo, Msig

ABSTRACT

The importance of providing support to people with catastrophic illnesses is evident when their needs are thoroughly understood. There are foundations such as "Desire Ambulance" who know the value of providing positive emotions, generating them, through the fulfillment of experiential dreams to people with catastrophic illnesses, that is where the significant value of helping foundations is understood. Among the multiple challenges faced by the "Desire Ambulance" foundation, the need to adequately relate the dreamer and the volunteer when it comes to fulfilling a dream stands out. Thus, this project aims, first of all, to distinguish which factors influence the relationship between dreamer and volunteer for the fulfillment of a dream. Second, to analyze the different methodologies that are used in data mining, it was established that the CRISP-DM methodology provides the necessary phases for the development of this work. In the different phases provided by the methodology, the data and, in parallel, the machine learning methods to be used were analyzed, defining decision trees as the algorithm selected for the development of the proposal due to its simplicity and effectiveness.

KEY WORDS: CATASTROPHIC, POSITIVE, FOUNDATIONS, ALGORITHM, MINING.



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT

Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE GRADUACIÓN

TÍTULO Y SUBTÍTULO:	Diseño de un modelo predictivo como apoyo de toma de decisiones en la relación entre soñador y voluntario de la Fundación "Ambulancia Deseo"		
AUTOR(ES):	Cindy Dennys Cuzco Ochoa		
REVISOR(ES)/TUTOR(ES)	Ing. Roberto Crespo, MSIG		
INSTITUCIÓN:	Universidad de Guayaquil		
UNIDAD/FACULTAD:	Ciencias Matemáticas y Físicas		
MAESTRÍA/ESPECIALIDAD:	Maestría Ingeniería de Software		
GRADO OBTENIDO:	Ingeniera en Sistemas Computacionales.		
FECHA DE PUBLICACIÓN:	Junio 2022	No. DE PÁGINAS:	107
ÁREAS TEMÁTICAS:	Modelo Predictivo para personas con enfermedades catastróficas.		
PALABRAS CLAVES/KEYWORDS:	CATASTRÓFICA-POSITIVAS-FUNDACIONES-ALGORITMO-MINERÍA.		
RESUMEN/ABSTRACT:	<p>La importancia de brindar apoyo a personas con enfermedades catastróficas es notaría cuando se conoce a profundidad de sus necesidades. Existiendo fundaciones como "Ambulancia Deseo" quien conoce el valor de brindar emociones positivas, generándolas, mediante el cumplimiento de sueños vivenciales a personas con enfermedades catastróficas, es ahí donde se logra comprender el valor significativo de ayudar a las fundaciones. Entre los múltiples desafíos que enfrenta la fundación "Ambulancia Deseo" se destaca la necesidad de relacionar adecuadamente el soñador y voluntario a la hora de cumplir un sueño. Así, este proyecto tiene como objetivo, en primer lugar, distinguir que factores influyen en la relación entre soñador y voluntario para el cumplimiento de un sueño. En segundo lugar, analizar las diferentes metodologías que son utilizadas en minería de datos, se dispuso que la metodología CRISP-DM proporciona las fases necesarias para el desarrollo de este trabajo. En las diferentes fases que brinda la metodología se analizaron los datos y paralelamente los métodos de machine Learning a utilizar, definiendo arboles de decisiones como el algoritmo seleccionado para el desarrollo de la propuesta por su simplicidad y eficacia</p>		
ADJUNTO PDF:	SI X	NO	
CONTACTO CON AUTOR/ES:	Teléfono: +593 997723277	E-mail: dennys.cuzco@gmail.com	
CONTACTO CON LA INSTITUCIÓN:	Nombre: universidad de Guayaquil		
	Teléfono: 2-283348		
	E-mail: fmatematicas@ug.edu.ec		



REGLAMENTO PARA EL PROCESO DE TITULACIÓN EN POSGRADO

CERTIFICADO PORCENTAJE DE SIMILITUD

Universidad de Guayaquil Facultad de Ciencias Matemáticas y Físicas Maestría en Ingeniería en Software

Habiendo sido nombrado Ing. Roberto Crespo Mendoza, MSIG., tutor del trabajo de titulación, certifico que el presente trabajo de titulación ha sido elaborado por la Ing. Cindy Dennys Cuzco Ochoa, con mi respectiva supervisión como requerimiento parcial para la obtención del título de Magister en Ingeniería en Software.

Se informa que el trabajo de titulación: “Diseño de un modelo predictivo como apoyo de toma de decisiones en la relación entre soñador y voluntario de la fundación “Ambulancia Deseo””, ha sido orientado durante todo el periodo de ejecución en el programa antiplagio Urkund quedando el 5 % de coincidencia.

Categoría	Enlace/nombre de archivo
	https://es.wikipedia.org/wiki/Fundaci%C3%B3n_Ambulancia_del_Deseo
	Fundació per a la Universitat Oberta de Catalunya / (null)
	UNIVERSIDAD TECNICA DE AMBATO / (null)
	Fundació per a la Universitat Oberta de Catalunya / (null)
	UNIVERSIDAD TECNICA DE AMBATO / (null)

<https://secure.urkund.com/view/132941590-897814-862136>



Firmado digitalmente por:
**ROBERTO CARLOS
CRESCO MENDOZA**

Roberto Crespo Mendoza.
CI: 0914642459
Director Trabajo de Titulación.



ANEXO V.- CERTIFICADO DEL DIRECTOR DE TRABAJO DE TITULACIÓN

Guayaquil, 30 de mayo del 2022

Sr. Ing. Douglas Iturburu MSc.

Decano de la Facultad de Ciencias Matemáticas y Físicas

Universidad de Guayaquil

De mis consideraciones:

Envío a Ud. el Informe correspondiente a la tutoría realizada al Trabajo de Titulación denominada **“Diseño de un modelo predictivo como apoyo de toma de decisiones en la relación entre soñador y voluntario de la fundación “Ambulancia Deseo”** del estudiante Ing. Cindy Dennys Cuzco Ochoa, de la maestría en Ingeniería en Software, indicando que ha cumplido con todos los parámetros establecidos en la normativa vigente:

- El trabajo es el resultado de una investigación.
- El estudiante demuestra conocimiento profesional integral.
- El trabajo presenta una propuesta en el área de conocimiento (opcional según la modalidad)
- El nivel de argumentación es coherente con el campo de conocimiento.

Adicionalmente, se adjunta el certificado de porcentaje de similitud (firmada), la versión aprobada del trabajo de titulación, el registro de tutorías y la rúbrica de evaluación del trabajo de titulación.

Dando por concluida esta tutoría de trabajo de titulación, CERTIFICO, para los fines pertinentes, que la estudiante está apta para continuar con el proceso.

Atentamente,



Firmado digitalmente por:
**ROBERTO CARLOS
CRESPO MENDOZA**

Ing. Roberto Crespo Mendoza. MSIG.
Ci: 0914642459



ANEXO X. – DECLARACIÓN DE AUTORÍA Y DE AUTORIZACIÓN DE LICENCIA GRATUITA INTRANSFERIBLE Y NO EXCLUSIVA PARA EL USO NO COMERCIAL DE LA OBRA CON FINES NOACADÉMICOS

FACULTAD DE CIENCIAS MATEMÁTICAS Y FÍSICAS PROGRAMA
MAESTRÍA EN INGENIERÍA EN SOFTWARE

LICENCIA GRATUITA INTRANSFERIBLE Y NO COMERCIAL DE LA OBRA
CON FINES NO ACADÉMICOS

Yo / Nosotros, **CINDY DENNYS CUZCO OCHOA**, (nombre (s) del/ de los estudiantes), con C.I. No: **0927789958**, certifico/amos que los contenidos desarrollados en este trabajo de titulación, cuyo título es **Diseño de un modelo predictivo como apoyo de toma de decisiones en la relación entre soñador y voluntario de la Fundación “Ambulancia Deseo”** son de mi/nuestra absoluta propiedad y responsabilidad, en conformidad al Artículo 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN*, autorizo/amos la utilización de una licencia gratuita intransferible, para el uso no comercial de la presente obra a favor de la Universidad de Guayaquil.

Atentamente,



Ingeniera
Cindy Dennys
Cuzco Ochoa

CINDY DENNYS CUZCO OCHOA
CI No.: 0927789958

ÍNDICE GENERAL

RESUMEN	ii
ABSTRACT	iii
CERTIFICADO PORCENTAJE DE SIMILITUD	v
INDICE DE ILUSTRACIONES	xi
INDICE DE TABLAS	xiii

CAPITULO I

Introducción

1.1	Justificación	3
1.2	Planteamiento del problema	4
1.3	Objetivos del problema	7
1.3.1	Objetivo General	7
1.3.2	Objetivos Específicos	7
1.4	Alcance del problema	8
1.5	Estructura de la memoria	9

CAPITULO II

Marco Teórico

2.1	Antecedentes de estudio	11
2.2	Marco conceptual	18
2.2.1	Inteligencia Artificial	18
2.2.2	Machine Learning	20
2.2.3	Algoritmo de Aprendizaje Supervisado	21
2.2.4	Algoritmo de Aprendizaje NO Supervisado	22
2.2.5	Algoritmos de Clasificación	23

2.2.6 Árboles de Decisión	25
2.2.7 Algoritmo de ID3	26
2.2.8 Algoritmo ID3 - Entropía de Shannon	27
2.2.9 Bosques Aleatorios - Random Forest	30
2.2.10 Regresión Logística	31
2.2.11 KNN – K vecinos cercanos	31
2.2.12 Naive Bayes	32
2.2.13 SVM - Máquinas de Vectores de Soporte	33
2.3 Marco legal	35
2.4 Marco institucional	40

CAPITULO III

Metodología

3.1 Modalidad de la investigación	44
3.2 Tipo de investigación	45
3.3 Metodología del trabajo	45
3.4 Desarrollo de la Metodología Elegida (CRISP-DM)	50
3.4.1 Fase 1: Comprensión del Negocio	50
3.4.2 Fase 2: Comprensión de los Datos	55
3.4.3 Fase 3: Preparación de los Datos	64
3.4.4 Fase 4: Modelado	83
3.4.5 Fase 5: Evaluación	96

Conclusiones y Recomendaciones

Conclusiones

Recomendaciones

Anexos

Bibliografía

Índice de Ilustraciones

Ilustración 1: Aprendizaje Supervisado y NO Supervisado.....	20
Ilustración 2: Aprendizaje Supervisado	22
Ilustración 3: Fases de Algoritmo Clasificación Supervisado.....	24
Ilustración 4: Ejemplo de Estructura de Arboles.....	26
Ilustración 5: Ejemplo de Entropía y Ganancia.....	29
Ilustración 6: Algoritmo - Bosques aleatorios	30
Ilustración 7: Algoritmo - KNN.....	32
Ilustración 8: Algoritmo - Navie Bayes.....	33
Ilustración 9: Algoritmo - SVM.....	34
Ilustración 10: Logo de la Fundación en Ecuador	42
Ilustración 11: Equipo de Apoyo.....	43
Ilustración 12: Comparativa de Metodologías	46
Ilustración 13: Encuesta de tipos de metodologías	47
Ilustración 14: Etapas de Metodología CRISP-DM.....	48
Ilustración 15: Datos de Soñador y Voluntario	64
Ilustración 16: Estadístico Descriptivo.....	67
Ilustración 17: Correlación de Variables.....	68
Ilustración 18: Análisis del Género - Soñador	72
Ilustración 19: Análisis de la Categoría - Soñador.....	73
Ilustración 20: Edades clasificado según Pareto	75
Ilustración 21: Análisis de Edades con Rango Establecido - Soñador.....	76
Ilustración 22: Análisis de Ciudad - Voluntarios	78
Ilustración 23: Análisis del Género - Voluntarios	79
Ilustración 24: Análisis de Edad – Voluntarios.....	81
Ilustración 25: Diagrama de Flujo.....	85
Ilustración 26: Nuevo Proyecto - Collab	87
Ilustración 27: Levantamiento de Unidad	87
Ilustración 28: Importar Datos	87
Ilustración 29: Previsualización de Datos cargados	88
Ilustración 30: Cambio de Variable	89
Ilustración 31: Variables Ajustadas	89

Ilustración 32: Distinción de Variables.....	90
Ilustración 33: Entrenamiento	91
Ilustración 34: Algoritmo de Árbol de Decisión	91
Ilustración 35: Ajuste del Modelo	92
Ilustración 36: Visualizar Modelo – Árbol de Decisión	92
Ilustración 37: Modelo ID3- Entropy	94
Ilustración 38: Análisis de los Algoritmos	96
Ilustración 39: Resultado de los Algoritmos	97
Ilustración 40: Matriz de Confusión	99
Ilustración 41: Matriz Confusión ID3 - Entropía	100
Ilustración 42: Matriz Confusión ID3 - Gini	101
Ilustración 43: Algoritmo - Random Forest	103
Ilustración 44: Visualizar Modelo - Random Forest	103
Ilustración 45: Comparativa	105

Índice de tablas

Tabla 1: Recursos Usados	52
Tabla 2: Restricciones.....	52
Tabla 3: Registro de Soñadores	56
Tabla 4: Registro de Voluntarios	57
Tabla 5: Tabla con Campos Originales de los Soñadores	61
Tabla 6: Tabla con Campos Originales de los Voluntarios	63
Tabla 7: Análisis del Género - Soñador	71
Tabla 8: Análisis de Categoría - Soñador	73
Tabla 9: Análisis de Edades - Soñador.....	74
Tabla 10: Rango de Edades	75
Tabla 11: Análisis de Edades con Rango - Soñador.....	76
Tabla 12: Análisis de Ciudad - Voluntario	77
Tabla 13: Análisis del Género - Voluntarios.....	79
Tabla 14: Análisis de Edad - Voluntarios	80

CAPÍTULO I

Introducción

En latino América existe un alto grado de presencia de personas con enfermedades Catastrófica. Siendo las enfermedades catastróficas aquellas que perjudican significativamente la salud en las personas. En estas enfermedades la característica principal, es el alto grado de complejidad para ser tratadas, llegando al punto de ser amenazantes para la vida de quien la padece.

Existen estudios que abalan los múltiples beneficios que aportan la generación de emociones positivas en las personas, siendo esta la razón principal para ser aplicado en diversas áreas con el propósito de ayudar de manera significativa a personas con enfermedades catastróficas.

En Ecuador existen fundaciones que priorizan el brindar apoyo desinteresado a personas con enfermedades catastróficas, puesto que, aquellas personas día a día están en la constante lucha de supervivencia por superar múltiples obstáculos, por ejemplo, salud, social, económico.

Fundación “Ambulancia Deseo” tiene como objetivo cumplir sueños vivenciales a personas con enfermedades catastróficas, pero, actualmente no posee herramientas tecnológicas que le brinde apoyo en el proceso de relación entre soñadores y voluntarios.

Encontrar una relación entre soñador y voluntario tiene una gran importancia puesto que, al existir una gran afinidad en ambos, el mismo influirá significativamente en el impacto emocional final por el cumplimiento del sueño.

El presente trabajo se enfoca en la elaboración de un modelo predictivo que brinde apoyo en la toma de decisión en la relación entre voluntario y soñador. Se analizará que factores influyen para encontrar una relación entre soñador y voluntario para así apoyar alcanzar del impacto emocional durante el cumplimiento de un sueño.

La información que se utilizará para la elaboración del modelo es la que actualmente maneja la fundación “Ambulancia Deseo”. La información será procesada, depurada, corregida, para posteriormente ser analizada. Después, en la etapa

correspondiente, se elegirá los campos y atributos que contribuyan significativamente en el análisis mediante el uso de árboles de decisiones donde se determinará su clasificación según la información procesada.

1.1 Justificación

En Ecuador existen pocas fundaciones que destinan sus actividades benéficas para el bienestar de las personas que sufren de enfermedades catastróficas.

La Fundación “Ambulancia Deseo” siendo su objetivo principal el cumplir sueños vivenciales a personas con enfermedades catastróficas, y al ser una fundación sin fines de lucro, cumple sus funciones, mediante el apoyo de voluntarios, profesionales y patrocinadores que gracias a su desinterés y buen corazón aporta con su granito de arena en las diversas actividades que tiene como finalidad cumplir el objetivo principal de la fundación.

En la actualidad, la fundación trabaja sus operaciones de forma manual, se busca con el desarrollo de este proyecto, aportar con una herramienta que apoye en la relación entre soñador y voluntario dentro de la fundación.

La presente investigación se asentará en identificar qué factores influyen en la relación entre soñador y voluntario, determinar las herramientas que se requieren para poder predecir una adecuada relación, la información que produzcan las herramientas ayudará al entrenamiento del modelo de datos y así lograr crear predicciones en la relación entre soñador y voluntario.

Durante el proceso de creación del modelo predictivo se comprenderá las necesidades de la fundación que surgen durante el proceso de relación entre soñador y voluntario. Finalmente, el modelo apoyará para que la fundación incremente su productividad en la logística para el cumplimiento de los sueños.

1.2 Planteamiento del Problema

El proceso de cumplir un sueño involucra diversas fases, una de ellas es, decidir cuál voluntario es el adecuado para que el soñador experimente una satisfacción plena por la compañía asignada durante el cumplimiento de su sueño.

Es de alta importancia elegir cual voluntario es el ideal para el soñador, dado que, el cumplimiento de un sueño involucra diversos aspectos, por ejemplo, no es lo mismo que un soñador de 10 años con cáncer y dificultad de movilidad sea acompañado a la playa por un voluntario femenino, a comparación de un voluntario masculino, ya que, el voluntario masculino puede apoyar en el traslado del niño por la complejidad de su movilidad a comparación de la voluntaria femenina.

La relación que se crea entre el soñador y voluntario, durante la ejecución de un sueño, es de vital importancia, puede este marcar una transcendencia en el éxito del cumplimiento de un sueño, por ejemplo, al crearse un vínculo emocional profundo después de cumplir un sueño, el voluntario puede decidir si acompaña al soñador en las diversas etapas que involucra la lucha contra su enfermedad catastrófica, la misma, beneficiaría adicionalmente el estado de salud del soñador gracias al lazo emocional que se generó en una previa experiencia compartida.

Por otro parte, aun recibiendo diferentes tipos de apoyo, fundación “Ambulancia Deseo” mantiene una deficiencia de

herramientas tecnológicas que estén ajustadas a sus necesidades.

En pleno siglo XX, fundación “Ambulancia Deseo” constantemente se encuentra en la búsqueda de herramientas que apoyen en los diversos procesos que se ejecuta para el cumplimiento de sus objetivos.

Dentro de la fundación “Ambulancia Deseo”, el proceso de relación entre soñador y voluntario para el cumplimiento de un sueño, se realiza de forma manual, a continuación, detallamos algunas dificultades:

Soñador

Identificar quienes están aún en la espera del cumplimiento de su sueño.

Clasificarlos por perfiles, por ejemplo, menor de edad o adulto.

Clasificar los sueños por complejidad o factores externos.

Voluntario

Clasificarlos por perfiles, por ejemplo, femenino o masculino.

Recordar que voluntarios están aptos para el cumplimiento de un sueño.

Identificar que voluntarios están activos dentro de la fundación.

Estas dificultades no son únicamente las que tiene que enfrentar la fundación durante el cumplimiento de un sueño, pero si tiene un alto grado de conexión frente al objetivo del desarrollo de este proyecto.

1.3 Objetivos del Problema

1.3.1 Objetivo General.

Proponer un modelo predictivo de aprendizaje automático supervisado como apoyo de toma de decisiones en la relación entre soñador y voluntario en la fundación “Ambulancia Deseo” de la ciudad de Guayaquil.

1.3.2 Objetivos Específicos.

Identificar los factores que influyen en la relación entre soñador y voluntario para instituir el orden de importancia aplicable en la fundación “Ambulancia Deseo”

Analizar diversas metodologías que permita gestionar los factores que influyen en la relación entre soñador y voluntario de la fundación “Ambulancia Deseo”

Diseñar un modelo predictivo de toma de decisiones que brinde apoyo en el vínculo entre soñador y voluntario de la fundación “Ambulancia Deseo”.

