



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática

Escuela Profesional de Ingeniería de Sistemas

**Sistema informático basado en analytics para la
prevención y monitoreo de casos de violencia contra la
mujer registrados en los Centros de Emergencia
Mujer, basado en indicadores de la norma ISO 37120**

TESIS

Para optar el Título Profesional de Ingeniero de Sistemas

AUTOR

Joselyn Mayté CCANTO BRAVO

José Antonio AIRA RETUERTO

ASESOR

Mg. Jorge Leoncio GUERRA GUERRA

Lima, Perú

2023



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Ccanto, J. & Aira, J. (2023). *Sistema informático basado en analytics para la prevención y monitoreo de casos de violencia contra la mujer registrados en los Centros de Emergencia Mujer, basado en indicadores de la norma ISO 37120.* [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Profesional de Ingeniería de Sistemas]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios autor/ asesor

Datos de autor 1	
Nombres y apellidos	Joselyn Mayté Ccanto Bravo
Tipo de documento de identidad	DNI
Número de documento de identidad	71645321
URL de ORCID	https://orcid.org/0000-0003-3193-6891
Datos de autor 2	
Nombres y apellidos	José Antonio Aira Retuerto
Tipo de documento de identidad	DNI
Número de documento de identidad	70558864
URL de ORCID	https://orcid.org/0000-0001-7582-3157
Datos de asesor	
Nombres y apellidos	Jorge Leoncio Guerra Guerra
Tipo de documento de identidad	DNI
Número de documento de identidad	08473333
URL de ORCID	https://orcid.org/0000-0003-2392-7821
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Fanny Sobero Rodríguez
Tipo de documento	DNI
Número de documento de identidad	20120467
Miembro del jurado 1	
Nombres y apellidos	Félix Armando Fermín Pérez
Tipo de documento	DNI
Número de documento de identidad	08736347
Datos de investigación	
Línea de investigación	Gobierno y Gestión de TIC

Grupo de investigación	Internet de las Cosas
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Latitud: -12.0431800 Longitud: -77.0282400
Año o rango de años en que se realizó la investigación	2017 - 2021
URL de disciplinas OCDE	Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04 Otras ingenierías y tecnologías https://purl.org/pe-repo/ocde/ford#2.11.02



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Profesional de Ingeniería de Sistemas

Acta Virtual de Sustentación de Tesis

Siendo las 16:00 horas del día 18 de enero del año 2023, se reunieron virtualmente los docentes designados como miembros de Jurado de Tesis, presidido por la Mg. Fanny Sobero Rodríguez, Ing. Félix Armando Fermín Pérez (Miembro) y el Mg Jorge Leoncio Guerra Guerra (Miembro Asesor), usando la plataforma Meet (meet.google.com/kyi-zeev-pdy), para la sustentación Virtual de la tesis Intitulada:

“SISTEMA INFORMÁTICO BASADO EN ANALYTICS PARA LA PREVENCIÓN Y MONITOREO DE CASOS DE VIOLENCIA CONTRA LA MUJER REGISTRADOS EN LOS CENTROS DE EMERGENCIA MUJER, BASADO EN INDICADORES DE LA NORMA ISO 37120”, de la Bachiller: Ccanto Bravo, Joselyn Mayté; para obtener el Titulo Profesional de Ingeniero de Sistemas.

Acto seguido de la exposición de la Tesis, el Presidente invitó a la Bachiller a responder las preguntas formuladas por los Miembros del Jurado.

La Bachiller, en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez las preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, la bachiller obtuvo la nota de 17 (diecisiete)

A continuación, la Presidenta del Jurado Mg. Fanny Sobero Rodríguez, declara a la Bachiller **Ingeniero de Sistemas**.

Siendo 17:00 horas, se levantó la sesión.

Mg. Fanny Sobero Rodríguez
Presidente

Ing. Félix Armando Fermín Pérez
Miembro

Mg Jorge Leoncio Guerra Guerra
Miembro Asesor



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Profesional de Ingenieria de Sistemas

Acta Virtual de Sustentación de Tesis

Siendo las 16:00 horas del dia 18 de enero del año 2023, se reunieron virtualmente los docentes designados como miembros de Jurado de Tesis, presidido por la Mg. Fanny Sobero Rodríguez, Ing. Félix Armando Fermín Pérez (Miembro) y el Mg Jorge Leoncio Guerra Guerra (Miembro Asesor), usando la plataforma Meet (meet.google.com/kyi-zeev-pdy), para la sustentación Virtual de la tesis Intitulada:

“SISTEMA INFORMÁTICO BASADO EN ANALYTICS PARA LA PREVENCIÓN Y MONITOREO DE CASOS DE VIOLENCIA CONTRA LA MUJER REGISTRADOS EN LOS CENTROS DE EMERGENCIA MUJER, BASADO EN INDICADORES DE LA NORMA ISO 37120”, del Bachiller: **Aira Retuerto, José Antonio**; para obtener el Título Profesional de Ingeniero de Sistemas.

Acto seguido de la exposición de la Tesis, el Presidente invitó al Bachiller a responder las preguntas formuladas por los Miembros del Jurado.

El Bachiller, en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez las preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, el bachiller obtuvo la nota de 17 (diecisiete)

A continuación, la Presidenta del Jurado Mg.Fanny Sobero Rodríguez, declara al Bachiller **Ingeniero de Sistemas**.

Siendo 17:00 horas, se levantó la sesión.

Mg. Fanny Sobero Rodríguez
Presidente

Ing. Félix Armando Fermín Pérez
Miembro

Mg Jorge Leoncio Guerra Guerra
Miembro Asesor



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, DECANA DE AMÉRICA
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Profesional de Ingeniería de Sistemas

INFORME DE EVALUACIÓN DE ORIGINALIDAD

1. Facultad de Ingeniería de Sistemas e Informática
2. Escuela Profesional de Ingeniería de Sistemas
3. Autoridad académica que emite el informe de originalidad

Directora (e) de la EPIS

4. Apellidos y Nombres de la autoridad académica

Dra. Luzmila E. Pró Concepción

5. Operador del programa informático de similitudes

Dra. Luzmila E. Pró Concepción

6. Documento evaluado

Título de pregrado “Sistema informático basado en Analytics para la prevención y monitoreo de casos de violencia contra la mujer registrados en los Centros de Emergencia Mujer, basado en indicadores de la norma ISO 37120”