1.4 Alcance del Problema

En esta sección se indicará la orientación del modelo predictivo, así como también las implicaciones de la misma como se detalla a continuación:

El modelo predictivo al implementarse en una herramienta tecnológica complementará el apoyo en el proceso de cumplimiento de un sueño, más no constituye como un sustituto completo para la toma de decisiones.

El modelo a predictivo se enfoca a las personas activas de la fundación, es decir, soñadores y voluntarios.

El modelo predictivo quedará listo para ser implementado según lo considere conveniente la fundación “Ambulancia Deseo”.

Los datos a tratarse se procesarán según la estructura del modelo diseñado.

Los datos a analizar y que se utilizaran para educar el modelo predictivo, son proveniente desde la base actual que maneja la fundación "Ambulancia Deseo".

1.5 Estructura de la memoria

En el presente trabajo, su estructura comienza desde los conceptos generales hasta lo específico, enfocado a la relación entre soñador y voluntario de la fundación "Ambulancia Deseo".

En el Capítulo I, se puntualiza en contexto la problemática que se encontró dentro de la fundación, específicamente en el proceso de relación entre soñador y voluntario de la fundación "Ambulancia Deseo" de la ciudad de Guayaquil. También, los objetivos, generales y específicos.

En el Capítulo II, se expone los estudios que son similares al presente trabajo. Las referencias utilizadas colaboran el comprender la necesidad del desarrollo y apoyo a temas que abarca el actual trabajo.

En el Capítulo III, comprende la explicación de las metodologías existentes y la que se consideró adecuada para desarrollo de este proyecto.

En el Capítulo IV, se presenta las conclusiones finales y se detallan los puntos relevantes definidos por el desarrollo del proyecto y finalmente encontraremos la bibliografía utilizada y anexos como aporte por el tema propuesto.

CAPÍTULO II

Marco teórico

En esta sesión del trabajo de investigación se explican los antecedentes del estudio y conceptualizaciones que utilizaremos para la generación de un modelo predictivo de aprendizaje automático supervisado para el apoyo en la toma de decisiones en la relación entre soñador y voluntario en la fundación “Ambulancia Deseo” de la ciudad de Guayaquil.

2.1 Antecedentes de Estudio

Al tratar temas como los que se abordarán en esta investigación, se debe tener un conocimiento básico del mismo, por eso se explicará lo que involucra las enfermedades catastróficas y sus casos opuestos, y así poder comprender los temas que lo involucra, acompañado de su problemática.

Enfermedades Catastróficas

Podemos decir que las enfermedades catastróficas son “aquellas que deterioran la salud de las personas, se caracterizan por su alto grado de complejidad, son agudas,

prolongadas, amenazantes para la vida pues en su mayoría son letales” (1260-TESIS-ALVAREZ Y MERCHAN.pdf, s. f., p. 33)

Considerando el aporte del párrafo anterior, se puede concluir que, se considera a las enfermedades catastróficas como casos muy complejos o casos diagnosticado en desahucio, necesitando permanente atención médica, preocupación emocional no solo para el paciente sino también para la familia.

En Ecuador, según registros estadísticos de la Alianza por el Derecho y la Protección de la Salud, el país registra 300.000 casos de personas con enfermedades catastróficas. Las cifras reflejan que 1 de cada 1.000 ecuatorianos padecen de alguna enfermedad rara o catastrófica. (*300 mil casos de enfermedades raras se registran en Ecuador*, s. f.). estas cifras se dividen entre hombres y mujeres. Adicionalmente 80 de cada 100 personas con alguna enfermedad catastrófica son niños, es decir, desde corta edad las personas con enfermedades catastróficas inician el proceso de supervivencia y lucha por conservar su vida.

Analizando el factor económico, para las personas con enfermedades catastróficas, la importancia se agudiza, ya que, se necesitan de muchos recursos para un tratamiento completo. Convirtiéndose en un peso muy complejo de superar y mantener tanto a la persona que padece una enfermedad como para la familia.

Aún con los aspectos previos mencionados, es posible enfrentar de una manera accesible o parcial las enfermedades, considerando que la salud es algo más que la ausencia de enfermedad y que las emociones positivas son algo más que la ausencia de emociones negativas.

Es posible plantear que las emociones positivas puedan prevenir enfermedades, o reducir la intensidad y duración de las mismas, así como también pueden ser útiles para poder alcanzar niveles elevados de bienestar subjetivo.

Felicidad y Bienestar

La felicidad, muy lejos de ser un sentimiento objetivo, dependiendo de las circunstancias, muchas veces es un reflejo de cómo vemos las cosas, es decir, de nuestra forma de pensar. Si bien es cierto que el dinero, la salud o el estatus pueden aumentar nuestro nivel de felicidad, sin embargo, una parte fundamental de nuestra felicidad depende de la cultura o idiosincrasia en la que vivimos.

Bienestar Emocional

Es apropiado dar a conocer que, el bienestar emocional se comprende como un estado subjetivo de los seres humanos, en el que existe el predominio de sensaciones optimistas o positivas, las cuales establecen satisfacción en uno mismo.

El bienestar emocional se vincula con el afrontamiento de dificultades y a la disminución de los factores de riesgo que amenazan la salud mental del sujeto. No obstante, el mantenimiento de este estado dependerá de diversos factores entre los que destaca el equilibrio entre las capacidades adaptativas de la persona y los factores estresores (Paola, s. f.).

Lo que nos ayuda a dar hincapié en la importancia de generar un bienestar emocional positivo en las personas y no sólo en aquellas que están frágiles ante dificultades de salud.

El bienestar emocional brinda la capacidad para razonar sobre las emociones y hacer uso de ellas para mejorar el pensamiento. Esto incluye las habilidades para percibir emociones con precisión, para acceder y generar emociones como ayuda al pensamiento, para comprender las emociones y el conocimiento emocional, y para regular reflexivamente las emociones de modo que promuevan un desarrollo emocional e intelectual.(Paola, s. f., p. 12)

Salud Mental

La Organización Mundial de la Salud define que la dimensión positiva de la salud mental ha sido subrayada en la definición de salud de la OMS, tal cual consta en la constitución misma: «La salud es un estado de completo bienestar físico, mental y social y no solamente la ausencia de afecciones o enfermedades». Los conceptos de salud mental incluyen bienestar subjetivo, autonomía, competencia, dependencia intergeneracional y

reconocimiento de la habilidad de realizarse intelectual y emocionalmente. (Dialnet-InteligenciaEmocionalYBienestarII-655308.pdf, s. f.)

Inteligencia Emocional

En la inteligencia emocional (IE) nos podemos referir a la capacidad de comprender, usar y manejar nuestras propias emociones de manera que reduzcan el estrés, nos ayuden a comunicarnos de manera efectiva y empatizar con los demás, superar desafíos y reducir conflictos.

Grupo de Personas Vulnerables

Durante la última década la atención a grupos vulnerables, también conocidos como grupos sociales en condiciones de desventaja, ocupa un espacio creciente en las agendas legislativas de las políticas públicas, con especial atención a los procesos de vulnerabilidad social de las familias, grupos y personas.(Grupos Vulnerables, s. f.) Siendo así, y desde hace mucho tiempo, se ha evidenciado desde diferentes perspectivas ya sea social, gubernamental, etc. como las personas vulnerables han sido olvidadas y hasta discriminadas. En la actualidad no hay una excepción frente a la lucha constante por

mantener una inclusión equitativa, donde podemos destacar, que se ha logrado una progresiva inclusión dentro de la sociedad, lo que nos demuestra la gran importancia de seguir luchando al cumplimiento de todas aquellas políticas creadas en beneficio de los grupos vulnerables.

Apoyo a Personas Vulnerables

La "evidente" necesidad de "humanizar" las actuaciones de los profesionales, así como de "personalizar" los casos clínicos, enfocándose en tres dimensiones: biológica, psicológica y social. La práctica diaria de la medicina actual resulta realmente incompleta. Nuestro sistema sanitario suele centrarse en un cuidado tipo que aborda la parte biológica de la enfermedad de forma altamente efectiva, pero descuida y relega a un segundo plano el factor psicológico y emocional de cada paciente. La fundación brinda a los enfermos la posibilidad de cerrar sus ciclos y paliar esas carencias atendiendo sus aspectos más íntimos, personales, familiares. Aunando esfuerzos entre ambas partes [Fundación y Sanidad pública], podremos hablar de un cuidado integral. (Gassó, 2019)

2.2 Marco conceptual

En esta sesión del trabajo se detallará en realidad una investigación bibliográfica que habla de las variables que se estudiarán en la investigación.

2.2.1 Inteligencia Artificial

Con la Inteligencia Artificial (IA) es posible indicar que es el punto de partida para la imitación de procesos de la inteligencia humana para así crear y aplicar algoritmos en entornos dinámicos de computación. Es decir, con la Inteligencia Artificial (IA) se trata de hacer que las computadoras u ordenadores piensen y actúen como los seres humanos.

En el ámbito laboral, la tecnología de IA mejora el rendimiento y la productividad de la empresa mediante la automatización de procesos o tareas que antes requerían esfuerzo humano. La IA también puede dar sentido a los datos a una escala que ningún humano jamás podría. Esta capacidad puede generar importantes ventajas. (*¿Qué es la inteligencia artificial (IA)?*, s. f.). Por tal, el uso y aplicación de la IA puede aportar productivamente de manera significativa en los diversos procesos donde se decida implementarla.

A continuación, las categorías que divide la IA:

Sistemas que piensan como humanos: Estos sistemas intentan emular la forma de pensar del ser humano, por ejemplo, redes neuronales artificiales, agentes inteligentes.

Sistemas que actúan como humanos: Estos sistemas intentan actuar de igual forma que el ser humano, por ejemplo, robótica.

Sistemas que piensan racionalmente: Se refiere, al uso de la lógica, aquí se intenta imitar el pensamiento lógico, por ejemplo, los sistemas expertos.

Sistemas que actúan racionalmente: Estos sistemas intentan emular la forma racional del ser humano. Por ejemplo, agentes inteligentes.

En cada categoría existente hay diferentes mecanismos, los mismos permiten resoluciones de las problemáticas en diferentes disciplinas. (*Categorías de la IA - Inteligencia Artificial*, s. f.)

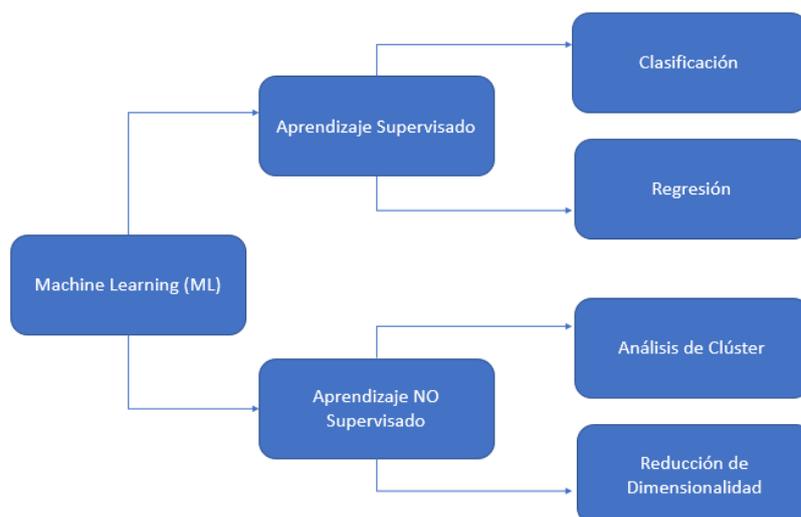
2.2.2 Machine Learning

Dentro de la IA está el Machine Learning (ML) siendo esta una de sus áreas más importantes. El Machine Learning son secuencias de instrucciones para conseguir resultados concretos y así poseer la habilidad de aprender sin la necesidad de una programación explícita.

El funcionamiento del ML se centra en entrenar un modelo, dar inicio con los datos de entrada, aprender de los mismos, y así, como obtener los resultados deseados, por tal, se da uso para tener predicciones o propuestas para la toma de decisiones.

Dentro del Machine Learning (ML) encontramos los siguientes algoritmos, que se aprecia en la ilustración 1.

Ilustración 1 Aprendizaje Supervisado y NO Supervisado



Fuente: (Calvo, 2017)
Autor: Diego Calvo

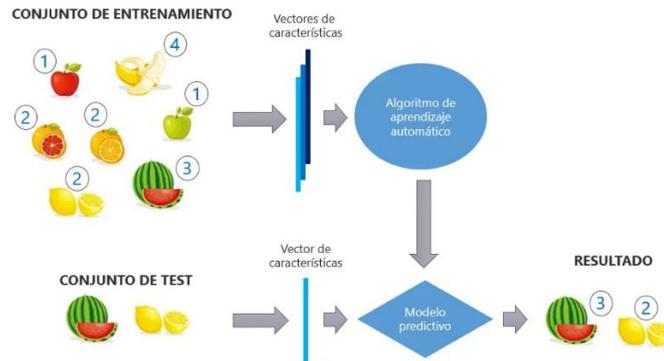
En la ilustración 1 muestra la clasificación dentro del Machine Learning (ML), los conceptos de cada clasificación se los explica de forma detallada durante la exploración de este trabajo.

2.2.3 Algoritmo de Aprendizaje Supervisado

En este tipo de aprendizaje parten de la generación de una función principal donde se asigna las entradas y las salidas deseadas. Para comprender el funcionamiento del aprendizaje supervisado se acudirá a un ejemplo donde tenemos un conjunto de datos etiquetados a priori: 1 manzanas, 2 cítricos, 3 sandias y 4 plátanos.

A este conjunto de datos le extrañamos sus características en vectores y con esto entrenamos el algoritmo de aprendizaje automático. El algoritmo nos genera un modelo el que podemos proporcionarle nuevos conjuntos de datos que nos clasifique los nuevos elementos que nunca ha visto con lo que ha aprendido en la fase de entrenamiento.(Calvo, 2019). A continuación, en la Ilustración 2 se hace referencia visual del ejemplo previamente mencionado.

Ilustración 2 Aprendizaje Supervisado



Fuente: (Calvo, 2019)
Autor: Diego Calvo

2.2.4 Algoritmo de Aprendizaje NO Supervisado

En este tipo de aprendizaje parten de la deducción de un conjunto de datos sin clasificación, es decir, su aprendizaje se realiza sin la necesidad de una pauta prevista.

Ejemplo: Imaginemos que una niña tiene como mascota a un gato. La niña identifica a su gato. Semanas después, un familiar trae un gato e intenta que juegue la niña con el gato. La niña quien no ha visto antes al gato del familiar, puede relacionar a su gato con la mascota del familiar pues tiene muchas características similares (orejas, forma, cola). La niña identificará a la mascota como un gato. Esto lo definimos como aprendizaje no supervisado, no se le enseña, su aprendizaje parte de los datos.

A diferencia del Aprendizaje Supervisado, los métodos de Aprendizaje no Supervisado no se pueden aplicar directamente a un problema de regresión o clasificación porque no tiene idea de cuáles pueden ser los valores de los datos de salida, lo que hace imposible que entrene el algoritmo de la forma en que lo haría normalmente.(Gonzalez, 2020).

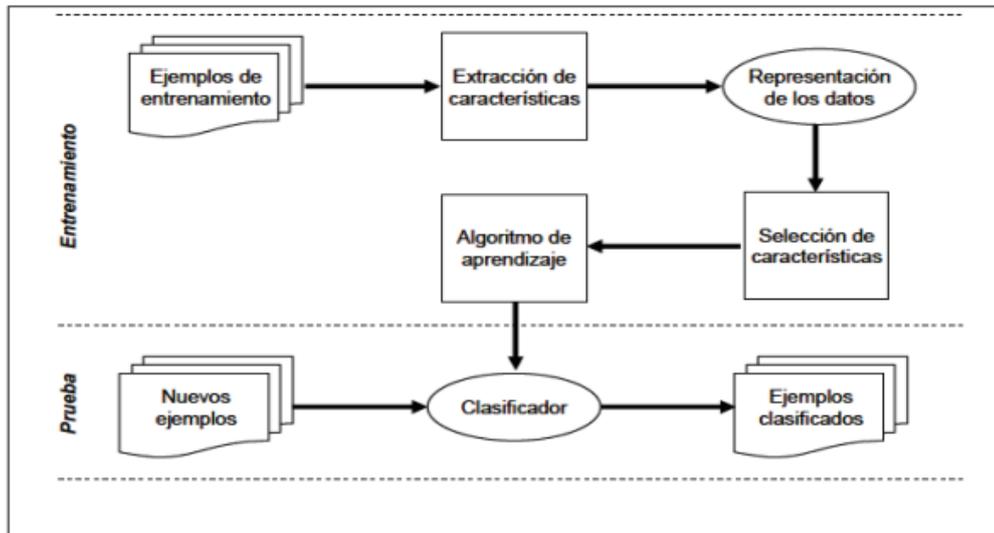
Después de distinguir el Aprendizaje Supervisado y NO supervisado, analizaremos los algoritmos del Aprendizaje Supervisado.

2.2.5 Algoritmos de Clasificación

Clasificación como un método de aprendizaje supervisado que tiene como objetivo clasificar en grupos en función de datos históricos etiquetados. Además, los algoritmos de clasificación son utilizados cuando lo que se busca son resultados que estén dentro de un conjunto finito de posibles resultados.

En la Ilustración 3 se aprecia como trabaja el algoritmo de Clasificación, su funcionalidad parte de la división de la data, siendo un porcentaje para el entrenamiento y otra para la prueba.

Ilustración 3 Fases de Algoritmo Clasificación Supervisado



Fuente: (Vera, s. f.)
Autor: Sebastián Ariel

Existen dos tipos principales de clasificaciones:

Binaria: Si el objetivo de las clases del modelo a entrenar es, verdadero o falso. El modelo a entrenarse es de tipo de clasificación binaria. Por ejemplo, decir si una fruta es cítrica o no, si un correo es spam o no.

Multiclase: Si el objetivo es la asignación de múltiples categorías. El modelo a entrenarse es de tipo de clasificación Multiclases. Por ejemplo, Reconocer los caracteres de una escritura manual (aquí las clases van de 0 a 9)

Tipos de algoritmos de Clasificación:

Arboles de Decisión (Algoritmo ID3).

Random Forest.

Regresión Logística

KNN

Naive Bayes.

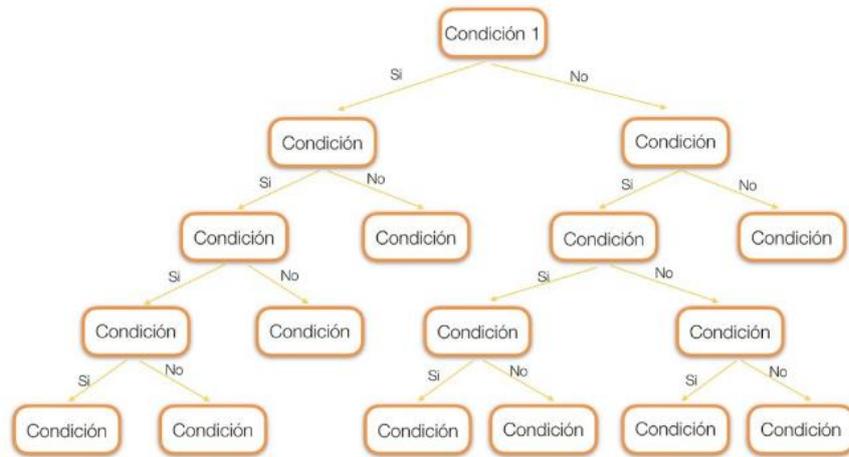
SVM

2.2.6 Arboles de Decisión

Dentro del Machine Learning y aprendizaje supervisado, tenemos el algoritmo de árboles de decisiones, siendo este uno de los más utilizados, su aplicabilidad principalmente se debe a que brinda una fácil interpretación y alta precisión.

Un árbol de decisión tiene una estructura similar a un diagrama de flujo donde un nodo interno representa una característica o atributo, la rama representa una regla de decisión y cada nodo u hoja representa el resultado. El nodo superior de un árbol de decisión se conoce como nodo raíz. («Árboles de Decisión Clasificación – Teoría», 2019). En la ilustración 4 se muestra la estructura que forma los árboles de decisiones.