7. Autores del documento

Bach. Joselyn Mayté Ccanto Bravo - Bach. José Antonio Aira Retuerto

8. Fecha de recepción del documento 02/12/2022

9. Fecha de aplicación del programa informático de similitudes 06/12/2022

10. Software utilizado

- Turnitin

11. Configuración del programa detector de similitudes

- Excluye textos entrecerrillados
- Excluye bibliografía
- Excluye cadenas menores a 40 palabras

12. Porcentaje de similitudes según programa detector de similitudes **08 (ocho)%**

13. Fuentes originales de las similitudes encontradas

Se adjunta en el anexo 1

-
14. Observaciones

-
15. Calificación de originalidad

- Documento cumple criterios de originalidad, sin observaciones
- Documento cumple criterios de originalidad, con observaciones
- Documento no cumple criterios de originalidad

16. Fecha de informe

09/12/2022



Firmado digitalmente por PRO
CONCEPCION Luzmila Elisa FAU
20148092282 soft
Motivo: Soy el autor del documento
Fecha: 14.12.2022 15:32:53 -05:00

Firma de evaluador

Dra. Luzmila E. Pró Concepción
Directora (e) de la EPIS

“Tesis presentada a la Universidad Nacional
Mayor de San Marcos, Lima, Perú, para obtener
el Título de Ingeniero de Sistemas”

© José Aira, Joselyn Ccanto, 2022

Todos los derechos reservados.

DEDICATORIA

A Dios por darme la sabiduría y fortaleza para concluir esta tesis, a mi papá y mamá por su aliento y apoyo constante y en memoria a mis abuelos Antonio Aira y Jorge Retuerto.

José Aira

DEDICATORIA

A Dios por darme fortaleza y guiar mi camino para poder concluir esta tesis, a mi mamá y mi tía por su apoyo constante, a mi abuela, mis tíos y en memoria de mi abuelo.

Joselyn Ccanto

AGRADECIMIENTOS

A Dios por ser el guía en cada etapa de la presente tesis, desde la elección del tema hasta el desarrollo de este.

A nuestra familia por su constante apoyo en todo el proceso de la tesis, aportando sus conocimientos y dándonos aliento.

A nuestros profesores de la universidad San Marcos, por todos los conocimientos brindados que nos sirvieron de base para el desarrollo de esta tesis. En especial a nuestro asesor de tesis, el profesor Jorge Guerra, por siempre estar dispuesto a apoyarnos, corregirnos y darnos aliento en todo el proceso.

RESUMEN

En los últimos años, el Perú presenta un gran porcentaje de mujeres víctimas de violencia física, psicológica, sexual y/o económica, que ha ocasionado que el personal disponible para el monitoreo continuo y/o prevención de los casos de violencia contra la mujer no sea suficiente en los Centros de Emergencia Mujer. Ante esta situación, la presente tesis desarrolla un sistema informático basado en las categorías predictiva y descriptiva de Analytics.

Para la parte predictiva, se ha desarrollado un modelo Random Forest, con una precisión del 97%, que predice si un caso de violencia contra la mujer reportado es tentativa de feminicidio. Finalmente, para la parte descriptiva, se implementó una plataforma de visualización alineada a los indicadores de los grupos temáticos de la norma ISO 37120:2018, enfocados a la igualdad de género y lucha contra la violencia, la cual permite realizar un seguimiento continuo y prevención de este tipo de casos.

Palabras Claves: Analytics, violencia contra la mujer, Norma ISO 37120, modelos predictivos, ETL.

ABSTRACT

In recent years, Peru has a large percentage of women victims of physical, psychological, sexual and/or economic violence, which has caused that the personnel available for continuous monitoring and/or prevention of cases of violence against women are not enough in the Women's Emergency Centers. Faced with this situation, this thesis develops a computer system based on the predictive and descriptive categories of Analytics.

For the predictive part, a Random Forest model has been developed, with an accuracy of 97%, which predicts whether a reported case of violence against women is attempted femicide. Finally, for the descriptive part, a visualization platform aligned with the indicators of the thematic groups of the ISO 37120: 2018 standard, focused on gender equality and the fight against violence, was implemented, which allows continuous monitoring and prevention of these types of cases.

Keywords: Analytics, violence against women, ISO 37120 standard, predictive models, ETL.

Índice

Introducción	13
Capítulo I: Planteamiento Metodológico	15
1.1 Antecedentes	15
1.2 Justificación	20
1.3 Definición del problema	21
1.4 Objetivos	21
1.4.1 Objetivo principal	21
1.4.2 Objetivos específicos	22
1.5 Hipótesis	22
Capítulo II: Marco Teórico	23
2.1 Centro de Emergencia Mujer – CEM	23
2.1.1 Violencia contra la mujer	23
2.1.2 Protocolo de actuación conjunta CEM – Comisarías	24
2.1.3 Instituciones relacionadas al CEM	25
2.2. Norma ISO 37120 – Desarrollo sostenible de ciudades inteligentes	30
2.3. Analytics	32
2.3.1 Analítica descriptiva	33
2.3.2 Analítica de diagnóstico	33
2.3.3 Analítica predictiva	34
2.3.4 Analítica prescriptiva	34
2.3.5 Analítica cognitiva	34
2.4. Algoritmos predictivos supervisados	35
2.4.1 Bosques Aleatorios (Random Forest)	35
2.4.2 Regresión Logística	38
2.4.3 Árbol de decisión	39
2.4.4 Máquinas de Soporte Vectorial (SVM)	41
2.4.5 KNN	42
Capítulo III: Estado del Arte	44
3.1 Trabajos previos	45
3.1.1 Internet de las Cosas - Monitoreo y Predicción	46
3.1.2 Analítica aplicada a monitoreo y predicción	51
3.1.3 Redes Neuronales - Seguimiento y Predicción	60
3.2. Determinación de la solución	68
Capítulo IV: Aporte Teórico	71

4.1. Esquema general de la solución	71
4.1.1 Fase de extracción y exploración de datos	72
4.1.2 Fase de preprocesamiento	73
4.1.3 Fase de interpretación y selección de features	73
4.1.4 Fases de analítica predictiva	74
4.1.5 Fase de analítica descriptiva	76
Capítulo V: Aporte Práctico	77
5.1. Selección de tecnologías basadas en hardware y software para la solución.	77
5.2. Desarrollo e implementación de la metodología propuesta	79
5.3. Validación de modelo predictivo	121
5.4. Discusión de resultados	126
5.5. Resumen de los resultados obtenidos	129
Capítulo VI: Conclusiones y recomendaciones	130
6.1. Conclusiones	130
6.2. Recomendaciones	131
ANEXO 1: Formulario de nuevos casos de tentativa	132
REFERENCIAS BIBLIOGRÁFICAS	150