Ilustración 4 Ejemplo de Estructura de Arboles



**Fuente: («Árboles de Decisión Clasificación – Teoría», 2019)
Autor: Aprende IA**

2.2.7 Algoritmo de ID3

Dentro los árboles de decisiones tenemos que la construcción del árbol de decisión no es única, y si aplicamos una estrategia u otra a la hora de decidir en qué orden se hacen las preguntas sobre los atributos podemos encontrar árboles muy dispares.

De entre todos los posibles árboles, estamos interesados en encontrar aquellos que cumplan las mejores características como máquinas de predicción, e intentaremos dar un mecanismo automático de construcción del árbol a partir de los ejemplos. (*Aprendizaje Inductivo: Árboles de Decisión - Fernando Sancho Caparrini, s. f.*). Siendo el ID3 el algoritmo más

utilizado y en el que se basa la creación de otros tipos de algoritmos mejorados.

2.2.8 Algoritmo ID3 - Entropía de Shannon

La entropía se la conoce por formarse por elementos que tienen características comunes relacionadas con su clase o naturaleza, lo que permite establecer una relación similar e idéntica entre ellos. Si la muestra es completamente homogénea, la entropía es cero y si la muestra está dividida en partes iguales, tiene una entropía de uno. («Clasificación de Aprendizaje automático supervisado», 2019a). Siendo la entropía una función de estados, es posible medir el grado de desorden de un sistema.

Fórmula para calcular entropía

$$\text{Entropia}(S) = \sum_{i=1}^n -p_i \log_2 p_i$$

S: Siendo la colección de objetos.

Pi: Probabilidad de los valores que son posibles.

I: Las posibles respuestas.

Si una muestra completamente homogénea (es decir, en la que todos se clasifican igual) tiene incertidumbre mínima, ya que no

hay dudas de cuál es la clasificación de cualquiera de sus elementos (si elegimos al azar cualquier de ellos, sabremos qué resultado tendremos).(*Aprendizaje Inductivo: Árboles de Decisión - Fernando Sancho Caparrini, s. f.*)

Algoritmo ID3 - Ganancia de Información o Gini

Dentro de la ganancia de información se enfoca en decrementar el valor de la entropía o incertidumbre de un sistema, esta es una medida de discriminación, un indicador del siguiente atributo a ser seleccionado para continuar con el proceso de división, discriminando el atributo seleccionado entre los demás atributos aún no clasificados.(«Árbol de decisión (modelo de clasificación ID3)», 2020)

Fórmula para calcular la ganancia:

$$G = \left| 1 - \sum_{k=1}^{n-1} (X_{k+1} - X_k)(Y_{k+1} + Y_k) \right|$$

G: Indicando el valor de Gini a obtener.

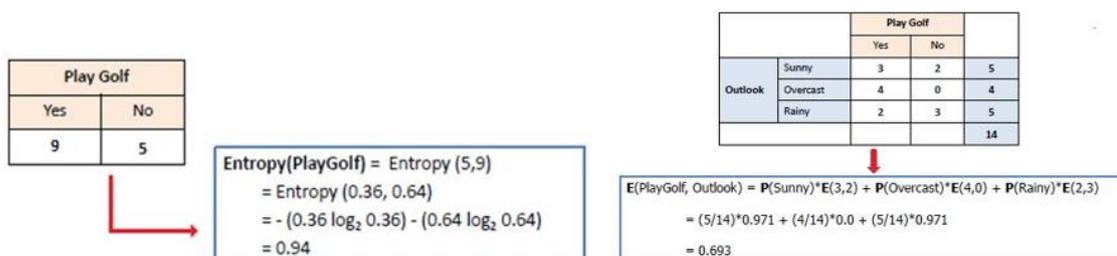
X: Hace referencia a la proporción acumulada de la población.

Y: Hace referencia a la proporción acumulada de los ingresos.

Si una muestra igualmente distribuida, es decir, que tiene el mismo número de ejemplos de cada posible clasificación, muestra una incertidumbre máxima, en el sentido de que es la peor situación para poder saber a priori cuál sería la clasificación de cualquiera de sus ejemplos elegido al azar. (Aprendizaje Inductivo: Árboles de Decisión - Fernando Sancho Caparrini, s. f.)

En la Ilustración 5 se aprecia un ejemplo comparativo de los algoritmos (ganancia y entropía), del lado izquierdo se aplica la **entropía** y del lado derecho se aplica la **ganancia**, donde se evidencia que el uso de la ganancia repercute significativamente en el valor final, y esto es, por las sub-divisiones que realiza el algoritmo.

Ilustración 5 Ejemplo de Entropía y Ganancia



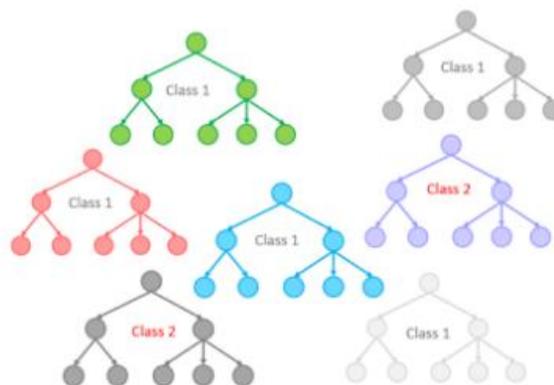
Fuente: Datos de la investigación
Autor: Dennys Cuzco

2.2.9 Bosques Aleatorios - Random Forest

Dentro del algoritmo supervisado del (ML) está el algoritmo de los bosques aleatorios o también conocido como RANDOM FOREST. Random Forest es un método versátil de aprendizaje automático capaz de realizar tanto tareas de regresión como de clasificación. También lleva a cabo métodos de reducción dimensional, trata valores perdidos, valores atípicos y otros pasos esenciales de exploración de datos. Es un tipo de método de aprendizaje por conjuntos, donde un grupo de modelos débiles se combinan para formar un modelo poderoso.(Gonzalez, 2018b)

En la Ilustración 6 se aprecia un ejemplo de la estructura que usa los bosques aleatorios.

Ilustración 6 Algoritmo - Bosques aleatorios



Fuente: Datos de la investigación
Autor: Dennys Cuzco

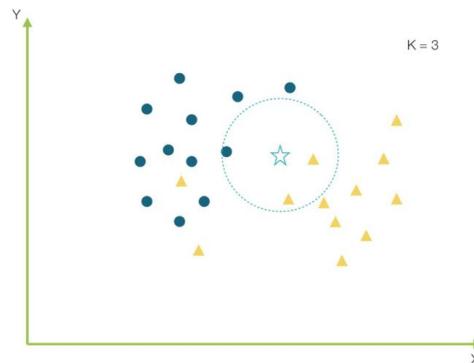
2.2.10 Regresión Logística

Siendo la Regresión Logística un algoritmo que pertenece al (ML), como todos los análisis de regresión, la regresión logística es un análisis predictivo. Se usa para describir datos y explicar la relación entre una variable binaria dependiente y una o más variables independientes nominales, ordinales, de intervalo o de nivel de razón.(Gonzalez, 2018a)

2.2.11 KNN – K vecinos cercanos

Se conoce al algoritmo K-Vecino por tener la clasificación más básica del (ML). Cabe señalar que la fase de formación mínima de KNN se realiza tanto a un coste de memoria, ya que debemos almacenar un conjunto de datos potencialmente enorme, como un coste computacional durante el tiempo de prueba, ya que la clasificación de una observación determinada requiere un agotamiento de todo el conjunto de dato. En la práctica, esto no es deseable, ya que normalmente queremos respuestas rápida.(«K Vecinos más Cercanos - Teoría», 2019)

Ilustración 7 Algoritmo - KNN



Fuente: Datos de la investigación
Autor: Dennys Cuzco

En la ilustración 7 se aprecia como trabaja del algoritmo KNN, los puntos que están fuera del círculo se los identifica como los vecinos más cercanos, generando que, los excluidos se clasifican por sus propias características.

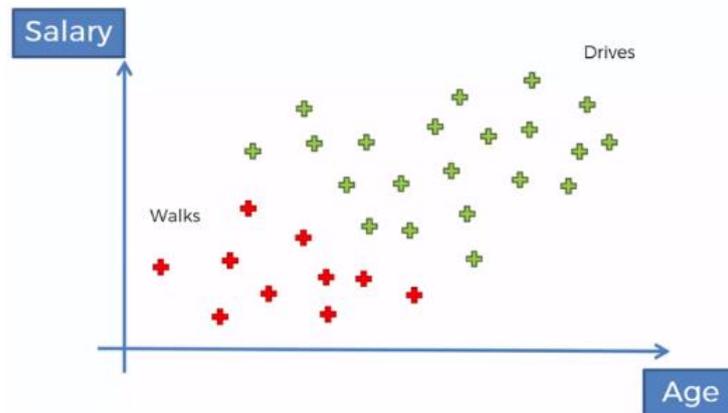
2.2.12 Naive Bayes

La aplicación del Naive Bayes es para la distribución binomial de datos, por tal, se puede decir que el clasificador Naive Bayes se basa en el teorema de Bayes con los supuestos de independencia entre los predictores, es decir, asume que la presencia de una entidad en una clase no está relacionada con ninguna otra característica.

Incluso si estas características dependen unas de otras o de la existencia de las otras características, todas estas propiedades

son independientes. («Clasificación de Aprendizaje automático supervisado», 2019b)

Ilustración 8 Algoritmo - Navie Bayes



Fuente: Datos de la investigación
Autor: Dennys Cuzco

En la ilustración 8 se aprecia como trabaja del algoritmo Naive Bayes, donde la clasificación de los datos los mantiene agrupados y ordenados en dirección a la edad y el salario.

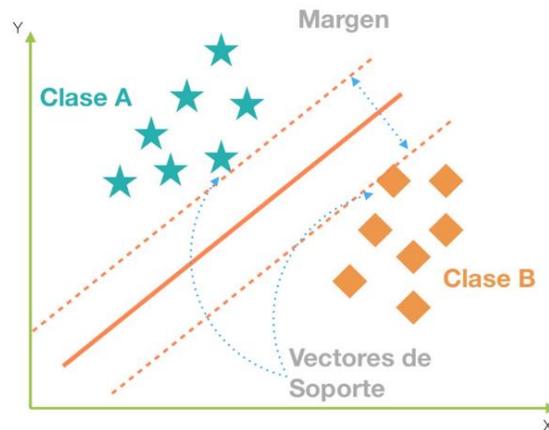
2.2.13 SVM - Máquinas de Vectores de Soporte

En el algoritmo supervisado del (ML) se encuentra el SVM (máquinas de vectores de soporte).

El SVM, por sus siglas en inglés, construye un hiperplano en un espacio multidimensional para separar las diferentes clases. El SVM genera un hiperplano óptimo de forma iterativa, que se utiliza para minimizar un error. La idea central de SVM es

encontrar un hiperplano marginal máximo que mejor divida el conjunto de datos en clases.(Gonzalez, 2019)

Ilustración 9 Algoritmo - SVM



Fuente: Datos de la investigación
Autor: Dennys Cuzco

En la ilustración 8 muestra el conjunto de datos que se encuentra de datos que se encuentra de forma paralela al margen, generando que estos datos se agrupen por 2 clases, clase a y clase B, estas clases por la estructura propia del SVM.

2.3 Marco Legal

Constitución de la República del Ecuador

La Carta Magna, como fuente principal de la autoridad jurídica ecuatoriana, a la actualidad sigue siendo la base legal que brinda validez la presente investigación, pues cuenta con los siguientes artículos como:

En el artículo 32 está establecido que: ““La salud es un derecho que garantiza el Estado, cuya realización se vincula al ejercicio de otros derechos, entre ellos el derecho al agua, la alimentación, la educación, la cultura física, el trabajo, la seguridad social, los ambientes sanos y otros que sustentan el buen vivir”

Ciencia, tecnología, innovación y saberes ancestrales

Art. 385.- El sistema nacional de ciencia, tecnología, Innovación y saberes ancestrales, en el marco del respeto al ambiente, la naturaleza, la vida, las culturas y la soberanía, tendrá como finalidad:

- Generar, adaptar y difundir conocimientos científicos y tecnológicos.

- Recuperar, fortalecer y potenciar los saberes ancestrales.
- Desarrollar tecnologías e innovaciones que impulsen la producción nacional, eleven la eficiencia y productividad, mejoren la calidad de vida y contribuyan a la realización del buen vivir.

Art. 386.- El sistema comprenderá programas, políticas, recursos, acciones, e incorporará a instituciones del Estado, universidades y escuelas politécnicas, institutos de investigación públicos y privados, empresas públicas y privadas, organismos no gubernamentales y personas naturales o jurídicas, en tanto realizan actividades de investigación, desarrollo tecnológico, innovación y aquellas ligadas a los saberes ancestrales.

El Estado, a través del organismo competente, coordinará el sistema, establecerá los objetivos y políticas, de conformidad con el Plan Nacional de Desarrollo, con la participación de los actores que lo conforman.

Art. 387.- Será responsabilidad del Estado:

- a) Facilitar e impulsar la incorporación a la sociedad del conocimiento para alcanzar los objetivos del régimen de desarrollo.
- b) Promover la generación y producción de conocimiento, fomentar la investigación científica y tecnológica, y potenciar los saberes ancestrales, para así contribuir a la realización del buen vivir.
- c) Asegurar la difusión y el acceso a los conocimientos científicos y tecnológicos, el usufructo de sus descubrimientos y hallazgos en el marco de lo establecido en la Constitución y la Ley.
- d) Garantizar la libertad de creación e investigación en el marco del respeto a la ética, la naturaleza, el ambiente, y el rescate de los conocimientos ancestrales.
- e) Reconocer la condición de investigador de acuerdo con la Ley.

Art. 388.- El Estado destinará los recursos necesarios para la investigación científica, el desarrollo tecnológico, la innovación, la

formación científica, la recuperación y desarrollo de saberes ancestrales y la difusión del conocimiento. Un porcentaje de estos recursos se destinará a financiar proyectos mediante fondos concursables. Las organizaciones que reciban fondos públicos estarán sujetas a la rendición de cuentas y al control estatal respectivo. (*Acuerdo 2018-040.pdf*, s. f.)

Constitución de la República Sección Quinta

Art. 44.- El Estado, la sociedad y la familia promoverán de forma prioritaria el desarrollo integral de las niñas, niños y adolescentes, y asegurarán el ejercicio pleno de sus derechos; se atenderá al principio de su interés superior y sus derechos prevalecerán sobre los de las demás personas.

Las niñas, niños y adolescentes tendrán derecho a su desarrollo integral, entendido como proceso de crecimiento, maduración y despliegue de su intelecto y de sus capacidades, potencialidades y aspiraciones, en un entorno familiar, escolar, social y comunitario de afectividad y seguridad.

Art. 46.- El Estado adoptará, entre otras, las siguientes medidas que aseguren a las niñas, niños y adolescentes:

- Atención a menores de seis años, que garantice su nutrición, salud, educación y cuidado diario en un marco de protección integral de sus derechos. (*mesicic4_ecu_const.pdf*, s. f.)

Condición Discapacitante

Art. 2.- De la persona con deficiencia o condición Discapacitante.- Se entenderá por persona con deficiencia o condición Discapacitante, aquella que presente disminución o supresión temporal de alguna de sus capacidades físicas, sensoriales o intelectuales, en los términos que establece la Ley, y que aun siendo sometidas a tratamientos clínicos o quirúrgicos, su evolución y pronóstico es previsiblemente (Garcés, s. f.)

Enfermedades Catastróficas y Raras o Huérfanas

Art 1.- El Estado ecuatoriano reconocerá de interés nacional a las enfermedades catastróficas y raras o huérfanas; y, a través de la autoridad sanitaria nacional, implementará las acciones necesarias para la atención en salud de las y los enfermos que las padezcan, con el fin de mejorar su calidad y expectativa de

vida, bajo los principios de disponibilidad, accesibilidad, calidad y calidez; y, estándares de calidad, en la promoción, prevención, diagnóstico, tratamiento, rehabilitación, habilitación y curación. Las personas que sufran estas enfermedades serán consideradas en condiciones de doble vulnerabilidad (Ricaurte et al., s. f.)

2.4 Marco Institucional

Las Fundaciones podrán ser constituidas por la voluntad de uno o más fundadores. Estas organizaciones buscan o promueven el bien común de la sociedad, incluyendo las actividades de promocionar, desarrollar e incentivar dicho bien en sus aspectos sociales, culturales, educacionales, ambientales, deportivas, así como actividades relacionadas con la filantropía y beneficencia pública; entre otras.(Nuevo reglamento para el otorgamiento de personalidad jurídica de organizaciones sociales, 2017)

La Fundación Ambulancia del Deseo fue creada en 2007 en Holanda por el enfermero y conductor de ambulancias, Kees Veldboer realizando un traslado interhospitalario a Mario Stefanuto, un paciente que había sido marinero. Al pasar por el

puerto, Kees observó que Mario se emocionaba profundamente. Al día siguiente y tras varias gestiones, consiguió los permisos para llevarle en una ruta por el puerto de Rotterdam.(«Fundación Ambulancia del Deseo», 2021). Convirtiéndose así, la experiencia con Mario el inicio de un gran proyecto social.

En una carta escrita antes de su muerte, Stefanutto dijo:

“Me hace bien saber que todavía hay personas a quienes le importan los demás. Les puedo asegurar desde mi propia experiencia que un pequeño gesto de alguien puede tener un gran impacto”(La fundación que cumple los últimos deseos de quienes están a punto de morir, 2015)

El objetivo de la Fundación es cumplir los anhelos de personas enfermas, hospitalizadas, con alguna enfermedad crónica, al final de la vida o en cuidados paliativos, que quieren cumplir un deseo y su enfermedad no se lo permite.

La Fundación, aunque surgió en Holanda con el cumplimiento del primer sueño, está presente en otros 15 países diferentes.

Esto ha generado múltiples beneficios que han cruzado fronteras entre países, por ejemplo, Trasladar a un soñador desde un país a otro para el cumplimiento su sueño, siendo este, en su fase terminal, poder estar en compañía de su familia.

Ilustración 10 Logo de la Fundación en Ecuador



Fuente <https://www.ambulanciadeseo.org.ec>

Autor Dennys Cuzco

La Fundación Ambulancia Deseo Ecuador es una institución sin fines de lucro, conformada por médicos y más de 80 voluntarios que trabajan de manera ardua para cumplir los últimos deseos o sueños a los pacientes que padecen de enfermedades catastróficas.

Contamos con un gran equipo de trabajo, quienes, a través de sus propios recursos, donaciones y buena voluntad, permiten cristalizar el sueño de los soñadores, brindándoles la posibilidad de alcanzar el deseo que tanto anhelan.(admin, s. f.)

Ilustración 11 Equipo de Apoyo



Fuente: <https://www.ambulanciadeseo.org.ec/nosotros/>

Autor: Dennys Cuzco

Aunque el objetivo de la fundación está activa en más de 15 países diferentes, la gestión operativa es de forma individual, es decir, cada fundación maneja de forma independiente su logística u operaciones que se considere de forma interna necesaria para lograr alcanzar sus objetivos.

CAPÍTULO III

Metodología

Para conseguir conclusiones concretas en el proceso de investigación, debe de definirse herramientas que ayuden a procesar los datos de manera objetiva. Mediante el uso de mecanismos computarizados de Machine Learning se pueden llegar a estudiar resultados. Algoritmos como Arboles de decisiones, Random Forest, simbolizan una relevante importancia debido al uso en trabajos científicos, concluyendo así, la importancia para la elaboración de este proyecto.

3.1 Modalidad de la Investigación

El diseño de un modelo predictivo que brinde apoyo en la toma de decisiones en la relación entre soñador y voluntarios de la Fundación "Ambulancia Deseo" corresponde a un proyecto factible, puesto que será una herramienta de apoyo para un problema presente en la fundación "Ambulancia Deseo".

Asimismo, al adaptarse a las características de un proyecto factible, también se adapta como estudio científico porque tiene en cuenta dos aspectos.

- Originalidad de proyecto.
- La realización metodológica y teórica al que se va a referir en esta investigación.

3.2 Tipo de Investigación

Aplicando los criterios de trabajos científicos, a continuación, se define los tipos y niveles a utilizar, que son: investigación exploratoria, comprobación de hipótesis e investigación experimental.

3.3 Metodología del Trabajo

Al utilizar la minería de datos para realizar el respectivo análisis, es posible seguir varios esquemas, tales como:

- KDD.
- SEMMA.
- CRISP-DM.

Siendo el KDD, SEMMA Y CRISP-DM utilizadas en proyectos enfocados a la minería de datos, es fundamental realizar la comparación entre metodologías y determinar cuál es la que mejor se ajusta a la necesidad del trabajo. En las metodologías

y modelos de los procesos, se tienen en cada una de ellas sus respectivas diferencias, los modelos indican las actividades que se realizan, mientras tanto que las metodologías incorporan fases de cómo son realizadas las actividades. El KDD y SEMMA están más enfocadas en las características técnicas del desarrollo de los procesos, pero no detallan el cómo gestionar o desarrollar cada fase. En cambio, CRISP-DM mantiene una perspectiva más amplia sobre los objetivos, actividades y fases de los proyectos.

Ilustración 12 Comparativa de Metodologías

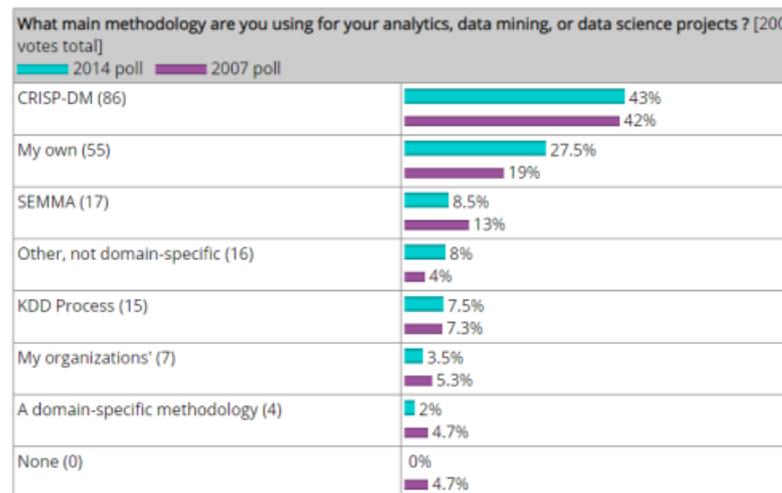
KDD	SEMMA	CRISP-DM
Pre KDD	xxxxx	Conocimiento del negocio
Selección	muestra	Conocimiento de los datos
Preprocesamiento	exploración	
Transformación	Modificación	
Minería de datos	Modelo	
interpretación / evaluación	evaluación	
Post KDD	xxxxx	

Fuente (*Sesion5_Metodologias.pdf*, s. f.)

Con la ilustración 12 muestra de forma resumida para comprender la forma que trabaja cada metodología, además

sustenta porque en la ilustración 13 es la metodología que se destaca en el área de minería de datos.

Ilustración 13 Encuesta de tipos de metodologías



Fuente Encuesta del 2007 al 2014 sobre metodologías utilizadas en minería de datos, («CRISP-DM, Still the Top Methodology for Analytics, Data Mining, or Data Science Projects», s. f.)

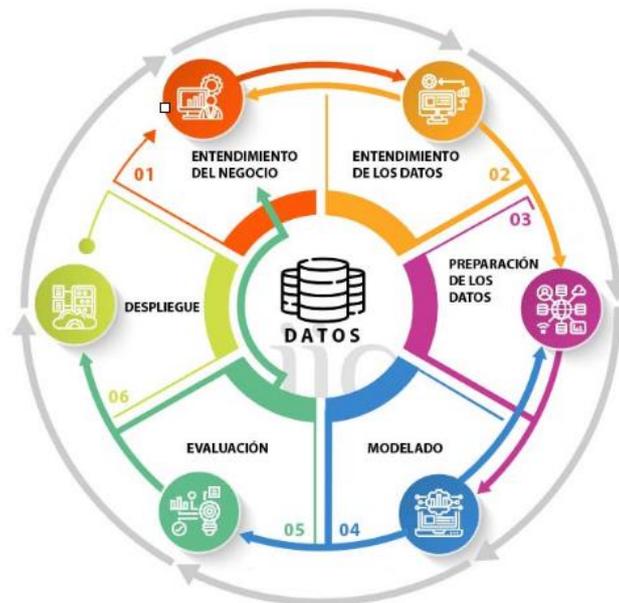
Con la ilustración 13 se puede decidir que, CRISP-DM es la metodología con el mayor porcentaje para el desarrollo de proyectos, por tal, se concluye que es aplicable para el desarrollo del actual proyecto.

Para el diseño del modelo predictivo, se aplica la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), siendo este uno de los modelos más utilizados y el que se ajusta con el objetivo del desarrollo de este trabajo.

CRISP-DM

La metodología posee 6 fases en el ciclo de vida del modelo. Cada fase de la metodología describe un conjunto de actividades o tareas que contribuyen con mejoras de desarrollo a la organización. A continuación, se detalla las 6 fases que posee la metodología: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue. En la ilustración 14 se demuestra de manera gráfica como trabaja CRISP-DM

Ilustración 14 Etapas de Metodología CRISP-DM



Fuente («La metodología CRISP-DM en ciencia de datos - IIC», 2021)

Autor Instituto de Ingeniería del Conocimiento

Fases de CRISP-DM:

Comprensión del Negocio. - Este es el punto de partida para los proyectos de minería de datos que utilizan la técnica CRISPDM. Comenzamos explorando la necesidades y expectativas del negocio, con el objetivo de identificar sus necesidades de forma técnica.

Compresión de los datos. - Aquí se procede con la adquisición de los datos y analizarlos, en el análisis se establecen características, cantidad, consistencia e identificar posibles errores o datos nulos. Esta se considera una fase importante, dado que, una mala interpretación de los datos puede cambiar significativamente el proyecto. Esta fase está relacionada con la preparación y modelado.

Preparación de los datos. – Aquí se realizan tareas como: Procesamiento de los datos, Cálculos de nuevos datos, Depuración de datos duplicados, limpieza de datos. Esta es la etapa preliminar al modelado, un error en esta etapa, afectará significativamente al resultado, por lo tanto, se considera una fase importante que demanda mucho tiempo.

Modelado. – Se examina y escoge el algoritmo que mejor se adapta al planteamiento del problema, se basa con los datos proporcionados en las fases previas. Se establece un plan de pruebas para especificar los parámetros que van a ser procesados por el algoritmo de aprendizaje. Aquí es donde se crea el modelo.

Evaluación. – En esta fase, evalúa los resultados obtenidos de su modelo para ver si cumplen con sus requisitos iniciales. Si no es así, debe volver al paso anterior y analizar si desea realizar una nueva iteración con resultados diferentes.

Despliegue. – Si la etapa anterior se completa con éxito, el modelo se pone a disposición del negocio para su aplicación. Esta fase no está considerada dentro del presente trabajo.

3.4 Desarrollo de la Metodología Elegida (CRISP-DM)

3.4.1 Fase 1: Comprensión del Negocio

Actual Situación y definición de los objetivos del negocio

En Ecuador existen fundaciones que brindan apoyo a personas con enfermedades catastróficas. En el presente proyecto se examinará el proceso de relación entre soñador y voluntario de la fundación “Ambulancia Deseo” quien se dedica al cumplimiento de sueños vivenciales en la ciudad de Guayaquil.

En el transcurso del desarrollo del proyecto no se publicará la información completa que permita identificar a los voluntarios y personal de la fundación, de igual forma, la información confidencial de las personas de quienes padecen enfermedades catastróficas.

La información no es publicada por el cumplimiento del Reglamento General de Protección de Datos que se detalla en el artículo 25, que especifica, que se debe de proteger los datos desde el diseño y por defecto.

A continuación, se detalla en la tabla 1, los recursos a utilizar en el presente proyecto:

Tabla 1 Recursos Usados

Recurso	Descripción	Tipos
Estudiante	El recurso humano que se utiliza para el desarrollo del proyecto es: el estudiante, el directo del proyecto, y el voluntario asignado por la fundación.	Persona
Director de Proyecto		
Voluntario de la Fundación		
Fuente: Base de Datos de la fundación "Ambulancia Deseo"	De esta fuente se obtendrá la información referente a los soñadores y voluntarios que actualmente posee registrada la fundación	Fuentes de datos, Archivos digitales y físicos.
Herramienta que se utilizará para la depuración, transformación para la aplicación de los algoritmos de minería de datos.	Para la realización de este proceso se utilizará, Python, siendo un lenguaje de programación con diversas librerías que apoyan en el proceso de minería de datos.	Software Tecnológico

**Fuente Datos de la Investigación
Autor Dennys Cuzco.**

En la tabla 2, se indicará las restricciones detectadas durante la elaboración del proyecto.

Tabla 2 Restricciones

Seguridad	Legal ó ética
Acceso a los registros: En la recolección de los datos se tendrá acceso a los datos que no contengan información que pueda identificar a los voluntarios de la fundación.	Privacidad de Datos: Se procede a utilizar datos que no permitan identificar a los soñadores ni a los voluntarios. Uso de Datos: Se utilizarán datos que permitan identificar la relación entre soñador y voluntario.

**Fuente Datos de la Investigación
Autor Dennys Cuzco.**

Los términos a utilizar en el proyecto se especifican a continuación:

Generales

Sueño: Deseo intenso de recibir algo por parte de la fundación, solicitado por una persona que padece alguna enfermedad catastrófica o estado de salud afectado.

Vivencial: Experiencia vivencial que se anhela adquirir a partir de un deseo.

Soñador: Persona que padece alguna enfermedad catastrófica, estado de salud en desahucio o similares.

Voluntario: Persona que ofrece desinteresadamente su ayuda para lograr un objetivo.

Relación: Conexión o vínculo que hay entre dos personas.

Técnicos

Python: Es un lenguaje de programación que contiene módulos, excepciones, tipos de datos dinámicos de muy alto nivel y clases, hace hincapié en la legibilidad de su código y contiene interfaces para cubrir el llamado recurrentes al sistema y bibliotecas.

El lenguaje Python dentro del Machine Learning, se distingue porque su aplicabilidad es muy frecuente, el motivo de su uso es, la gran facilidad y sencillez en el uso de las instrucciones. Otro punto que lo destaca frente a otros lenguajes es que permite el uso de modelos matemáticos, característica que no son común en otros lenguajes de programación.

A continuación, se detalla las librerías a usarse para la manipulación de datos y generación de gráficos:

Pandas: Biblioteca que permite la manipulación y análisis de datos de código abierto eficiente, potente y muy flexible en su uso.

Scikit-learn: Biblioteca de código abierto, maneja diversos algoritmos de última generación, muy usado dentro del aprendizaje automático con Python.

Matplotlib: Biblioteca de código abierto, permite la creación de gráficos en dos dimensiones, para la generación de gráficos parte de datos contenidos en listas o arrays.

Seaborn: Librería de código abierto, permite la visualización de datos ofreciendo interfaces de alto nivel.

Google Collab: Se lo conoce con el nombre de "Colaboratory", con el uso de esta herramienta es posible programar y ejecutar Python

en el navegador. Su finalidad es de facilitar trabajo relacionados al estudio de los datos científicos.

3.4.2 Fase 2: Comprensión de los Datos

Esta fase se centra en la capacidad de comprender los datos iniciales. Esto va a permitir identificar los problemas, descubrir conocimiento preliminar, subconjuntos de datos que permitan formar una hipótesis de la información que se está analizando, entre otros aspectos relacionados a los datos.

Recolección de los datos iniciales

En la recolección de datos, están los registros de los soñadores y de los voluntarios, los mismos están en los archivos digitales y físicos que posee actualmente la fundación "Ambulancia Deseo" de la ciudad de Guayaquil.

Los datos digitales que fueron compartidos están en archivos CSV, la información que se encuentra en documentos físicos, los mismos fueron migrados a archivos digitales con el propósito de unificar ambos registros.

Se describe en la tabla 3 y en la tabla 4, los campos que contiene los archivos compartidos por la fundación, después se describirá detalladamente el tipo de información que posee cada campo para una mejor comprensión de los datos.

Los archivos compartidos con la información de los soñadores, contiene los campos que se detalla en la tabla 3.

Tabla 3 Registro de Soñadores

Dato	Tipo de Dato	Tipo de Variable
Id	Numérico	Cuantitativo-Discreto
Fecha de Registro	String	Cualitativo-Nominal
Descripción del Sueño	String	Cualitativo-Nominal
Edad	Numérico	Cuantitativo-Discreto
Nombre Completo	String	Cualitativo-Nominal
Género	String	Cualitativo-Nominal
Datos de Familiar	String	Cualitativo-Nominal
Nombre de Albergue	String	Cualitativo-Nominal
Ciudad	String	Cualitativo-Nominal
Número Telefonico	String	Cualitativo-Nominal

Fuente Datos de la Investigación
Autor Dennys Cuzco.

Los archivos compartidos con la información de los Voluntarios, contiene los campos que se detalla en la tabla 4:

Tabla 4 Registro de Voluntarios

Dato	Tipo de Dato	Tipo de Variable
Id	Numérico	Cuantitativo-Discreto
Fecha de Registro	String	Cualitativo-Nominal
Nombre Completo	String	Cualitativo-Nominal
Correo	String	Cualitativo-Nominal
Número Telefónico	String	Cualitativo-Nominal
Nacionalidad	String	Cualitativo-Nominal
Ciudad	String	Cualitativo-Nominal
Género	String	Cualitativo-Nominal
Edad	Numérico	Cuantitativo-Discreto

Fuente: Datos de la Investigación
Autor: Dennys Cuzco.

A continuación, las definiciones de cada campo:

Id: Variable de valor único e incremental, variable que se utiliza para su identificación.

Fecha de Registro: Variable que contiene el día, mes y el año del registro, según al tipo que aplica, sea este, soñador o voluntario.

Descripción del Sueño: Variable de tipo texto, el soñador describe el tipo de sueño que desea recibir.

Edad: Variable de tipo numérico, refleja la edad que tiene a la fecha de registro.

Nombre Completo: Variable de tipo texto que contiene los nombres completos, según al tipo que aplica, soñador o voluntario.

Género: Definición del tipo de persona, Femenino o Masculino, según al tipo que aplica, sea este, soñador o voluntario.

Datos de Familiar: Variable de tipo texto, contiene información relevante del familiar que acompaña generalmente al soñador.

Nombre del Albergue: Variable de tipo texto, contiene el nombre de la fundación o institución que apoya en el tratamiento del soñador.

Ciudad: Corresponde al nombre de la ciudad, ubicación donde reside, según al tipo que aplica, soñador o voluntario.

Número Telefónico: Variable de tipo texto, contiene el número de contacto, según al tipo que aplica, soñador o voluntario.

Correo: Variable de tipo texto, información de contacto vía e-mail.

Nacionalidad: Variable de tipo texto, especifica el país en el que nació, este campo lo recopilan debido a que la fundación tiene fundaciones aliadas en otros países.

La información que se utiliza para el desarrollo de este trabajo, fueron obtenidos de la fundación "Ambulancia Deseo" de la ciudad de Guayaquil.

Internamente la fundación clasifica a las personas según el aporte o finalidad con la que se vincula a la fundación, a continuación, se lo describe:

Voluntario: Persona que brinda su apoyo a la fundación de forma desinteresada.

Soñador: Persona que padece alguna complicación de salud, quien acude a la fundación con el objetivo de recibir el cumplimiento de un sueño vivencial.

Administrativo: Persona que gestiona actividades de alto rango, por ejemplo, acuerdo con instituciones o representación en reuniones, etc.

La Fundación "Ambulancia Deseo" posee un proceso de evaluación interna, del cual, ellos mismos realizan una preclasificación según datos del soñador y del sueño solicitado, y los mismos se asignan por los siguientes conceptos:

Bajo: Determina el cumplimiento del sueño como factible y con poca complejidad.

Medio: Determina el cumplimiento del sueño con dificultad, ya sea, por movilidad de parte del soñador, complejidad por el sueño que solicitan y otros.

Actualmente la clasificación se realiza según el soñador, voluntario y el sueño a cumplir, paralelamente con el criterio del encargado del sueño (otro voluntario).

Descripción de Datos

Los campos recibidos en archivos de EXCEL permiten el respectivo análisis de los mismos. En la tabla 5 se muestra la información compartida que pertenece a los soñadores de la Fundación "Ambulancia Deseo" de la ciudad de Guayaquil.

Tabla 5 Tabla con Campos Originales de los Soñadores

N O	FEC HA	SUEÑO	EDA D	NOMB RE	Se xo	FAMILI AR	ALBERG UE	CIUDA D	TELEFO NO	CORREO
1	2021-09-29	Conocer y actuar con el elenco de don day y los pequeños actores pasar la tarde con ellos	8	Zuleyka	F	Peña vargas	Ronald	Guayaquil	098804x	Andy1519@xx
2	2021-09-29	Quisiera conocer focas y todos los animales que habitan en galápagos viajar en avión a Quito	10	Misael Mathius	F	VERA MITE	Ronald	Guayaquil	099646x	
3	2021-09-29	Conocer las Islas Galápagos	8	Evelyn Charlotte	F	Lara Reyes	Ronald	Guayaquil	098210x	evelyn_re@xx
4	2021-09-29	Conocer Galápagos	6	Dylan Reynaldo	M	Alvarado Alvarado	Ronald	Daule	096891x	divettez@xx
5	2021-09-29	Viajar a las islas Galápagos o conocer al jugador Lionel Messi	11	Robinson Santiago	M	Quiroz Villon	Ronald	San Isidro	098116x	alexands@xx
6	2021-09-29	Conocer las Islas Galápagos	8	Evelyn Charlotte	F	Lara Reyes	Ronald	Guayaquil	098210x	evelyn_r@xx
7	2021-09-29	Conocer las Islas Galápagos	14	Alex Raul	M	Lindao Avegno	Ronald	Guayaquil	099058x	Cristyo9@xx
8	2021-09-29	Poder viajar y conocer las islas Galápagos	7	Ronny Yeshua	M	Vinces Albet	Ronald	Manta	099008x	albet_ma@xx
9	2021-09-29	Tener una habitación con comodidad	10	Tifanny Anahi	F	Gaibor Chimbó	Ronald	Montalvo	096851x	axelmed2@xx
10	2021-09-29	Conocer Galápagos	6	Nixon Andres	M	Rodríguez velasco	Ronald	Guayaquil	098713x	carla9290@xx

11	2021-09-29	Adecuar el cuarto por su enfermedad	19	David Joshua	M	Lopez Coronel	Ronald	Guayaquil	093996x	susana092@xx
12	2021-09-29	Celebrar los 17 años con los jugadores de Barcelona en el estadio y meter un gol junto a ellos.	16	Víctor Bolívar	M	Montalvo Sarmiento	Ronald	Machala	099396x	anasarmien@xx
13	2021-09-29	Viajar a la Playa en familia	6	Andrés Alejandro	M	verdesoto Varas	Soldaditos	Quevedo	096895x	varascather@xx
14	2021-09-29	Conocer México Acapulco o las islas Galápagos	19	Alex steven	M	Etsamp Sacon	Ronald	Santo Domingo	098017x	etsampalex@xx
15	2021-09-29	Conocer Galápagos	6	Dylan Reynaldo	M	Alvarado Alvarado	Ronald	Daule	096891	divetteaz@xx

**Fuente: Datos de la Fundación “Ambulancia Deseo” GYE
Voluntaria: Mariela Carchi.**

En la tabla 6 se muestra la información compartida que pertenece a los voluntarios de la Fundación "Ambulancia Deseo" de la ciudad de Guayaquil.