Lista de figuras

Figura 1.1 Evolución del número de casos de violencia contra la mujer atendidos en los CEM en el periodo 2017 – 2021	16
Figura 1.2. Evolución del número de feminicidios registrados en los CEM en el periodo 2017 – 2021.	16
Figura 1.3. Evolución del número de tentativa de feminicidios registrados en los CEM en el periodo 2017 – 2021.	17
Figura 2.1. Ruta de atención del protocolo de atención conjunta CEM y PNP.	25
Figura 2.2. MIMP-CEM e Instituciones que actúan de manera conjunta en el ámbito integral y protección frente a la violencia contra las mujeres y los integrantes del grupo familiar.	29
Figura 2.3. Grupos temáticos Norma ISO 37120:2018.	31
Figura 2.4 Tipos de analítica.	35
Figura 2.5 Random Forest.	36
Figura 2.6 Esquema del proceso de funcionamiento del Random Forest.	37
Figura 2.7. Parámetros de sklearn.ensemble.RandomForestClassifier.	37
Figura 2.8 Regresión Logística para 2 variables.	38
Figura 2.9. Parámetros de sklearn.linear_model.LogisticRegression.	39
Figura 2.10 Árbol de decisión.	40
Figura 2.11 Parámetros de sklearn.tree.DecisionTreeClassifier.	41
Figura 2.12 Máquina de Soporte Vectorial.	41
Figura 2.13 Parámetros de sklearn.svm.SVC.	42
Figura 2.14 Representación del algoritmo KNN.	43
Figura 2.15 Parámetros de sklearn.neighbors.KNeighborsClassifier.	43
Figura 3.1. Arquitectura del sistema propuesto.	48
Figura 3.2. Modelo de monitoreo y predicción del crimen propuesto.	50
Figura 3.3. Pantallas de la implementación del modelo.	51
Figura 3.4. Flujo para predecir el índice de violencia de pareja.	53
Figura 3.5. Proceso metodológico del modelo predictivo.	58
Figura 3.6. Gráficos de barras con el número de delitos por estado y mapa de calor para representar el mayor número de asaltos en un estado.	60
Figura 3.7. Matriz de confusión de cada conjunto de datos.	62
Figura 3.8. Análisis comparativo del método propuesto con métodos avanzados.	64
Figura 3.9. Descripción general de la red bicanal propuesto.	65
Figura 3.10. Validación cruzada del modelo predictivo propuesto.	66
Figura 3.11. Tablas de comparación del método propuesto vs métodos avanzados para los dos conjuntos de datos.	67
Figura 4.1. Esquema propuesto para el desarrollo de la solución.	71
Figura 4.2. Esquema de extracción y exploración.	72
Figura 4.3. Esquema de preprocessamiento.	73
Figura 4.4. Esquema de interpretación y selección de features.	74
Figura 4.5. Esquema de balance de datos y selección de modelos predictivos.	74
Figura 4.6. Esquema de balance de datos y selección de modelos predictivos.	76
Figura 5.1. Base de datos del registro de casos del Centro Emergencia Mujer.	78
Figura 5.2. Código Python de extracción de archivos fuente.	80
Figura 5.3. Identificación de valores perdidos.	81
Figura 5.4. Identificación de outliers.	82
Figura 5.5. Distribución de los datos de algunos features.	82

Figura 5.6. Comportamiento del valor del parámetro n_neighbors de acuerdo con el accuracy obtenido.	93
Figura 5.7. Validación cruzada y desviación estándar del modelo predictivo con KNN.	94
Figura 5.8. Validación de modelo elegido SVM frente a casos nuevo.	98
Figura 5.9. Comportamiento del valor del parámetro Max_depth y Max_Samples de acuerdo con el accuracy obtenido.	101
Figura 5.10. Validación cruzada y desviación estándar del modelo predictivo con Random Forest.	103
Figura 5.11. Comportamiento del valor del parámetro criterion y max_depth de acuerdo con el accuracy obtenido.	106
Figura 5.12. Validación cruzada y desviación estándar del modelo predictivo con Decision Tree.	108
Figura 5.13. Modelo de datos copo de nieve.	114
Figura 5.14. Indicadores propuestos basados en la ISO 37120:2018 plasmados en gráficos de la herramienta Power BI.	116
Figura 5.15. Panel de seguimiento de los casos de violencia contra la mujer reportados en Lima Metropolitana.	119
Figura 5.16. Formulario de apoyo para la predicción de nuevos casos de tentativa de feminicidio.	120
Figura 5.17. Porcentaje de precisión del modelo predictivo (Random Forest) de los nuevos casos de violencia.	125
Figura 5.18. Matriz de confusión frente a nuevos casos de violencia contra la mujer.	128

Lista de tablas

Tabla 3.1. Resultados de la aplicación de los métodos predictivos.	54
Tabla 3.2. Resumen de resultados de los modelos.	58
Tabla 3.3 Comparación detallada de los resultados obtenidos en el conjunto de datos Violent Flow, Hockey Fight y Surveillance Flight.	63
Tabla 3.4 <i>Comparación del rendimiento entre los diferentes conjuntos de datos.</i>	66
Tabla 3.5 Resultados de la evaluación.	70
Tabla 4.1 Técnica y algoritmos de balanceo de datos.	75
Tabla 5.1. Número de columnas de los archivos por cada año.	81
Tabla 5.2. Preprocesamiento aplicado a los features del registro de casos del Centro Emergencia Mujer 2017-2021.	83
Tabla 5.3. Concepto de variables dummies aplicado a los features faltantes del registro de casos del Centro Emergencia Mujer 2017-2021.	87
Tabla 5.4. Concepto de variables dummies aplicado a los features faltantes del registro de casos del Centro Emergencia Mujer 2017-2021.	88
Tabla 5.5. Resultado de la evaluación de los algoritmos de predicción con las técnicas de balanceo (accuracy)002E	90
Tabla 5.6. Resultado de la evaluación del parámetro n_neighbors en KNeighborsClassifier.	92
Tabla 5.7. Resultado de la evaluación de los parámetros max_iter y class_weight en LogisticRegression.	95
Tabla 5.8. Resultado de la evaluación de los parámetros gamma y C en SVM.	97
Tabla 5.9. Resultado de la evaluación de los parámetros n_estimators (A), max_depth (B), max_samples (C) y random_state en RandomForestClassifier.	99
Tabla 5.10. Resultado de los mejores valores de los parámetros n_estimators (A), max_depth (B), max_samples (C) y random_state en relación con el accuracy.	102
Tabla 5.11. Resultado de la evaluación de los parámetros criterion (A), max_depth (B), max_samples (C) y random_state en DecisionTreeClassifier.	104
Tabla 5.12. Resultado de los mejores valores de los parámetros criterion (A), max_depth (B), max_samples (C) y random_state) en relación con el accuracy.	107
Tabla 5.13. Resumen de resultados de los algoritmos de predicción de acuerdo con su respectivo accuracy.	109
Tabla 5.14. Indicadores propuestos en base a la Norma ISO 37120-2018.	111
Tabla 5.15. Tabla de validación de los resultados del valor predicho por modelo predictivo vs el valor real de los nuevos casos de violencia de tentativa y no tentativa de feminicidio.	122
Tabla 5.16. Tabla de validación de los resultados de algoritmos predictivos.	126
Tabla 5.17. Tabla con la interpretación de las métricas que permiten medir el rendimiento del modelo frente a casos de tentativa de feminicidio	127

Introducción

La violencia contra la mujer es uno de los problemas sociales de gran preocupación para la mayoría de las autoridades de países en el mundo, violencia que se manifiesta muchas veces dentro del hogar de forma física, sexual, psicológica y/o económica, sin distinción de raza, credo y condición social. Estos casos de violencia en su mayoría son realizados por hombres (esposos, convivientes, padres), lamentablemente algunos terminan en tentativa de feminicidio o en feminicidio.

Muchas organizaciones a nivel mundial tratan de frenar el avance de violencia hacia las mujeres mediante la creación de políticas, leyes, medidas y entidades especializadas para ello, por ejemplo, UN (2015) incluye dentro de los Objetivos de Desarrollo Sostenible, la Igualdad de Género (Objetivo 5), enfocado en la eliminación de toda forma de violencia contra las mujeres y niñas; otro ejemplo son los ámbitos de la norma ISO 37120:2018, los cuales hacen énfasis en la inclusión de las mujeres y el papel fundamental que cumple en la sociedad (ISO, 2018).

Dentro de los esfuerzos realizados en el Perú en torno a esta problemática está la creación de los Centro Emergencia Mujer (CEM) y del programa Aurora para frenar y controlar esta violencia; sin embargo, a pesar del esfuerzo realizado presenta un deficiente sistema de atención a las víctimas, según (DP, 2018).

Ante la problemática presentada en el Perú, la presente tesis propone el desarrollo de un sistema informático basado en Analytics para un adecuado monitoreo y prevención de los casos de violencia contra la mujer registrados en los CEM de Lima Metropolitana, para ello se utiliza las técnicas y/o herramientas necesarias para: la extracción de los datos de casos de violencia registrados en los CEM, el desarrollo de un proceso ETL para la limpieza de los datos, la implementación de un modelo

predictivo para casos de tentativa de feminicidio y la presentación de los resultados alineados a la norma ISO 37120:2018 en una plataforma de visualización.

La presente tesis está organizada en 6 capítulos descritos brevemente a continuación: El capítulo I presenta el Planteamiento Metodológico donde se detalla los antecedentes, definición del problema, justificación, los objetivos principales y específicos de esta tesis. El capítulo II se enfoca en el Marco Teórico donde se definen los términos necesarios para el entendimiento de los capítulos siguientes de esta investigación. El capítulo III aborda el Estado de Arte, donde se revisa artículos y tesis que plantean una solución con técnicas, herramientas y/o métodos para la problemática de violencia, de los cuales se seleccionan algunos, de tal manera que sirvan de apoyo al planteamiento de la solución de la presente tesis. El capítulo IV, Aporte Teórico, en este capítulo se describe la metodología propuesta con sus fases y técnicas para el desarrollo de la solución, La aplicación de dicha metodología y sus fases se detallan en el Capítulo V, donde se describe minuciosamente las técnicas, algoritmos y herramientas necesarias para la implementación de la solución propuesta, así como los resultados obtenidos de la plataforma de visualización y la validación del modelo predictivo. Finalmente, en el Capítulo VI se plasma las conclusiones de esta investigación y las recomendaciones para futuros trabajos de investigación.

Capítulo I: Planteamiento Metodológico

1.1 Antecedentes

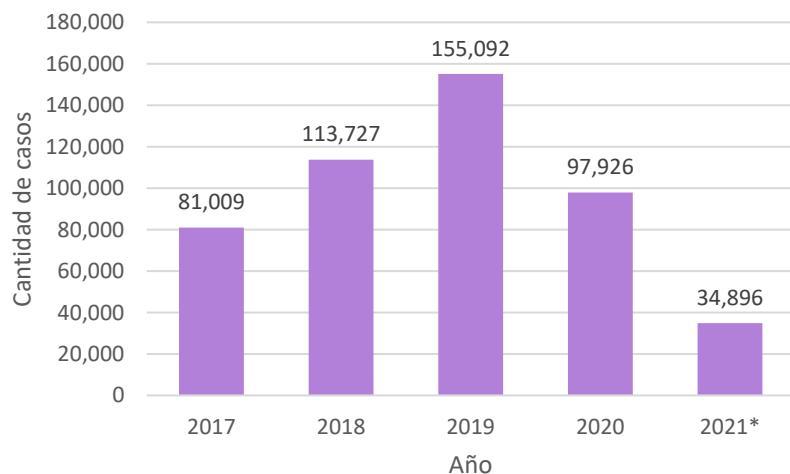
Dentro de los 17 Objetivos de Desarrollo Sostenible (ODS), planteados por las Naciones Unidas, el Objetivo 5 - Igualdad de Género menciona como meta 5.2: “Eliminar todas las formas de violencia contra todas las mujeres y las niñas en los ámbitos público y privado, incluidas la trata y la explotación sexual y otros tipos de explotación” (UN, 2015).

A nivel mundial según el Programa de las Naciones Unidas para el Desarrollo, PNUD (2021), la violencia física y/o sexual se ha presentado en 1 de cada 3 mujeres; y en el Perú, según el Instituto Nacional de Estadística e Informática, INEI (2019), 6 de cada 10 mujeres entre 15 a 49 años han sido víctimas de violencia psicológica y/o verbal, física o sexual.