Tabla 6 Tabla con Campos Originales de los Voluntarios

Id Registro	Fecha de Registro	Nombres Completo	Correo electrónico:	Número de teléfono de contacto:	Nacionalidad	Ciudad	Sexo	Edad
1	14/11/2017	Paola	Pbonaffin@xx	59398414X	Ecuatoriana	Guayaquil	F	33
2	14/11/2017	Maria Ines	maonzaga@xx	59399907X	Ecuatoriana	Guayaquil	F	59
3	14/11/2017	Diana Elizabeth	dianelijurcs@xx	96788340X	Ecuatoriana	Guayaquil	F	40
4	14/11/2017	Larissa	lvillavicenc@xx	99502918X	Ecuatoriana	Guayaquil	F	35
5	14/11/2017	Mariela Elizabeth	arielacarc@xx	59396960X	Ecuatoriana	Guayaquil	F	35
6	14/11/2017	Ana	Anama_1@xx	59398731X	Ecuatoriana	Guayaquil	F	26
7	14/11/2017	Alexis lorena	alexischae@xx	09947822X	Ecuatoriana	Guayaquil	M	34
8	14/11/2017	Paola Cecilia	pc260886@xx	98571751X	Ecuatoriana	Guayaquil	F	31
9	16/11/2017	shirley briggitte	shirley_me@xx	93922736X	Ecuatoriana	Guayaquil	F	20
10	16/11/2017	daive	barcetego@xx	98588997X	Ecuatoriana	Guayaquil	M	32
11	16/11/2017	Mabel	maribelcriol@xx	59393996X	Ecuatoriana	Guayaquil	F	24
12	16/11/2017	Brigitte Rubi	kuku_94@xx	59396747X	Ecuatoriana	Guayaquil	F	23
13	16/11/2017	Nathaly Romina	Natty_4ev7@xx	98266848X	Ecuatoriana	Guayaquil	F	20
14	16/11/2017	Julia	izurietajulia@xx	59399907X	Ecuatoriana	Quito	F	25
15	17/11/2017	Mariyayirina	ryalayana@xx	93966306X	Ecuatoriana	Guayaquil	F	18
16	17/11/2017	Alexa Ariana	arianarive@xx	09390253X	Ecuatoriana	Guayaquil	F	20
17	17/11/2017	Jimena Patricia	menitaxy@xx	59399006X	Ecuatoriana	Guayaquil	F	25
18	17/11/2017	Steven Josué	stevenchaax@xx	53988995X	Ecuatoriana	Guayaquil	M	22
19	17/11/2017	Estefanía	angi_rocio@xx	59399183X	Ecuatoriana	Guayaquil	F	25
20	18/11/2017	Carolina	Carolinabla@xx	59398777X	Extranjera	Guayaquil	F	33

**Fuente: Datos de la Fundación "Ambulancia Deseo" GYE
Facilitador: Mariela Carchi.**

La descripción del sueño, en algunos casos, son nuevamente consultados para tener una mejor comprensión del sueño que están solicitando.

La edad, tanto en el soñador como del voluntario, está regida a la fecha de registro, para un posterior procesamiento de los datos, la edad debe de ser reajustada.

3.4.3 Fase 3: Preparación de los Datos Análisis de Datos

Para la preparación de datos y su respectivo análisis, se unificó los datos en un solo archivo de Excel, tanto los digitales como los que estaban de forma física. Con el objetivo de facilitar el análisis que cumple cada campo y así determinar cuáles son los que influyen de manera significativa en el proceso de relación entre soñador y voluntario.

Ilustración 15 Datos de Soñador y Voluntario

SOÑADOR										VOLUNTARIO												
NO.	STRO	SUEÑO	EDAD	NOMBRE	SEXO	FAMILIA	ALBET	CIUDAD	TELEF	CORREO	Id	Fecha	Nombres	Apellido	Correo	Número	Nacimiento	Ciudad	Dirección	Sexo	Edad	
1	1	Conocer y actuar con	3	Zuleyk	F	Pafia	Casa R	Guayaq	080804	Andyf9@oz	1	1	Padra	Bonafini	bonafini	59399414	Ecuador	Guayaquil	Pradera 10da	F	33	
3	2	Quisiera conocer focar	8	Michael	F	VEPA	Casa R	Guayaq	080443	399	2	2	Maria Ines	Gonzaga	maines	59399907	Ecuador	Guayaquil	Las Fibras	F	59	
4	3	Conocer las Islas Galá	8	Evelyn	F	Lara R	Casa R	Guayaq	080200	evelyn_rejos	3	3	Diana Elizab	Jurado Yl	dianeliz	59789834	Ecuador	Guayaquil	Villa del rey e	F	40	
5	4	Conocer Galápagos	6	Dylan	F	M	Alvara	Casa R	Daule	080891	divetteaz@i	4	4	Larissa	Villavicenc	villavicenc	59900239	Ecuador	Guayaquil	Km. 2.5 Vía S	F	35
7	5	Viajar a las Islas Galápa	11	Robinet	M	Quiroc	Casa R	San Isidro	080182	alejandrovill	5	5	Mariela Elizab	Carchi Gd	marizab	59396860	Ecuador	Guayaquil	Urb. Casita	F	35	
8	6	Conocer las Islas Galápa	3	Evelyn	F	Lara R	Casa R	Guayaq	080200	evelyn_rejos	6	6	Ana	Guajala	Anama	59398738	Ecuador	Guayaquil	Cdla Bella su	F	26	
9	7	Conocer las Islas Galápa	14	Alex R	M	Lindae	Casa R	Guayaq	080902	Citargony99	7	7	Alexis Iorena	Chavez M	alexisca	08947822	Ecuador	Guayaquil	la togo	M	14	
10	8	Poder viajar y conocer	7	Ronny	M	Vinceo	Casa R	Manta	090008	abet_mariar	8	8	Paola Cecilia	Quifoneo	pc28088	59671913	Ecuador	Guayaquil	Km. 24 vía s	F	31	
11	9	Tener una habitación co	10	Tiffany	F	Gaboi	Casa R	Montal	080858	asemed2006	9	9	shirley briggit	Mendoza	shirley_m	59382736	Ecuador	Guayaquil	Guasmo Sur	F	20	
12	10	Conocer Galápagos	6	Nixon	M	Rodri	Casa R	Guayaq	080713	carla8230@p	10	10	salve	moralis	barceteg	59588997	Ecuador	Guayaquil	dignidad pop	M	32	
13	11	Adeusar el cuarto por s	19	David	M	Lopez	Casa R	Guayaq	083994	susana0822	11	11	Mabel	Cincho	marbelco	59332996	Ecuador	Guayaquil	ciudadela ju	F	24	
14	12	Celebrar los 15 años co	16	Victor	M	Norie	Casa R	Machul	080926	anasarmec9	12	12	Erizma Pab	Borbor L	lukia_34	59398747	Ecuador	Guayaquil	guasmo sur	F	23	
15	13	Visitar a la Playa en Fam	6	Andres	M	verde	Albergu	Gueved	080896	varacatheri	13	13	Nahaly Flor	Pillagasa	Nahay_4	59266848	Ecuador	Guayaquil	Pascalez	F	20	
16	14	Conocer Mexico Acaga	19	Alex st	M	Etsam	Casa R	Santo C	080077	etsampalesf	14	14	Julia	Izueta	izuetar	59339390	Ecuador	Quito	Tanda	F	25	
17	15	Conocer Galápagos	6	Dylan	F	M	Alvara	Casa R	Daule	080891	divetteaz@i	15	15	Mariya irina	Layana ar	Marizab	59386306	Ecuador	Guayaquil	Villa españa	F	18
18	16	Visitar alaz Islas Galápa	4	Pedro	M	Carral	Casa R	Mocaco	080804093		16	16	Alexia Ariana	Piorea C	arianamy	59390251	Ecuador	Guayaquil	Duran Pezre	F	20	
19	17	Conocer Galápagos, Is	11	Adrian	M	Santill	Casa R	El Babulo	080200	angelherber	17	17	Jimena Pate	Cueva C	mentiam	59399086	Ecuador	Guayaquil	Miloto	F	26	
20	18	Reunión de 15 años en	16	Angelín	F	Arzob	Casa R	Guayaq	080248	maiagon84	18	18	Steven Josu	Changot	stevens	59388995	Ecuador	Guayaquil	Cdla. Flores	M	22	
21	19	Visitar y conocer a Shad	19	Elkin Iv	M	Haro	Casa R	Guayaq	080944	ivaniv4458@	19	19	Estefania	Cevallos	angl_roc	59399834	Ecuador	Guayaquil	Guayacanes	F	28	
22	20	Visitar Quito	16	Andrea	F	Diaz G	Casa R	Guayaq	080954	guerracarolin	20	20	Carolina	Bianco	Carolina	59396777	Extranje	Guayaquil	Alborada IIV	F	33	
23	21	Ir a Galápagos	10	Janny	F	Melg	Casa R	Santa E	080077	naryo1989@	21	21	Allisson Mdel	Pérez Ont	alliperez	08380707	Ecuador	Oña	Loreno de G	F	21	
24	22	Quisiera su sueño es cono	11	Alex P	M	Mite	Casa R	Guayaq	080930	lademite@	22	22	Legny	Lopez	legnylat	59372962	Ecuador	Guayaquil	Urb.la joya e	F	47	

Fuente: Datos de la Investigación
Autor: Dennys Cuzco.

En la ilustración 15, se analiza la información de los soñadores y los voluntarios, en conjunto con un voluntario de la fundación, se empieza a pre seleccionar los campos que se consideran relevantes y que podrían influir en la relación entre soñador y voluntario, a lado de cada campo pre seleccionado se especifica como se identificará cada campo durante el desarrollo del proyecto, cada campo pre seleccionado se detallan a continuación:

Soñador

- Edad (S_Redad).
- Sexo (S_Sexo).
- Categoría (S_Categoria).

Voluntario

- Edad (V_Redad).
- Sexo (V_Sexo).
- Ciudad (V_Ciudad).

La pre selección de los campos se la realizó bajo los siguientes soportes trabajados en equipo con voluntarios de la fundación “Ambulancia Deseo”:

Edad: Con esta información es posible identificar frente al tipo de persona o personalidad básica que tiene la persona, esa esta,

soñador o voluntario. Por ejemplo. si es un niño, un joven ó un adulto.

Sexo: Se considera prioridad, dentro de la fundación, que el acompañante y el soñador sea del mismo género o sexo, aunque no se considerada como una restricción.

Categoría: Es determinada por una preclasificación de un voluntario de la fundación, donde puede involucrar la complejidad del sueño solicitado o el estado del soñador por su estado de salud.

Ciudad: Permite identificar a los voluntarios según la ubicación a donde se vaya a cumplir el sueño, la identificación de la ciudad se solicita por el proceso de expansión que está trabajando internamente la Fundación “Ambulancia Deseo”

Técnicas estadísticas para el procesamiento de la información

En esta sección se desarrolla las técnicas aplicadas para el análisis y proceso de la información de la fundación "Ambulancia Deseo".

La correlación de Pearson nos guía como punto de partida para cubrir con la comparación de manera grupal de los datos, también incluye las relaciones que existen en las variables.

La información que se detalla a continuación se obtiene por medio del uso de la herramienta estadística SPSS, el mismo brinda una facilidad para el procesamiento y análisis de los datos.

Ilustración 16 Estadístico Descriptivo

Estadísticos descriptivos

	Media	Desv. Desviación	N
V_Ciudad	1,04	,217	206
V_Sexo	1,33	,470	206
V_Redad	1,46	,581	206
S_Redad	2,70	1,356	206
S_Sexo	1,44	,497	206
S_Categoria	1,35	,479	206

Fuente: Datos de la Investigación

Autor: Dennys Cuzco.

En la ilustración 16 se aprecia las variables a analizar durante el desarrollo de este trabajo, cada variable muestra su media, desviación y el total de registros (N)

Ilustración 17 Correlación de Variables

		Correlaciones					
		V_Ciudad	V_Sexo	V_Redad	S_Redad	S_Sexo	S_Categoria
V_Ciudad	Correlación de Pearson	1	-,077	-,064	,023	-,022	-,039
	Sig. (bilateral)		,274	,362	,747	,750	,577
	N	206	206	206	206	206	206
V_Sexo	Correlación de Pearson	-,077	1	,151*	,029	,517**	-,081
	Sig. (bilateral)	,274		,031	,675	,000	,247
	N	206	206	206	206	206	206
V_Redad	Correlación de Pearson	-,064	,151*	1	-,385**	,151*	,275**
	Sig. (bilateral)	,362	,031		,000	,030	,000
	N	206	206	206	206	206	206
S_Redad	Correlación de Pearson	,023	,029	-,385**	1	,171*	-,093
	Sig. (bilateral)	,747	,675	,000		,014	,184
	N	206	206	206	206	206	206
S_Sexo	Correlación de Pearson	-,022	,517**	,151*	,171*	1	,268**
	Sig. (bilateral)	,750	,000	,030	,014		,000
	N	206	206	206	206	206	206
S_Categoria	Correlación de Pearson	-,039	-,081	,275**	-,093	,268**	1
	Sig. (bilateral)	,577	,247	,000	,184	,000	
	N	206	206	206	206	206	206

*. La correlación es significativa en el nivel 0,05 (bilateral).

**.. La correlación es significativa en el nivel 0,01 (bilateral).

Fuente: Datos de la Investigación
Autor: Dennys Cuzco.

Análisis: La tabla de correlación, ilustración 17, nos permite comprender y determinar si “existe” o “no existe” una relación entre las variables que se va a analizar, la relación se la determina por medio del nivel de significancia. Además, mediante el coeficiente “r” de Pearson tenemos el grado de correlación entre las variables, en este caso se analiza las variables pre seleccionadas.

Es posible indicar que las variables tienen un valor significativo entre sí cuando su valor relacional es menor a 0.05.

Adicionalmente, si el valor es menor a 0.01 significa que hay una relación muy significativa entre las variables.

Analizando la variable V_Sexo frente a las otras variables, es posible indicar en la variable V_Redad que su nivel de confianza es del 95%, es decir, el dato de coeficiente de la correlación “r” es significativo. La variable S_Redad no ingresa en el rango de significancia, pero al existir un alto grado de significancia con las otras variables es posible su aplicabilidad. Finalmente, la variable S_Sexo se considera que su nivel de significancia es del 99% es decir los datos del coeficiente de correlación “r” es “muy significativo”, concluyendo así que es posible emplear la variable en el desarrollo de este proyecto.

Analizando la variable V_Redad frente a las otras variables, es posible indicar en la variable V_Sexo que su nivel de confianza es del 95%, es decir, el dato de coeficiente de la correlación “r” es significativo. Siendo muy similar el análisis con la variable S_Sexo. Por otro lado, las variables S_Redad y S_Categoria se considera que su nivel de significancia es del 99% es decir los datos del coeficiente de correlación “r” es “muy significativo”, concluyendo

así que es posible emplear la variable en el desarrollo de este proyecto.

Analizando la variable S_Redad frente a las otras variables, es factible mencionar que en la variable V_Redad y S_Sexo, se considera que su nivel de significancia es del 99% es decir los datos del coeficiente de correlación “r” es “muy significativo”, concluyendo así que es posible emplear la variable en el desarrollo de este proyecto.

Analizando la variable S_Sexo frente a las otras variables, es factible mencionar que en las variables V_Sexo, S_Redad y S_Categoria se consideran que su nivel de significancia es del 99% es decir los datos del coeficiente de correlación “r” es “muy significativo”. Con la variable V_Redad se aprecia que su nivel de confianza es del 95%, es decir, el dato de coeficiente de la correlación “r” es significativo, logrando así a la conclusión que es posible emplear la variable en el desarrollo de este proyecto.

Finalmente analizando la variable S_Categoria frente a las otras variables, es posible mencionar que V_Redad y S_Sexo se consideran que su nivel de significancia es del 99% es decir los

datos del coeficiente de correlación “r” es “muy significativo” ingresando este último campo para el desarrollo del proyecto.

Análisis de Datos

Soñadores

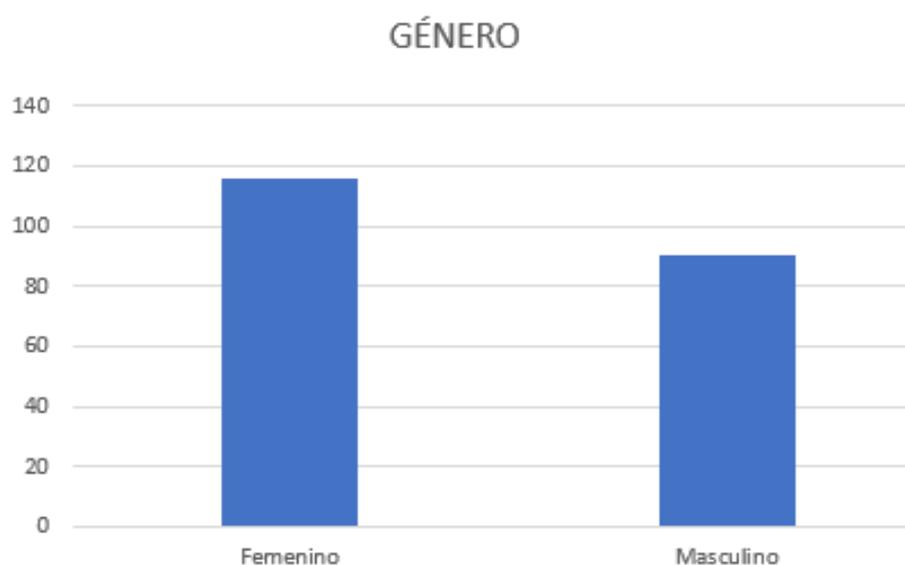
Género: Se presentan la información en la tabla 7 de estadística y los histogramas de las estadísticas del campo género que pertenece al grupo de Soñadores de la fundación “Ambulancia Deseo”.

Tabla 7 Análisis del Género - Soñador

	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Femenino	116	56,31%	56,31%	56,31%
Masculino	90	43,69%	43,69%	100,00%
Total	206	100,00%	100,00%	

**Fuente: Datos de la Investigación
Autor: Dennys Cuzco**

Ilustración 18 Análisis del Género - Soñador



Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Análisis: En la ilustración 18 se aprecia que el género es un factor no modificable que indican en las enfermedades catastróficas. Se aprecia que hay un 56,31% del género Femenino y un 43,69% del género Masculino. Estos datos serán usados para el entrenamiento del algoritmo.

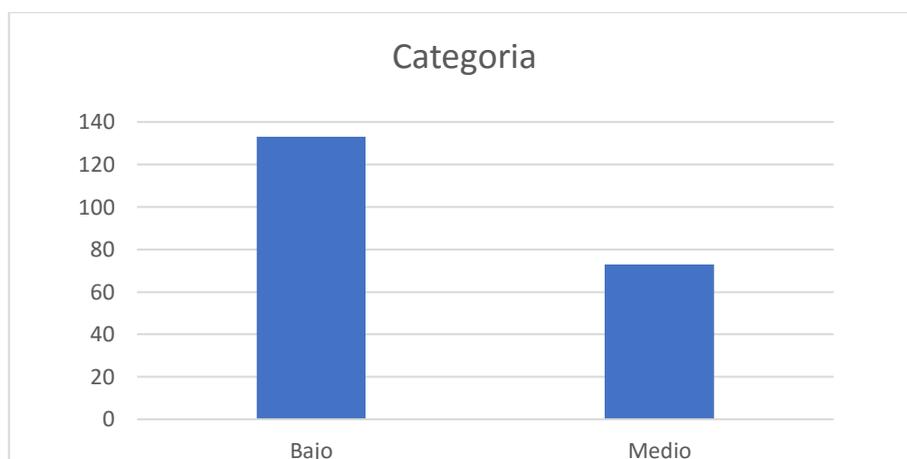
Categoría: Se presentan la información en la tabla 8 de estadística y los histogramas de las estadísticas del campo categoría, que pertenece al grupo de Soñadores de la fundación “Ambulancia Deseo”.

Tabla 8 Análisis de Categoría - Soñador

	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Bajo	133	64,56%	64,56%	64,56%
Medio	73	35,44%	35,44%	100,00%
Total	206	100,00%	100,00%	

Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Ilustración 19 Análisis de la Categoría - Soñador



Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Análisis: En la ilustración 19 se demuestra que la categoría es un factor que permite la identificación de casos con mayor complejidad que no necesariamente puede ser por la enfermedad catastrófica. Se aprecia que hay un 64,56% de categoría tipo bajo y un 35,44% de categoría tipo medio, estos datos serán usados para el entrenamiento del algoritmo.

Edad: Se presentan la información en la tabla 9 de estadística y los histogramas de las estadísticas del campo Edad, que pertenece al grupo de Soñadores de la fundación “Ambulancia Deseo”.

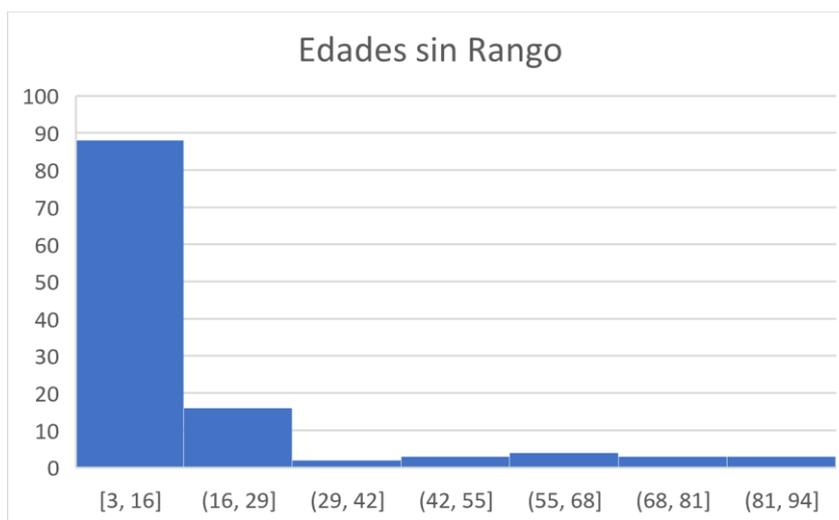
Tabla 9 Análisis de Edades - Soñador

Datos Agrupados		
Edades	Frecuencia	Porcentaje
3 - 16	89	43%
17 - 29	40	19%
30 - 42	10	5%
43 - 55	7	3%
56 - 68	30	15%
69 - 81	15	7%
82 - 94	15	7%
	206	100%

Fuente: Datos de la Investigación
Autor: Dennys Cuzco

La información que se muestra en el campo Edad se encuentra de forma dispersa, para lo cual, damos uso al diagrama de Pareto traza de distribución de datos, para evaluar su clasificación y estimar el uso de rangos y definirlos para utilizarlos dentro del modelado.

Ilustración 20 Edades clasificado según Pareto



Fuente: Datos de la Investigación
Autor: Dennys Cuzco

En la ilustración 20 se muestra la clasificación según el diagrama de Pareto, para lo cual, se procede con la reducción de los rangos. El propósito es, evitar la dispersión de datos por el campo de Edad.

Con el apoyo de un voluntario administrativo de la fundación “Ambulancia Deseo”, se definió los siguientes rangos para la edad.

Dentro de la Tabla 10 se visualizar los rangos definidos.

Tabla 10 Rango de Edades

Rango	Categoría
0 - 17	Menor
18 - 30	Joven
31 - 52	Adulto
> 53	Avanzada

Fuente: Datos de la Investigación
Autor: Dennys Cuzco

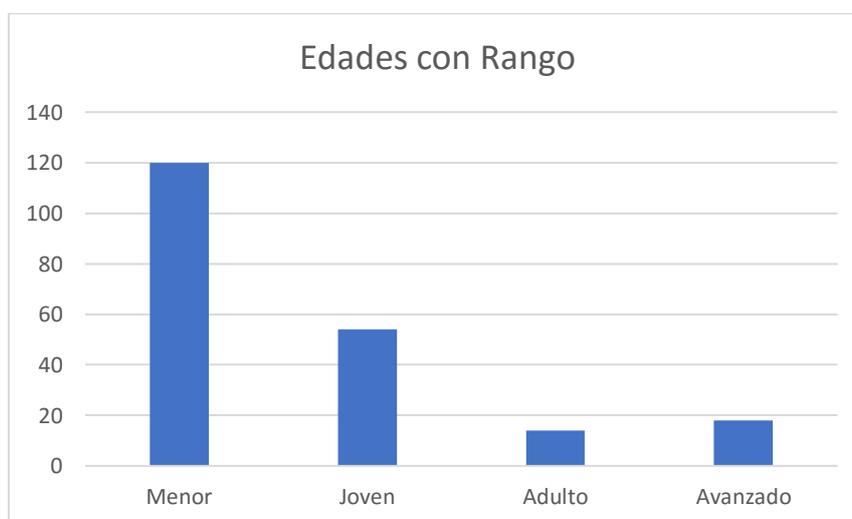
Se asigna los rangos establecidos en las edades de los soñadores, se muestra en la tabla 11 reajustada la información.

Tabla 11 Análisis de Edades con Rango - Soñador

Datos Agrupados		
Edades	Frecuencia	Porcentaje
0 – 17	120	58%
18 – 30	54	26%
31 - 52	14	7%
> 53	18	9%
	206	100%

Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Ilustración 21 Análisis de Edades con Rango Establecido - Soñador



Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Análisis: La ilustración 21 demuestra que, con la asignación de rangos, en las edades de los soñadores, la población que la fundación “Ambulancia Deseo” cumple sueños, recae en su mayoría en la categoría menor, es decir, a personas de corta edad, niños, superando el 50%. Seguido de la categoría Joven con un 26%. Estos datos serán usados para el entrenamiento del algoritmo.

Voluntarios

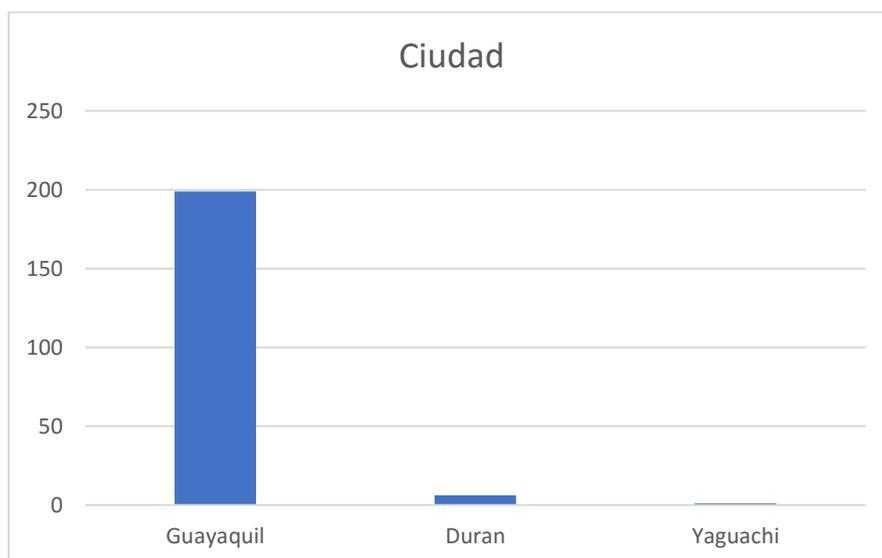
Ciudad: Se presentan la información en la tabla 12 de estadística y los histogramas de las estadísticas del campo Ciudad, que pertenece al grupo de Voluntarios de la fundación “Ambulancia Deseo”.

Tabla 12 Análisis de Ciudad - Voluntarios

	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Guayaquil	199	96,60%	96,60%	96,60%
Duran	6	2,91%	2,91%	2,91%
Yaguachi	1	0,49%	0,49%	100,00%
	206	100%		

Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Ilustración 22 Análisis de Ciudad - Voluntarios



Fuente: Datos de la Investigación

Autor: Dennys Cuzco

Análisis: La ilustración 22 demuestra que la categoría, en su mayoría pertenece a Guayaquil con más del 96%, diferencia significativa en comparación a las otras ciudades que muestran los registros. Estos datos serán usados para el entrenamiento del algoritmo.

Género: Se presentan la información en la tabla 14 de estadística y los histogramas de las estadísticas del campo género perteneciente al grupo de Voluntarios de la fundación “Ambulancia Deseo”. El considerar el género en los voluntarios al igual que los soñadores se debe al hincapié de relacionarlos por el mismo

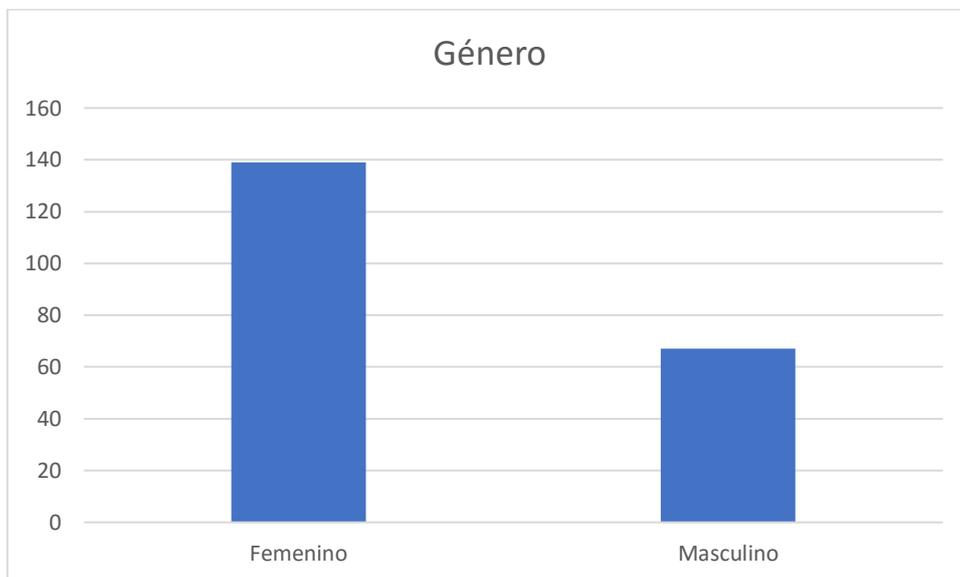
género, por ejemplo, una dama con una niña, un caballero con un niño.

Tabla 13 Análisis del Género - Voluntarios

	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Femenino	139	67,48%	67,48%	67,48%
Masculino	67	32,52%	32,52%	100%
Total	206	100%	100%	

Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Ilustración 23 Análisis del Género - Voluntarios



Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Análisis: La ilustración 23 nos demuestra que el género es un factor no modificable y que incide de forma significativa a la hora de asignación del soñador con un voluntario. Se aprecia que hay un 67,58% del género Femenino, es decir, el género Femenino

presenta un mayor interés comparado al género Masculino en actividades que cumplen las fundaciones como fundación “Ambulancia Deseo”, estos datos serán usados para el entrenamiento del algoritmo.

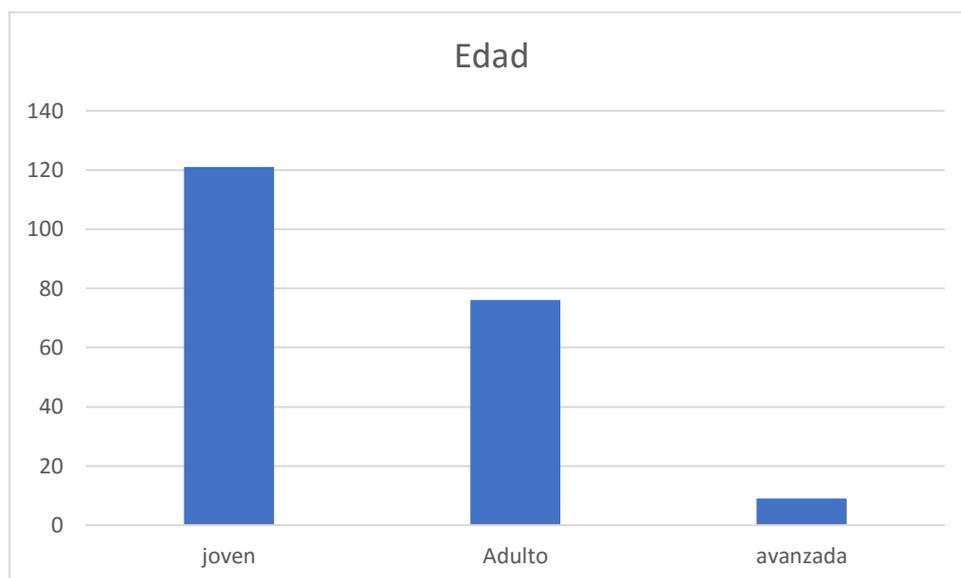
Edad: Se presentan la información en la tabla 15 de estadística y los histogramas de las estadísticas del campo Edad, que pertenece al grupo de Voluntarios de la fundación “Ambulancia Deseo”. La información se procesa con el rango previamente definido en el análisis de la edad en la sección de soñadores.

Tabla 14 Análisis de Edad - Voluntarios

	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Joven	121	58,74%	58,74%	58,74%
Adulto	76	36,89%	36,89%	36,89%
Avanzada	9	4,37%	4,37%	100%
	206	100,00%	100,00%	

Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Ilustración 24 Análisis de Edad – Voluntarios



Fuente: Datos de la Investigación
Autor: Dennys Cuzco

Análisis: La ilustración 24 demuestra que, con la asignación de rangos dentro de la información del campo edad, en la sección de Voluntarios, la mayor población son personas jóvenes, sobrepasando el 50%, seguido de la categoría adulto con un 36,89%. Aunque no es sorpresa para la fundación, existen voluntarios con edad avanza que apoyan aún en el objetivo de cumplir sueños vivenciales. Estos datos serán usados para el entrenamiento del algoritmo.

Limpeza de Datos

Previo a la creación del modelado es necesario la inspección respectiva de la data y así depurar los datos de la fundación, con

el propósito de evitar, datos inválidos, información inexacta, datos duplicados, información errónea, a continuación, desglosamos las acciones aplicadas en la data:

Eliminación de registros con campos en vacío: Los datos que no contenían información, se aplicó este tipo de limpieza.

Eliminación de registros no acordes al campo: Los datos que no contenían la información acorde al campo, se aplicó este tipo de limpieza.

Eliminación de registros duplicados: Los datos que contenían información repetida, independiente el campo que sea, se aplicó este tipo de limpieza.

3.4.4 Fase 4: Modelado

Finalizada la preparación y el depurado de los datos para ser procesados en el modelado, después, es necesario identificar que datos van a contribuir para la creación y evaluación del modelado donde se los vincula con los algoritmos.

Para comenzar a desarrollar el modelado por la propuesta del trabajo, es necesario escoger los algoritmos a utilizar, de igual forma, evaluar el mecanismo para la visualización de los resultados.

La información proporcionada por la fundación "Ambulancia Deseo" excede de los 500 registros, siendo esta, la información de los soñadores y los voluntarios. Este es el punto de partida para encontrar la relación entre soñador y voluntario, y, el motivo a desarrollarse este trabajo, que es, el encontrar una afinidad o relación en 2 variables (soñador y voluntario).

Si dentro de los algoritmos de clasificación existen 2 tipos de subclasificación, binaria y multiclase, y con la información proporcionada por la fundación, donde se identifica cuando SI y cuando NO hay una relación entre soñador y voluntario, podemos

concluir que, es aplicable la subclasificación binaria, por el resultado de verdadero o falso.

Por lo antes expuesto, se define que, mediante el uso de algoritmos dentro del Machine Learning es posible identificar si hay o no afinidad entre un soñador y un voluntario. Árboles de Decisiones será el algoritmo aplicado por su simplicidad y fácil interpretación.

Como resultado, la relación entre soñador y voluntario se define por verdadero o falso, siendo este:

Verdadero: Cuando SI existió una relación entre soñador y voluntario.

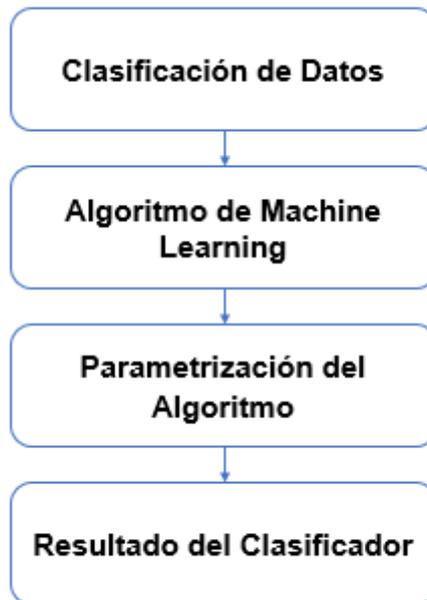
Falso: Cuando NO existió una relación entre soñador y voluntario.

Construcción de los modelos

Para la creación y evaluación del modelo es necesario establecer el flujo de desarrollo con el que se va a trabajar. Adicionalmente, se observará el comportamiento de los datos según en los algoritmos que se esté evaluando.

Flujo de Desarrollo para la implementación del Algoritmo

Ilustración 25 Diagrama de Flujo



Fuente: Datos de la Investigación
Autor: Dennys Cuzco

En la ilustración 25 se aprecia el flujo de desarrollo creado para la implementación del algoritmo de clasificación.

Procesamiento de Entrenamiento

La forma de trabajar el proceso de clasificación es, los datos a procesar deben de estar debidamente etiquetados. Analizando la información, es posible identificar la cantidad de clases que hay con sus respectivos ejemplos.

Al aplicar clasificación supervisada, el algoritmo construye un modelo. Del 100% de la información, se asigna un porcentaje de información al entrenamiento, esto significa que, el entrenamiento ayudará a predecir cuándo hay una relación entre un soñador y un voluntario.

Hardware

Para el desarrollo y uso del modelo se necesita de una computadora que posea mínimo las siguientes características:

Conexión a internet.

2 GB de memoria de RAM

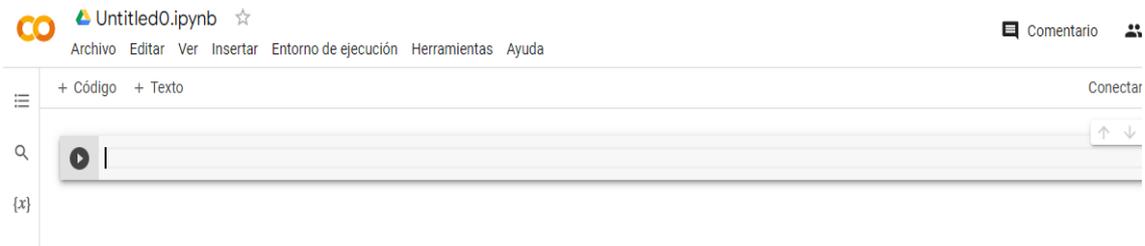
500 GB en Disco Duro

Navegador instalado, Chrome versión > 101; Firefox versión > 95

Ambiente de Trabajo

Usando el navegador de preferencia, se busca el ambiente de desarrollo que ofrece Google, Google Collab, <https://colab.research.google.com/?hl=es>. Dentro de la página de Collab, se apertura un nuevo proyecto, donde se asigna el nombre del proyecto y se inicia sesión con una cuenta de Gmail. El propósito de usar la cuenta de Gmail es guardar los proyectos dentro de la cuenta del Drive.

Ilustración 26 Nuevo Proyecto - Collab



Fuente: (Google Colaboratory, s. f.)
Autor: Google

La ilustración 26 muestra el inicio del proyecto donde se creará el modelo predictivo.

Ilustración 27 Levantamiento de Unidad

```
from google.colab import drive  
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive;
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

La ilustración 27 muestra donde se alojará el modelo a crearse, en este caso, dentro del Drive. Esta opción permite que el modelo sea probado, migrado por quien considera necesitarlo.

Ilustración 28 Importar Datos

```
import pandas as pd  
df = pd.read_excel('/content/sample_data/fundacionAmbulanciaDeseo.xlsx')
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 28 se aprecia el uso de la librería Panda, propia de Python, esta librería permite importar los archivos que se necesita para la predicción, y que será usado para el entrenamiento de modelo. En la variable DF se almacenará la información del archivo importado.