En las cartillas estadísticas publicado por el Portal Estadístico Programa Nacional AURORA, PEPNA (2021a), se puede concluir de la figura 1.1 (casos de violencia contra la mujer), figura 1.2 (feminicidios) y figura 1.3 (tentativas de feminicidio), reportados al Centros de Emergencia Mujer (CEM) en el 2020 (periodo de confinamiento a causa del COVID), evidencian una baja a comparación del 2018 y 2019, sin embargo, a comparación de los casos reportados en el 2017 presentan un alza.

Figura 1.1

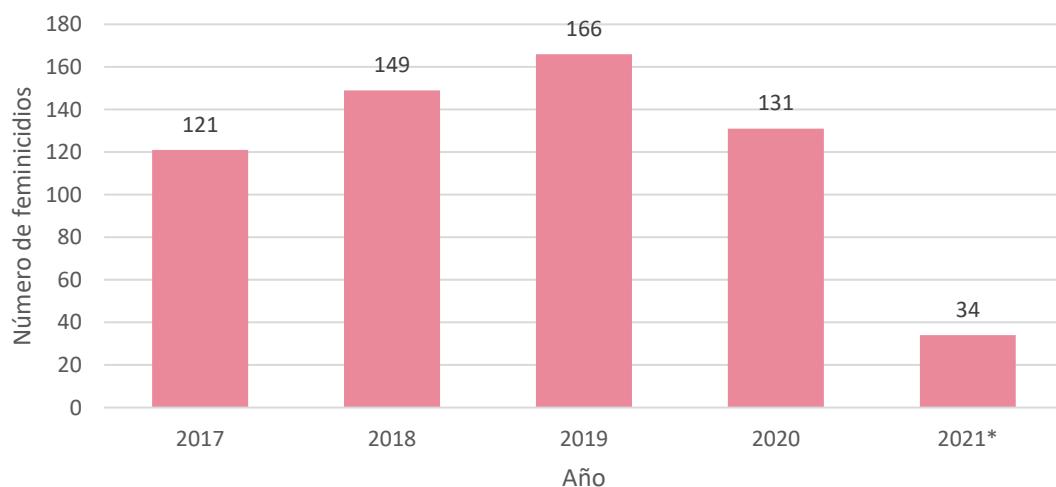
Evolución del número de casos de violencia contra la mujer atendidos en los CEM en el periodo 2017 – 2021



Nota: (*) El periodo comprende el primer trimestre del 2021. Fuente: Elaboración propia

Figura 1.2.

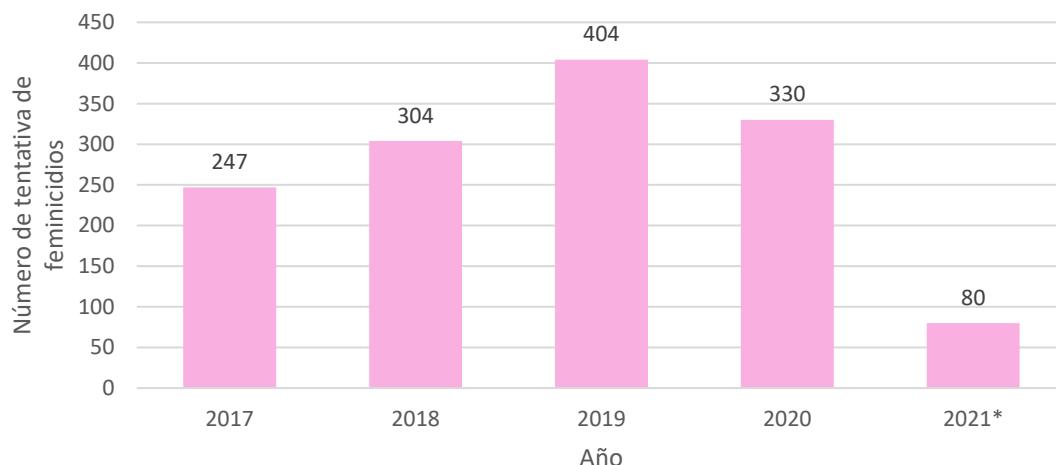
Evolución del número de feminicidios registrados en los CEM en el periodo 2017 – 2021.



Nota: (*) El periodo comprende el primer trimestre del 2021. Fuente: Elaboración propia

Figura 1.3.

Evolución del número de tentativa de feminicidios registrados en los CEM en el periodo 2017 – 2021.



Nota: (*) El periodo comprende el primer trimestre del 2021. Fuente: Elaboración propia

Para contrarrestar dicha violencia en el Perú, según el Ministerio de la Mujer y Poblaciones Vulnerables, MIMP (2019); las instituciones que se hacen cargo son: Ministerio de la Mujer y Poblaciones Vulnerables a través del CEM, Ministerio del Interior a través de la Policía Nacional del Perú (PNP), Ministerio de Justicia, Poder Judicial (PJ), Ministerio Público (MP), Ministerio de Salud, Ministerio de Educación entre otros; con funciones de prevención, monitoreo, asistencia e investigación de los casos, las cuales se realizan de manera articulada a través de un Protocolo Interinstitucional.

Según el Observatorio Nacional de la Violencia Contra las Mujeres y los Integrantes del Grupo Familiar, ONVCMIGF (2017), una de las deficiencias de las instituciones que combaten la lucha contra la violencia a la mujer es la desproporcionalidad existente entre la atención de los casos presentados con el número

de policías para tratarlos, donde se señala que un total de 3 781 de los casos no se les otorgaba medidas de protección aun cuando el Poder Judicial había dictaminado ello.

Otras de las deficiencias encontradas, según la Defensoría del Pueblo, DP (2018), menciona que la supervisión de los casos de violencia contra la mujer admitidos por los CEM tiene acciones iniciales poco después que el caso ingrese; sin embargo, este seguimiento va disminuyendo conforme se va complicando ya sea por tiempo o complejidad. Además, según PEPNA (2020), en el periodo de enero a septiembre del 2020, el 15% de las mujeres víctimas de feminicidio denunció acciones violentas por parte de su pareja.

La presente tesis se enfoca en el CEM, ya que es una de las primeras instancias a donde recurren las víctimas, MIMP (2018), y donde se presentan deficiencias para el monitoreo y acompañamiento de las mujeres. Según INEI (2019), el 77,6 % de las mujeres violentadas recurren por ayuda a la PNP, y esta articula con el CEM para la atención y seguimiento oportuno (MIMP, 2018).