Ilustración 29 Previsualización de Datos cargados

```
df.sample(10)
```

	V_Ciudad	V_Sexo	V_Redad	S_Redad	S_Sexo	S_Categoria	Afinidad
84	Guayaquil	M	Adulto	menor	M	Bajo	1
184	Guayaquil	F	joven	menor	F	Bajo	1
133	Guayaquil	F	Adulto	joven	F	Medio	0
185	Guayaquil	M	Avanzada	Avanzada	M	Medio	0
48	Guayaquil	M	joven	joven	F	Bajo	1
98	Guayaquil	M	joven	Avanzada	M	Bajo	1
52	Guayaquil	M	joven	joven	M	Bajo	1
34	Guayaquil	F	joven	menor	F	Bajo	1
106	Guayaquil	F	joven	menor	F	Bajo	1
76	Guayaquil	F	joven	menor	F	Bajo	1

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 29 se aprecia parte de la información importada, la función Sample muestra aleatoriamente 10 registros. Esta función permite revisar que la información importada es la que se usar en el modelo.

Ilustración 30 Cambio de Variable

```
df = pd.get_dummies(data=df)
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 30, se da muestra el uso de la función `Get_Dummies`, la función permite eliminar la primera de las columnas generadas para cada característica codificada para evitar la denominada colinealidad (que una de las características sea una combinación lineal de las otras), lo que dificulta el correcto funcionamiento de los algoritmos. (La función `get_dummies` | *Interactive Chaos*, s. f.)

Ilustración 31 Variables Ajustadas

	Afinidad	V_Ciudad_Duran	V_Ciudad_Guayaquil	V_Ciudad_Yaguachi	V_Sexo_F	V_Sexo_M	V_Redad_Adulto	V_Redad_Avanzada
0	1	0	1	0	0	1	0	0
1	0	0	1	0	1	0	0	0
2	1	0	1	0	1	0	1	0
3	1	0	1	0	1	0	0	0
4	1	0	1	0	1	0	0	0
...
201	0	0	1	0	1	0	1	0
202	0	0	1	0	1	0	0	0
203	1	0	1	0	1	0	0	0
204	0	0	1	0	1	0	0	0
205	0	0	1	0	1	0	0	0

206 rows x 17 columns

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 31 se aprecia el cambio aplicado en las variables, el objetivo de usar esta función es de cambiar el tipo de variable,

de nominal a ordinal. La información que posee el archivo de la fundación contiene datos que la máquina no puede intuir, este inconveniente es superado gracias al uso de la fundación `get_dummies`.

Importada la información en el Collab, donde se está armando el modelo, se procede a separar las variables, tanto la explicativa como la variable objetivo. Con la identificación de las variables se comienza a preparar los parámetros que usará el algoritmo.

Ilustración 32 Distinción de Variables

```
explicativas = df.drop(columns='Afinidad')  
objetivo = df.Afinidad
```

Fuente: Datos de Investigación

Autor: Dennys Cuzco

En la ilustración 32 se aprecia la identificación de las variables a usarse en el modelo, y son las siguientes:

Explicativa: Los campos que construyen las diversas clases que se necesita que el modelo aprenda para predecir.

Objetivo: El resultado esperado por las clases, determinado por las variables explicativas.

Ilustración 33 Entrenamiento

```
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(explicativas, objetivo, test_size = 0.20)
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 33 se aprecia el uso de la función `train_test_split` que permite el inicio del entrenamiento, como hiper parámetros asignamos lo siguiente:

Variable X.- Variable Explicativa.

Variable Y.- Variable Objetivo

Tamaño del Test. – La cantidad de datos que se usará para el testeo, reflejada en porcentaje, es decir, el 0.20 representa el uso del 20% de los todos datos. Empíricamente el 0.80% restante se usará para el entrenamiento.

Algoritmo ID3 - Entropy

Ilustración 34 Algoritmo de Árbol de Decisión

```
from re import X
from sklearn.tree import DecisionTreeClassifier

#creamos el arbol de decision.
model = DecisionTreeClassifier(criterion = 'entropy')
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 34 se inicia la creación del algoritmo, se especifica inicialmente la librería sklearn para aplicar el algoritmo de Árbol de Decisión, como hyper parámetro se especifica el tipo, siendo este caso, la entropía. La variable model, contendrá la estructura del modelo de árbol de decisión.

Ilustración 35 Ajuste del Modelo

```
model.fit(X=explicativas,y=objetivo)
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 35 se especifica el uso de la función fit, la función permite realizar ajuste del algoritmo previamente inicializado, es decir, árbol de decisión bajo el criterio de entropía.

Ilustración 36 Visualizar Modelo – Árbol de Decisión

```
from sklearn.tree import plot_tree
import matplotlib.pyplot as plt

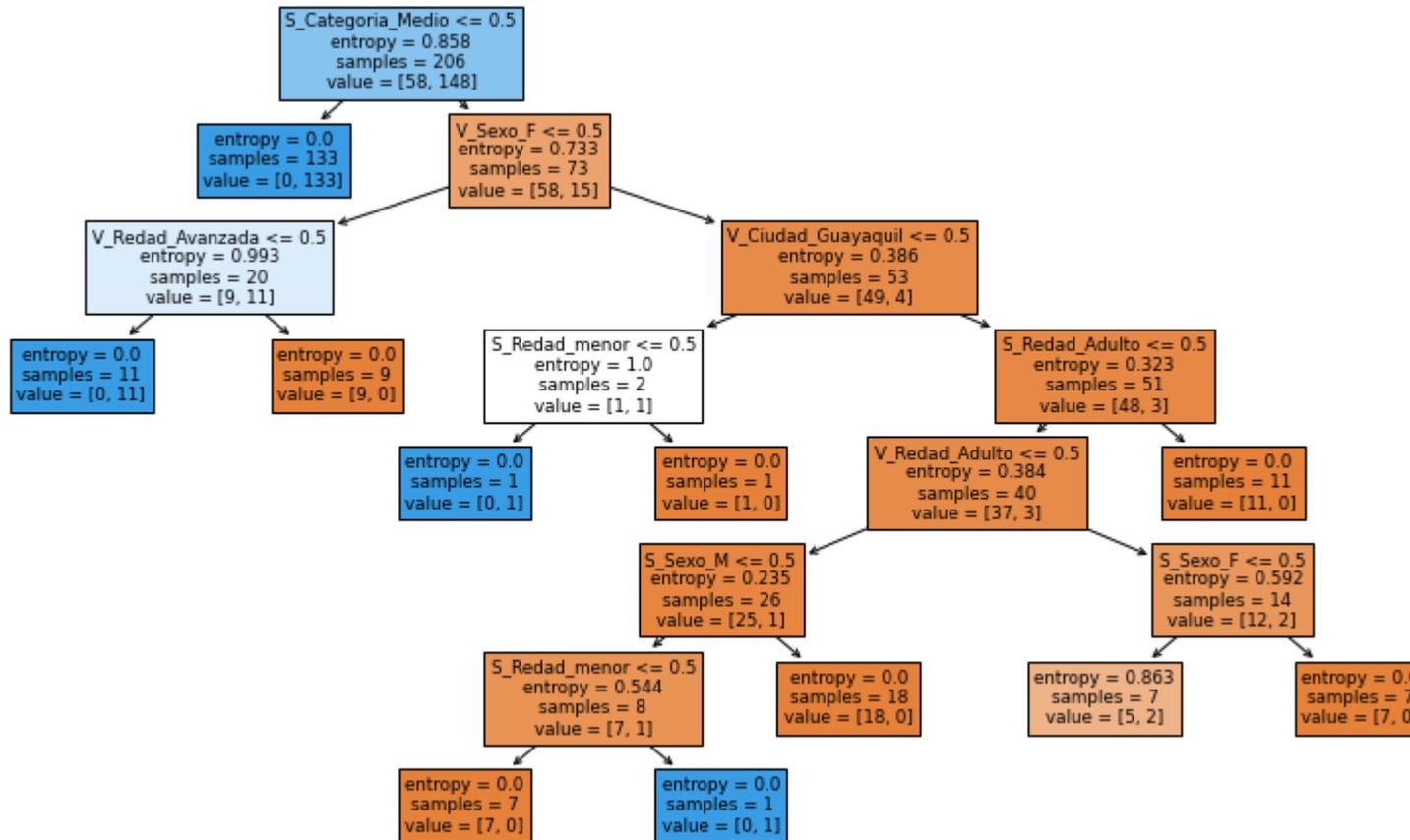
plt.figure(figsize=(14,8))
plot_tree(decision_tree=model,feature_names=explicativas.columns,filled=True,fontsize=7);
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 36 se grafica el modelo en forma de árbol, la biblioteca `matplotlib` permite la visualización y facilita la comprensión de la estructura del algoritmo empleado en el modelo.

La importancia de analizar de forma visual el algoritmo es para comprender el su funcionamiento y posteriormente poder realizar comparaciones con otros hyper parámetros o con otros algoritmos de clasificación.

Ilustración 37 Modelo ID3- Entropy



Fuente: Datos de Investigación
 Autor: Dennys Cuzco

En la ilustración 37 se visualiza el árbol de decisión que armó el algoritmo por la información que recibió. El algoritmo de árboles de decisiones con el hyper parámetro entropía, arma el modelo según la homogeneidad de los datos.

La entropía arma el modelo según la homogeneidad de los datos, como se puede apreciar el árbol inicia por el campo categoría tipo Medio, es decir el que tiene menor entropía, después se estructura diferentes (ramas) siendo estas las soluciones creadas por el algoritmo, los nodos que se encuentran al final (hojas) es lo que permite al algoritmo predecir lo que estamos buscando.

3.4.5 Fase 5: Evaluación

Para esta fase se inicia con una evaluación genérica de los diversos algoritmos que posee machine Learning en la categoría de aprendizaje supervisado.

Ilustración 38 Análisis de los Algoritmos

```
# Spot Check Algorithms
models = []
models.append(('ArbolesDecisiones-Entropy', DecisionTreeClassifier(criterion = 'entropy')))
models.append(('ArbolesDecisiones-Gini', DecisionTreeClassifier(criterion = 'gini')))
models.append(('Random Forest', RandomForestClassifier()))
models.append(('Regresión Logística', LogisticRegression()))
models.append(('Análisis discriminante lineal', LinearDiscriminantAnalysis()))
models.append(('KNN (vecino cercano)', KNeighborsClassifier()))
models.append(('Naive Bayes', GaussianNB()))
models.append(('SVM', SVC()))

# evaluate each model in turn
results = []
names = []
for name, model in models:
    kfold = model_selection.KFold(n_splits=10) #, random_state=seed
    cv_results = model_selection.cross_val_score(model, X_train, Y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 38 se aprecia el uso de los algoritmos del aprendizaje supervisado sin la especificación de hyper parámetros, es decir, sin aplicar ajustes para determinar el comportamiento de se busca en el algoritmo.

Ilustración 39 Resultado de los Algoritmos

ArbolesDecisiones-Entpy: 0.987868
ArbolesDecisiones-Gini: 0.987868
Random Forest: 0.987868
Regresión Logística: 0.987868
KNN (vecino cercano): 0.975000
Naive Bayes: 0.933088
SVM: 0.987868

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 39 se evidencia que es posible aplicar diferentes algoritmos para alcanzar una buena predicción. En su mayoría los algoritmos presentan más de una 95% como resultado final, es decir, la información que recibe el algoritmo es posible predecirlo con una efectividad mayor del 95%.

El algoritmo KNN (vecinos cercanos) posee un porcentaje de efectividad del 97.50% siendo menor por un 0.01% en comparación a los otros algoritmos, esto se debe al comportamiento que cubre el algoritmo, siendo este, el algoritmo inicia seleccionando el valor K y los datos que se acerquen al valor de K será la predicción buscada.

El algoritmo Naive Bayes es el que posee el menor porcentaje en la comparación de los algoritmos, siendo su porcentaje del 93.30%. Su porcentaje demuestra que el comportamiento del algoritmo al

procesar los datos afecta su predicción debido a que el algoritmo aplica independencia condicional en cada clase.

En la sección de anexos se encuentra el código completo de la Comparativa de Algoritmos de Aprendizaje Supervisado con el propósito de facilitar su uso.

Matriz de Confusión

Ilustración 40 Matriz de Confusión

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix

#class_names = df.target_names
class_names = ['Real', 'Predicción']
# crear matriz de confusión
matrix = confusion_matrix(Y_test, target_predicted)

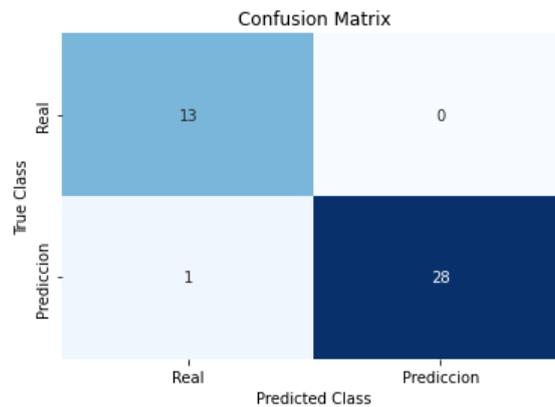
# crear marco de datos de pandas Crear un conjunto de datos
dataframe = pd.DataFrame(matrix, index=class_names, columns=class_names)

# crear mapa de calor dibujar mapa de calor
sns.heatmap(dataframe, annot=True, cbar=None, cmap="Blues")
plt.title("Confusion Matrix") #, plt.tight_layout()
plt.ylabel("True Class"), plt.xlabel("Predicted Class")
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 40 se aprecia la elaboración de la matriz de confusión, esta matriz permite visualizar el desempeño del algoritmo elegido, es decir, el algoritmo de árbol de decisiones – entropía.

Ilustración 41 Matriz Confusión ID3 - Entropia



Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 41 se aprecia que la matriz de confusión posee verdaderos positivos, verdaderos negativos y falsos positivos (mal clasificado). No posee falsos negativos. La cantidad de registros identificados dentro de la matriz corrobora con la alta efectividad del algoritmo elegido.

Se detalla la información de los 3 casos identificados:

Verdaderos Positivos (13): Hace referencia a la data de testeo, donde un total de 13 registros, el algoritmo los identificó como positivos y la respuesta fue correcta.

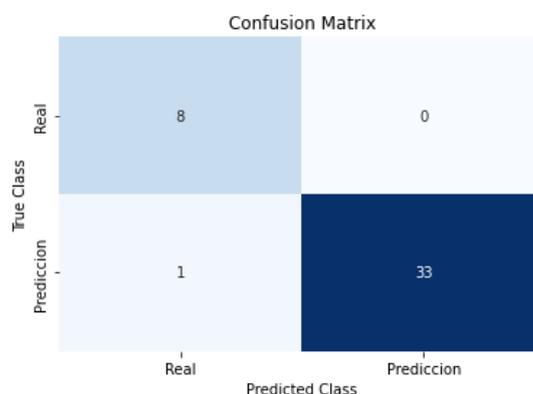
Verdaderos Negativos (28): Hace referencia a la data de testeo, donde un total de 28 registros, el algoritmo los identificó como negativos y la respuesta fue correcta.

Falso Positivo (1): Hace referencia a la data de testeo, donde un registro el algoritmo lo identificó como positivo y su respuesta no fue correcta, es decir, la respuesta era falsa.

La cantidad de registro es de 206 en total, al aplicar el 20% para el testeo, ese 20% representa un total de 42 registro, es decir, si sumamos $13 + 28 + 1$ da un total de 42 registros. Concluimos que los registros detectados en la matriz coinciden con el porcentaje de registros especificados en la fase de entrenamiento del algoritmo.

En la sección de anexos se encuentra el código completo del modelo predictivo (Gini y Entropía) con el propósito de facilitar su uso.

Ilustración 42 Matriz Confusión ID3 - Gini



Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 42 se aprecia la matriz de confusión para el análisis del desempeño del algoritmo de árbol de decisiones – Gini. Esta matriz es similar a la matriz de algoritmo de árbol de decisiones – entropía. Lo que corrobora que los resultados de datos a predecir serán similares aun usando diferentes tipos de algoritmos de aprendizaje supervisado.

Se detalla la información de los 3 casos identificados en la matriz de árbol de decisiones - Gini:

Verdaderos Positivos (8): Hace referencia a la data de testeo, donde un total de 8 registros, el algoritmo los identificó como positivos y la respuesta fue correcta.

Verdaderos Negativos (33): Hace referencia a la data de testeo, donde un total de 33 registros, el algoritmo los identificó como negativos y la respuesta fue correcta.

Falso Positivo (1): Hace referencia a la data de testeo, donde un registro el algoritmo lo identificó como positivo y su respuesta no fue correcta, es decir, la respuesta era falsa.

Algoritmo Random Forest

Ilustración 43 Algoritmo - Random Forest

```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(criterion = 'entropy', random_state = 42)
rfc.fit(X_train, Y_train)

# Evaluating on Training set
rfc_pred_train = rfc.predict(X_train)
print('Training Set Evaluation F1-Score=>', f1_score(Y_train, rfc_pred_train))

Training Set Evaluation F1-Score=> 0.9959183673469388
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 43, se demuestra el diseño del algoritmo Random Forest para evaluar su comportamiento y realizar la respectiva comparación con los dos previos algoritmos evaluados.

Ilustración 44 Visualizar Modelo - Random Forest

```
import matplotlib.pyplot as plt
import numpy as np
feature_importance=pd.DataFrame({
    'rfc':rfc.feature_importances_,
    'dt':model.feature_importances_
}),index=df.drop(columns=['Afinidad']).columns
feature_importance.sort_values(by='rfc',ascending=True,inplace=True)

index = np.arange(len(feature_importance))
fig, ax = plt.subplots(figsize=(18,8))
rfc_feature=ax.barh(index,feature_importance['rfc'],0.4,color='purple',label='Random Forest')
dt_feature=ax.barh(index+0.4,feature_importance['dt'],0.4,color='lightgreen',label='Decision Tree')
ax.set(yticks=index+0.4,yticklabels=feature_importance.index)

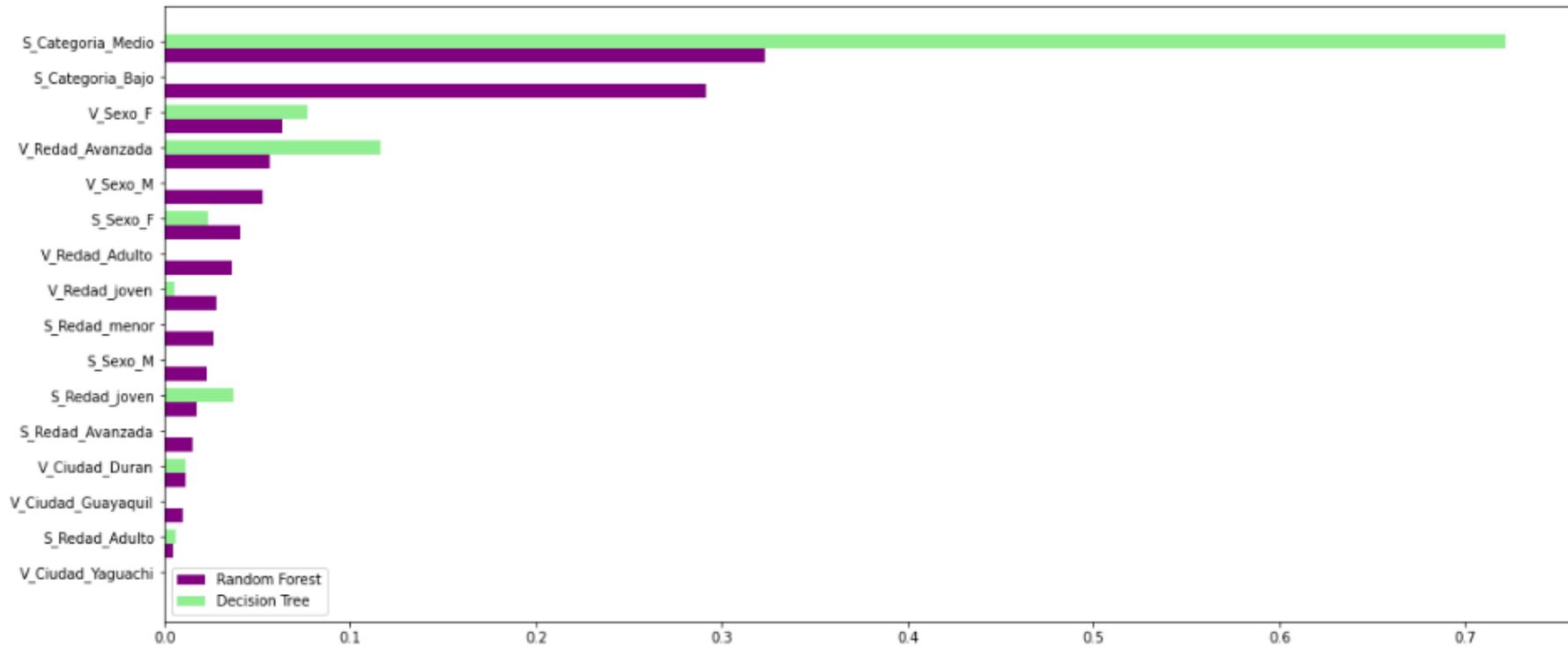
ax.legend()
plt.show()
```

Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 44, se demuestra el uso de información referente al algoritmo de Random Forest, para visualizar el modelo con una gráfica diferente.

En la sección de anexos se encuentra el código completo del modelo predictivo (Random Forest) con el propósito de facilitar su uso.

Ilustración 45 Comparativa



Fuente: Datos de Investigación
Autor: Dennys Cuzco

En la ilustración 45 se muestra la gráfica del algoritmo Random Forest. El algoritmo priorizo su estructura por el campo categoría tipo medio, seguido del campo género de los voluntarios tipo Femenino. Evaluando ambos gráficos, es posible afirmar que la estructura de ambos algoritmos inicia de igual forma. Los campos restantes poseen una pequeña variación, pero se debe al enfoque que brinda el uso de cada algoritmo.

Aunque el algoritmo Random Forest, posee una estructura diferente al momento de clasificar, en comparación de los árboles de decisiones, es posible afirmar la gran similitud que existe a la hora de predecir.

Como métricas tenemos los siguientes resultados, dentro del modelo usado.