Cabe mencionar también la importancia del servicio del MIMP, línea 100, para reportar casos de violencia contra la mujer e integrantes del grupo familiar, el cual durante el confinamiento obligatorio a causa de la pandemia del COVID 19, según DP (2020), del número de llamadas recibidas, solo priorizaron los que reportaban casos cuyo riesgo es severo, los que ocurrieron en flagrancia, las violaciones sexuales y los feminicidios; los de riego moderado o leve su atención no se trató de forma inmediata. Además, según PEPNA (2021a), entre enero a diciembre del 2020 se reportaron 235 791 llamadas a la Línea 100, de los cuales 79% informaron que la víctima era mujer.

Alineado a las actividades realizadas por las instituciones para mejorar la atención y seguimiento ante algún caso de violencia contra la mujer y los integrantes del grupo familiar, Viviano (2019) implementa el registro piloto integrado de

feminicidio obteniendo información básica representada por 6 variables comunes (Ubigeo, medio o modalidad utilizado, lugar de ocurrencia, edad, nacionalidad y vínculo relacional), extraídas a partir de un proceso de integración y limpieza de datos de los repositorios de la PNP, MIMP y MP. Dicho artículo concluye que se debe definir una metodología de integración fiable, y obtener un mayor número de variables para enriquecer la información. Adicionalmente, Viviano (2018) indica que la limitación para la presentación de datos radica en la heterogeneidad de estos.

Por otro lado, en relación con las medidas de prevención a través de las predicciones, Sivanagaleela y Rajesh, (2019), desarrollan una investigación usando el algoritmo difuso de C-Mean para agrupar los datos delictivos y de esa forma realizar una predicción identificando las áreas donde se cometieron frecuentemente dichos delitos y poder reducirlos. De igual forma Nuraeni et al., (2019), utiliza el modelo CRISP-DM para clasificar y determinar patrones de violencia hacia las mujeres y niños, a través del algoritmo K-means, con el objetivo de dar seguimiento a cada uno de los casos de violencia. Además, Kmuar et al., (2017) indica que a través del proceso de análisis del crimen y predicción del delito se puede planificar estrategias de prevención y reducción de la tasa de ocurrencia de delitos contra la mujer, esta investigación propone predecir el número de delitos que pueden ocurrir contra las mujeres en los próximos años, a través de 2 técnicas de predicción: Clasificación Ingenua de Bayes y el algoritmo de series de tiempo.

Finalmente, el proyecto piloto descrito por Viviano en el 2019, es un paso a la visión de Smart city, la cual está alineado a Amagasa et al., (2017), quienes proponen una caja de herramientas analíticas para una ciudad inteligente, la cual se divide en 3 fases importantes: Preparación, Análisis y Visualización, que transforman los datos obtenidos en información útil para la ciudad de Fujisawa. Dando así una base para la

implementación de una ciudad inteligente a partir del análisis y procesamiento de datos. En la misma línea, Moschen et al., (2019) describe la importancia de la norma ISO 37120 alineado a los objetivos de desarrollo sostenible de la ONU resaltando la importancia de las medidas, indicadores y herramientas para monitorear el desarrollo urbano e indicando que la dimensión de seguridad de la norma ISO 37120 se alinea al objetivo 5 de los objetivos de desarrollo sostenible de la ONU (ODS). Además, Pagliaro et al., (2016) proponen una metodología de 6 fases enfatizando de manera transversal la importancia de los datos en: Planificación preliminar, identificación de los campos de acción, adquisición de datos, análisis de datos, categorización de problemas, definición de las estrategias.

En resumen, por todas las deficiencias encontradas en las instituciones, se desea plantear, en base a los artículos de investigación, un sistema de Analytics usando indicadores de la norma ISO 37120:2018 alineado a la ODS, para un monitoreo y prevención efectiva de los casos violencia contra la mujer en Lima.

1.2 Justificación

Dentro de los desafíos del Perú para la adhesión a la OCDE (Organización para la cooperación y el Desarrollo económico), se encuentra el desarrollo de políticas de gobierno en torno a las perspectivas propuestas por dicha institución, entre ellas el capital humano, el cual está alineado al objetivo 5: Igualdad de género del ODS.

De acuerdo con los datos consignados en la sección de antecedentes, la violencia contra la mujer es un problema social que involucra a todo el país, independientemente de raza, credo y condición social, por lo tanto, es una de las principales trabas para construir una sociedad más justa e inclusiva.

Parte de la justificación también se encuentra en el desarrollo de una plataforma analítica, utilizada para las tareas de prevención y monitoreo que abarca la propuesta, a través de los datos obtenidos de los diferentes repositorios de los CEM, donde se consignan casos de violencia contra la mujer.

1.3 Definición del problema

La gran cantidad de casos de violencia contra la mujer, reportados en los últimos años, acompañado además de un deficiente sistema de atención a las víctimas, ha ocasionado un manejo poco adecuado de este tipo de delito, lo cual ha quedado más evidente durante el periodo de la pandemia del COVID19, conllevando a una agudización de este problema social y a una ineficiencia de las políticas existentes relacionado a la violencia contra la mujer. Por lo tanto, ¿Es posible que las TIC's puedan brindar seguimiento de los casos actuales de violencia contra la mujer y apoyar en la prevención de futuros casos de feminicidio?

1.4 Objetivos

1.4.1 Objetivo principal

Desarrollar un sistema informático basado en analytics usando los indicadores de la norma ISO 37120:2018 que contribuirá en la prevención y monitoreo de actividades relacionadas a los casos de violencia contra la mujer.

1.4.2 Objetivos específicos

- Desarrollar un proceso ETL para obtener data integra y limpia utilizando los registros de repositorios relacionados a las instituciones encargadas de la lucha contra la violencia a la mujer.
- Realizar un análisis eficiente y alineado a los indicadores de la norma ISO 37120:2018 de los grupos temáticos enfocados a la igualdad de género y lucha contra la violencia.
- Implementar un sistema predictivo de los casos de violencia contra la mujer, que apoye a alertar a las instituciones de forma estratégica.
- Implementar en una plataforma de visualización los resultados de los análisis realizados.

1.5 Hipótesis

El sistema informático basado en analytics usando los indicadores de la norma ISO 37120:2018, permitirá realizar tareas efectivas en la prevención y monitoreo de actividades relacionadas a los casos de violencia contra la mujer.