```
accuracy 0.94 (+/- 0.10)
balanced_accuracy 0.70 (+/- 0.49)
precision 0.94 (+/- 0.10)
recall 1.00 (+/- 0.00)
f1 0.97 (+/- 0.05)
```

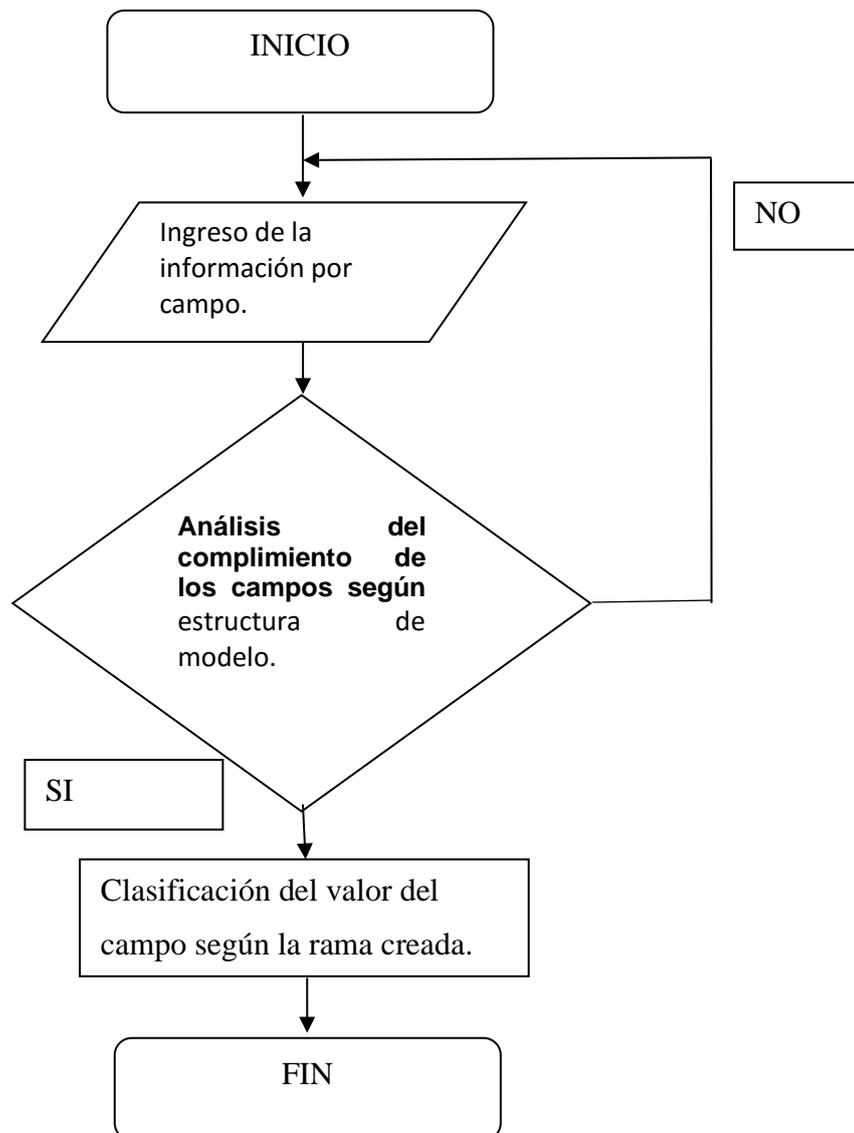
Con los resultados es posible manifestar lo siguiente:

El modelo presenta un 94% de precisión.

La predicción de los verdaderos positivos se dio en un 100%, es decir, la identificación se dio sin errores.

Con las anteriores métricas detalladas, el score F1 brinda como resultado un 97%, siendo este un porcentaje que nos demuestra la alta efectividad del modelo.

Diagrama de Flujo del Modelo Predictivo ID



Conclusiones y Recomendaciones

Conclusiones

En el presente apartado, se menciona las conclusiones por la elaboración del modelo predictivo, además de recomendaciones.

Al identificar los factores que influyen en la relación entre soñador y voluntario se logró evidenciar la repercusión que cumple cada factor, lo que permitió estructurar la información imprescindible para el aprendizaje del modelo.

Luego de identificar los factores se procede con el análisis de las diversas metodologías existentes, lo que conlleva al uso de la metodología CRISP-DM porque integra las tareas necesarias para proyectos predictivos logrando así un análisis completo en el proceso de cumplir un sueño.

Finalmente, con el diseño y ensayo del modelo predictivo se evidencia la efectividad del uso de árboles de decisiones para este proyecto. El modelo creado queda apto para la implementación como herramienta tecnológica y brindará un apoyo significativo por su alto porcentaje de predicción en el proceso de toma de decisiones en la relación entre un soñador y un voluntario en la fundación "Ambulancia Deseo".

Recomendaciones

En este apartado se indica los consejos a tomar en cuenta para el desempeño y uso del proyecto desarrollado.

Si el modelo se desea aplicar en otras instituciones o fundaciones, es de vital importancia evaluar la situación actual de la institución y concluir que se ajustan a las fases descritas en el proyecto.

Al obtener la información de la fundación “Ambulancia Deseo” se evidenció la falta de herramientas tecnológicas que le permitan crear bases de datos con márgenes bajos de errores para el mantenimiento de la data, por lo que se recomienda un análisis completo de los datos antes de iniciar los procesos que cubre el proyecto.

Debido al manejo de información personal sensible, por políticas de privacidad, se recomienda antelar la solicitud de la información que se necesite.

Se recomienda que el modelo sea manejado bajo las herramientas tecnológicas descrita en el Capítulo III, Fase de Modelado sección Hardware para alcanzar los resultados que ofrece el desarrollo de este proyecto.

ANEXOS

Código - Modelo Predictivo (Ganancia – Entropía)

```
#dirección donde se alojan los archivos
from google.colab import drive
drive.mount('/content/drive')

#importar datos
import pandas as pd
df = pd.read_excel('/content/sample_data/fundacionAmbulanciaDeseo.xlsx')

#Cambio de variable nominal a ordinal
df = pd.get_dummies(data=df)

#identificación de variables
explicativas = df.drop(columns='Afinidad')
objetivo = df.Afinidad

#división de los datos, entrenamiento y test
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(explicativas, objetivo, test_size = 0.20)

# Construir Arbol de Decisión
from re import X
from sklearn.tree import DecisionTreeClassifier

#creamos el arbol de decisión (gini - entropy).
model = DecisionTreeClassifier(criterion = 'gini')

#Ajustamos el modelo.
model.fit(X=explicativas,y=objetivo)

# Predecir la respuesta para el conjunto de datos de prueba.
target_predicted = model.predict(X_test)

print(model.score(X_test,Y_test))
DecisionTreeClassifier()
```

```
#Importar librerías para la creación del modelo
from sklearn.tree import plot_tree
import matplotlib.pyplot as plt

#configuración de la imagen, expandimos
plt.figure(figsize=(14,8))
plot_tree(decision_tree=model,feature_names=explicativas.columns,filled=True);

#importación de librerías para creación de matriz
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import confusion_matrix

#distinción de datos reales y predicción
class_names = ['Real','Predicción']

# crear matriz de confusión
matrix = confusion_matrix(Y_test, target_predicted)

# crear marco de datos de pandas Crear un conjunto de datos
dataframe = pd.DataFrame(matrix, index=class_names, columns=class_names)

# asignación de estilo al mapa.
sns.heatmap(dataframe, annot=True, cbar=None, cmap="Blues")
plt.title("Confusion Matrix") #, plt.tight_layout()
plt.ylabel("True Class"), plt.xlabel("Predicted Class")
```

Código - Modelo Predictivo (Random Forest)

```
from sklearn.metrics import f1_score
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(criterion = 'entropy', random_state = 42)
rfc.fit(X_train, Y_train)

# Evaluating on Training set
rfc_pred_train = rfc.predict(X_train)
print('Training Set Evaluation F1-Score=>', f1_score(Y_train, rfc_pred_train))

# Entrenamiento con evaluación de F1-score
rfc_pred_test = rfc.predict(X_test)
print('Testing Set Evaluation F1-Score=>', f1_score(Y_test, rfc_pred_test))

import matplotlib.pyplot as plt
import numpy as np
feature_importance=pd.DataFrame({
    'rfc':rfc.feature_importances_,
    'dt':model.feature_importances_
},index=df.drop(columns=['Afinidad']).columns)
feature_importance.sort_values(by='rfc',ascending=True,inplace=True)

index = np.arange(len(feature_importance))
fig, ax = plt.subplots(figsize=(18,8))
rfc_feature=ax.barh(index,feature_importance['rfc'],0.4,color='purple',label='Random Forest')
dt_feature=ax.barh(index+0.4,feature_importance['dt'],0.4,color='lightgreen',label='Decision Tree')
ax.set(yticks=index+0.4,yticklabels=feature_importance.index)

ax.legend()
plt.show()
```

Comparativa de Algoritmos de Aprendizaje Supervisado

```
import pandas as pd
from pandas.plotting import scatter_matrix
import matplotlib.pyplot as plt
from sklearn import model_selection
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC

import pandas as pd
df = pd.read_excel('/content/sample_data/fundacionAmbulanciaDeseo.xlsx')

#visualizar la cantidad de información
print(df.shape)

#definición de variables
from sklearn.model_selection import train_test_split
df = pd.get_dummies(data=df)
explicativas = df.drop(columns='Afinidad')
objetivo = df.Afinidad

#entrenamiento
validation_size = 0.20
seed = 7
X_train, X_validation, Y_train, Y_validation = model_selection.train_test_split(explicativas, objetivo, test_size=validation_size, random_state = 0)

# Testeo evaluación de metricas
seed = 7
scoring = 'accuracy'
```

```

# Spot Check Algorithms
models = []
models.append(('ArbolesDecisiones-
Entpy', DecisionTreeClassifier(criterion = 'entropy')))
models.append(('ArbolesDecisiones-
Gini', DecisionTreeClassifier(criterion = 'gini')))
models.append(('Random Forest', RandomForestClassifier()))
models.append(('Regresión Logística', LogisticRegression()))
models.append(('KNN (vecino cercano)', KNeighborsClassifier()
))
models.append(('Naive Bayes', GaussianNB()))
models.append(('SVM', SVC()))

# resultado de evaluación por algoritmos
results = []
names = []
for name, model in models:
    kfold = model_selection.KFold(n_splits=10) #, random_stat
e=seed
    cv_results = model_selection.cross_val_score(model, X_tra
in, Y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_result
s.std())
    print(msg)

```

BIBLIOGRAFÍA

300 mil casos de enfermedades raras se registran en Ecuador. (s. f.).

Recuperado 9 de junio de 2022, de

<https://www.edicionmedica.ec/secciones/salud-publica/300-mil-casos-de-enfermedades-raras-hay-en-el-ecuador--87269>

1260-TESIS-ALVAREZ Y MERCHAN.pdf. (s. f.). Recuperado 18 de abril de

2022, de <http://repositorio.ug.edu.ec/bitstream/redug/36062/1/1260->

TESIS-

[ALVAREZ%20Y%20MERCHAN.pdf#page=33&zoom=100,148,273](http://repositorio.ug.edu.ec/bitstream/redug/36062/1/1260-ALVAREZ%20Y%20MERCHAN.pdf#page=33&zoom=100,148,273)

Acuerdo 2018-040.pdf. (s. f.). Recuperado 28 de mayo de 2022, de

<https://www.gob.ec/sites/default/files/regulations/2019->

[04/Acuerdo%202018-040.pdf](https://www.gob.ec/sites/default/files/regulations/2019-04/Acuerdo%202018-040.pdf)

admin. (s. f.). Nosotros. *Ambulancia Deseo Ecuador.* Recuperado 22 de agosto

de 2021, de <https://ambulanciadeseo.org.ec/nosotros/>

Aprendizaje Inductivo: Árboles de Decisión—Fernando Sancho Caparrini. (s. f.).

Recuperado 27 de mayo de 2022, de

<http://www.cs.us.es/~fsancho/?e=104>

Árbol de decisión (modelo de clasificación ID3). (2020). En *Wikipedia, la enciclopedia libre.*

https://es.wikipedia.org/w/index.php?title=%C3%81rbol_de_decisi%C3%

[B3n_\(modelo_de_clasificaci%C3%B3n_ID3\)&oldid=131555722](https://es.wikipedia.org/w/index.php?title=%C3%81rbol_de_decisi%C3%B3n_(modelo_de_clasificaci%C3%B3n_ID3)&oldid=131555722)

Árboles de Decisión Clasificación – Teoría. (2019, octubre 18). □ *Aprende IA.*

<https://aprendeia.com/arboles-de-decision-clasificacion-teoria-machine-learning/>

- Calvo, D. (2017, julio 18). Aprendizaje Supervisado y No supervisado. *Diego Calvo*. <https://www.diegocalvo.es/aprendizaje-supervisado-y-no-supervisado/>
- Calvo, D. (2019, marzo 23). Aprendizaje supervisado. *Diego Calvo*. <https://www.diegocalvo.es/aprendizaje-supervisado/>
- Categorías de la IA - Inteligencia Artificial*. (s. f.). Recuperado 26 de mayo de 2022, de <https://sites.google.com/site/sitwebia/categorias-de-la-inteligencia-artificial>
- Clasificación de Aprendizaje automático supervisado. (2019a, diciembre 24). *sitiobigdata.com*. <https://sitiobigdata.com/2019/12/24/clasificacion-de-aprendizaje-automatico-supervisado/>
- Clasificación de Aprendizaje automático supervisado. (2019b, diciembre 24). *sitiobigdata.com*. <https://sitiobigdata.com/2019/12/24/clasificacion-de-aprendizaje-automatico-supervisado/>
- CRISP-DM, still the top methodology for analytics, data mining, or data science projects. (s. f.). *KDnuggets*. Recuperado 23 de mayo de 2022, de <https://www.kdnuggets.com/crisp-dm-still-the-top-methodology-for-analytics-data-mining-or-data-science-projects.html/>
- Dialnet-InteligenciaEmocionalYBienestarII-655308.pdf*. (s. f.).
- Fundación Ambulancia del Deseo. (2021). En *Wikipedia, la enciclopedia libre*. https://es.wikipedia.org/w/index.php?title=Fundaci%C3%B3n_Ambulancia_del_Deseo&oldid=140585324
- Garcés, L. M. (s. f.). *Decreto Ejecutivo 194 Registro Oficial Suplemento 109 de 27-oct.-2017 Estado: Vigente*. 10.

- Gassó, R. (2019, diciembre 28). *Una ambulancia y un equipo médico para cumplir deseos de los pacientes: «Nuestro sistema descuida las emociones»*. ElDiario.es. https://www.eldiario.es/desalambre/deseos-sanitarios-medicos-ambulancias_1_1177672.html
- Gonzalez, L. (2018a, marzo 22). Aprendizaje Supervisado: Logistic Regression. □ *Aprende IA*. <https://aprendeia.com/aprendizaje-supervisado-logistic-regression/>
- Gonzalez, L. (2018b, marzo 23). Aprendizaje Supervisado: Random Forest Classification. □ *Aprende IA*. <https://aprendeia.com/aprendizaje-supervisado-random-forest-classification/>
- Gonzalez, L. (2019, agosto 16). Máquinas Vectores de Soporte Clasificación – Teoría. □ *Aprende IA*. <https://aprendeia.com/maquinas-vectores-de-soporte-clasificacion-teoria/>
- Gonzalez, L. (2020, febrero 11). Aprendizaje no Supervisado. □ *Aprende IA*. <https://aprendeia.com/aprendizaje-no-supervisado-machine-learning/>
- Google Colaboratory. (s. f.). Recuperado 28 de mayo de 2022, de <https://colab.research.google.com/?hl=es>
- Grupos Vulnerables. (s. f.). Recuperado 16 de abril de 2022, de http://archivos.diputados.gob.mx/Centros_Estudio/Cesop/Comisiones/d_gvulnerables.htm
- K Vecinos más Cercanos—Teoría. (2019, julio 19). □ *Aprende IA*. <https://aprendeia.com/k-vecinos-mas-cercanos-teoria-machine-learning/>
- La función `get_dummies` | Interactive Chaos. (s. f.). Recuperado 28 de mayo de 2022, de <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/la-funcion-getdummies>

La fundación que cumple los últimos deseos de quienes están a punto de morir.

(2015, octubre 1). BBC News Mundo.

https://www.bbc.com/mundo/noticias/2015/10/150928_sociedad_ultimos_deseos_fundacion_muerte_holanda_jirafa_mr

La metodología CRISP-DM en ciencia de datos—IIC. (2021, noviembre 29).

Instituto de Ingeniería del Conocimiento.

<https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/>

Mesicic4_ecu_const.pdf. (s. f.). Recuperado 28 de mayo de 2022, de

https://www.oas.org/juridico/pdfs/mesicic4_ecu_const.pdf

Nuevo reglamento para el otorgamiento de personalidad jurídica de

organizaciones sociales. (2017, noviembre 6). Naranjo Martínez &

Subía. <https://nmsslaw.com.ec/reglamento-personalidad-juridica-organizaciones-sociales/>

Paola, D. S. D. (s. f.). *Bienestar Emocional.* 72.

¿Qué es la inteligencia artificial (IA)? (s. f.). Recuperado 19 de abril de 2022, de

<https://www.oracle.com/mx/artificial-intelligence/what-is-ai/>

Ricaurte, W. A. G., Chedraui, M. A. D., Guamán, J. C. A., Díaz, N. J. G., Jaime,

P. I. H., Morante, M. A. O., Muñoz, J. S. P., Cedeño, G. D. S., Sánchez,

Á. R. S., Flores, W. V. V., Rodríguez, C. A. V., & Reinoso, R. I. V. (s. f.).

ASAMBLEÍSTAS QUE INTEGRAN LA COMISIÓN: 56.

Sesion5_Metodologias.pdf. (s. f.). Recuperado 9 de junio de 2022, de

https://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf

Vera, S. A. P. (s. f.). *Técnicas Semi Supervisadas Aplicado a.* 54.