Capítulo II: Marco Teórico

2.1 Centro de Emergencia Mujer – CEM

Una de las instituciones de lucha contra la violencia hacia la mujer son los CEM, según Plataforma digital única del Estado Peruano [PDUEP] (2021) los define como:

Servicios públicos especializados y gratuitos, de atención integral y multidisciplinaria, para víctimas de violencia contra la mujer e integrantes del grupo familiar. Para procurar la recuperación del daño sufrido, los CEM brindan orientación legal, defensa judicial, consejería psicológica y asistencia social.

El programa está dirigido a todas las personas afectadas por violencia familiar y sexual de su ámbito de responsabilidad, sin importar condición social, edad o sexo.

La cantidad total de CEM en el Perú es 416, de los cuales 245 son CEM regulares, 170 CEM-Comisarías y 1 CEM-Centro de salud. Lima metropolitana cuenta con 39 de ellos (PEPNA, 2021a).

2.1.1 Violencia contra la mujer

Según la Organización Mundial de la Salud [OMS], la violencia contra la mujer se define como: “Todo acto de violencia de género que resulte, o pueda tener como resultado un daño físico, sexual o psicológico para la mujer” (OMS, 2021).

2.1.1.1. Violencia contra la mujer en el marco legal nacional e internacional

Debido a que la violencia contra la mujer es uno de los problemas que enfrenta el mundo, se han creado leyes, convenciones, declaraciones con el objetivo de defender los derechos de la mujer, y según Ministerio de Justicia y Derechos Humanos [MINJUS] (s.f.) el Perú se ha adherido a las siguientes:

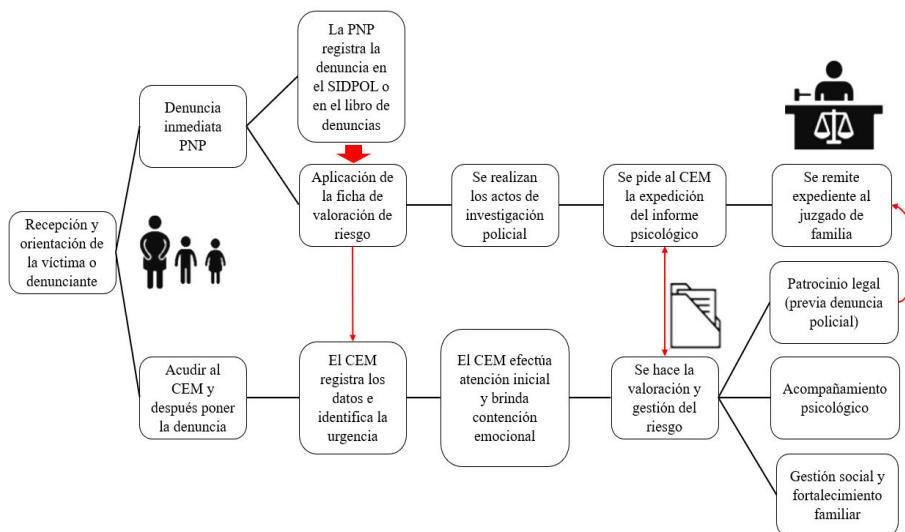
- Declaración Universal de los Derechos Humanos (1948)
- Convención sobre eliminación de todas las formas de discriminación contra la mujer (CEDAW)

2.1.2 Protocolo de actuación conjunta CEM – Comisarías

Las actividades que realizan el Centro de Emergencia Mujer (CEM) y la Policía Nacional del Perú (PNP), está basado en el protocolo de atención conjunta creado por ambas instituciones debido a que son las primeras instancias de atención cuando ocurre algún caso de violencia contra la mujer, MIMP (2018). Dicho proceso se da mediante formularios o puntuaciones de riesgos, según la casuística de los hechos de violencia como se muestra en la figura 2.1.

Figura 2.1.

Ruta de atención del protocolo de atención conjunta CEM y PNP.



Fuente: MIMP (2018)

2.1.3 Instituciones relacionadas al CEM

En el Perú, en base a la (Ley N° 30364, 2015, art. 45), norma publicada con el objetivo de erradicar la violencia contra la mujer, se desarrolló el Protocolo de base de actuación conjunta en el ámbito integral y protección frente a la violencia contra las mujeres y los integrantes del grupo familiar, MIMP (2019), el cual establece 4 servicios:

1. Asistencia jurídica y defensoría pública

- Servicio brindado por el MIMP a través de CEM y complementado de ser necesario por el MINJUSDH a través de Defensa Pública
- Al recibir una denuncia la PNP, Ministerio Público, Poder Judicial derivan al CEM para que sea atendido de forma integral, de no poder brindar el servicio el CEM deriva al MINJUSDH para garantizar que la víctima pueda recibir asistencia legal y atención integral
- Medidas de protección y/o cautelares

- Patrocina los casos de violencia en cualquier etapa del proceso
- Gestiones para albergar a la víctima de riesgo moderado o severo en un hogar de refugio temporal
- La PNP, Ministerio Público, Poder Judicial aplica la ficha de valoración de riesgo

2. Atención y protección especializada

- La PNP tienen el deber de recibir las denuncias de violencia contra la mujer e integrantes del grupo familiar y aplicar la ficha de valoración de riesgo, a la vez es el encargado de coordinar con el CEM para su respectiva atención, y de ser el caso el ingreso de la víctima a un hogar de refugio temporal. En el caso de flagrancia la PNP detiene al agresor y realiza un relevamiento de información y elabora un acta que servirá para que el Ministerio Público siga con las investigaciones pertinentes
- El juez de familia emite las medidas de protección y el Ministerio Público lo registra en el Registro único de Víctimas y agresores (RUVA)
- En el caso de que las víctimas sean niños, niñas y/o adolescentes que se encuentren en situación de riesgo el juzgado de familia comunica a la DEMUNA o a Unidad de Protección Especial para que puedan salvaguardar a los niños, niñas y/o adolescentes

3. Promoción, prevención, atención y recuperación de víctimas de violencia

- Servicio brindado a través de la comisaría, Ministerio Público, Poder Judicial, CEM, establecimientos de salud, Unidad de Protección Especial- UPE, Instituto de Medicina Legal o Ciencias Forenses, la comunidad y campañas de difusión,
- La derivación e ingreso de las víctimas de violencia se puede realizar a través del CEM según los criterios de derivación establecidos por el MIMP involucrando a los sistemas de justicia y hogares temporales o UPE.
- De acuerdo con el tipo de violencia y situación de riesgo de la víctima se elabora un plan de seguridad para proteger a la víctima
- Brindan orientación y soporte socioemocional.

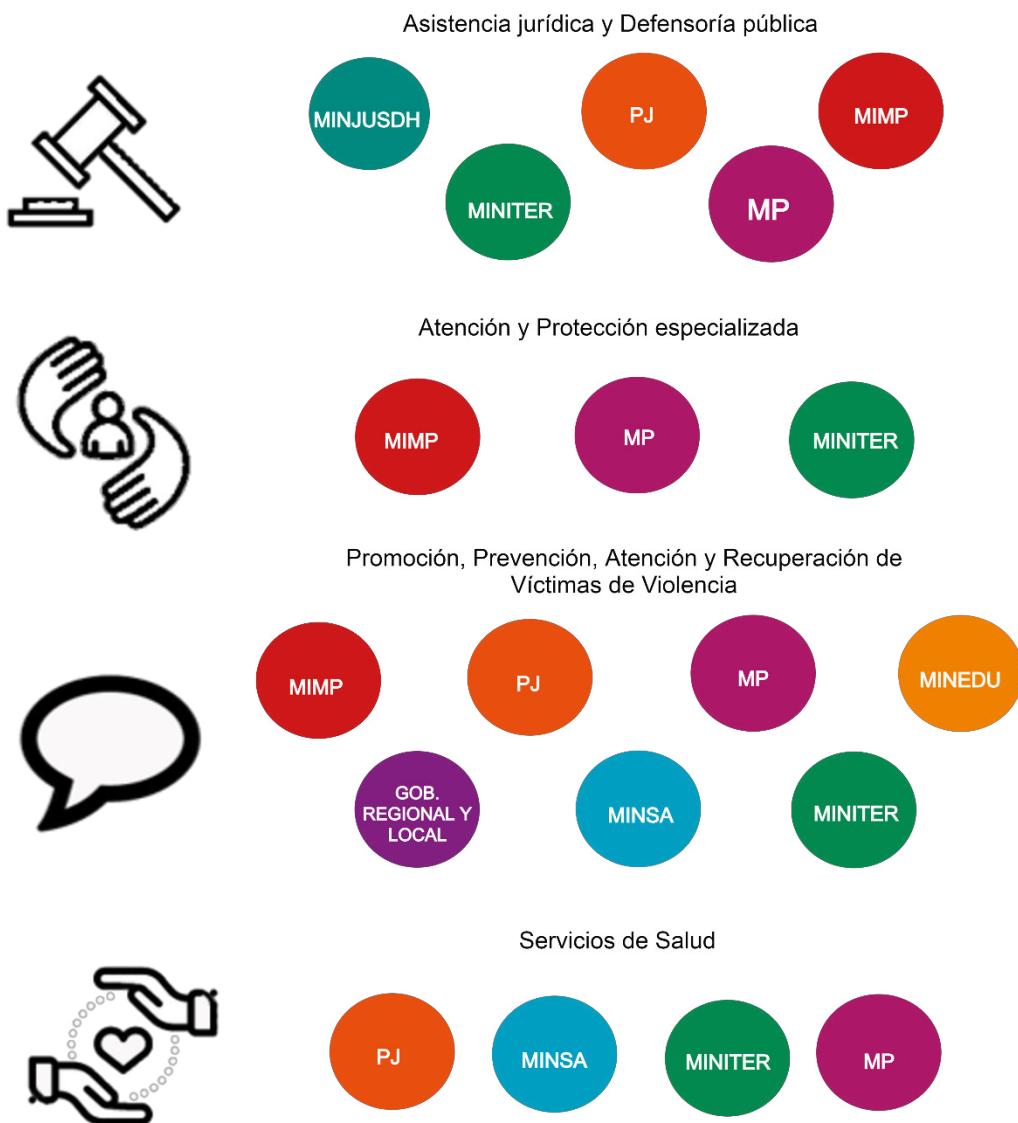
4. Servicios de salud

- La PNP, Ministerio Público, Poder Judicial u otros pueden derivar los casos de violencia contra la mujer para su atención en el servicio de salud.
- Los centros de salud pueden atender u orientar casos de violencia a través del personal de psicología o similares, de igual forma al tomar conocimiento de un caso de violencia contra la mujer debe de presentar una denuncia a las instituciones correspondientes.
- En caso de detectar que la víctima está en estado de desprotección debe de comunicar de forma inmediata a las instituciones correspondientes.
- El SAMU es el encargado de atender derivaciones de afectación física o mental y coordinar con el establecimiento más cercano para la atención oportuna.

Dichos servicios son cumplidos de manera coordinada por las siguientes instituciones, el Ministerio de la Mujer y Poblaciones Vulnerables (MIMP) a través del Centro de Emergencia Mujer (CEM), el Ministerio de Salud (MINSA), el Ministerio del Interior a través de la Policía Nacional del Perú (PNP), Ministerio de Justicia y Derechos Humanos (MINJUSDH), Ministerio de Educación (MINEDU), Poder Judicial, Ministerio Público, y los gobiernos locales y regionales, mostradas en la figura 2.2. Donde se puede resaltar que el MIMP a través del CEM participa de manera activa en 3 de ellas cumpliendo funciones de prevención, monitoreo, asistencia, investigación de los casos.

Figura 2.2.

MIMP-CEM e Instituciones que actúan de manera conjunta en el ámbito integral y protección frente a la violencia contra las mujeres y los integrantes del grupo familiar.



Fuente: Elaboración Propia

2.2. Norma ISO 37120 – Desarrollo sostenible de ciudades inteligentes

La Organización Internacional de Normalización (ISO) es la institución encargada de crear estándares que sirvan como guía y orientación a las empresas u organizaciones a implementar buenas prácticas en sus procesos para brindar productos y servicios con parámetros de calidad definida. Entre las normas ISO publicadas, según Lockett et al., (2020) existe una relacionada a las Ciudades y comunidades sostenibles- Indicadores para los servicios urbanos y la calidad de vida, llamada norma ISO 37120.

La cual cuenta con 2 ediciones, siendo la primera publicada en el año 2014 denominada Norma ISO 37120:2014, que contempla 17 grupos temáticos (Economía, Educación, Energía, Media Ambiente, Finanzas, Respuesta ante Incendios y Emergencias, Gobierno, Salud, Esparcimiento, Seguridad, Residuos Sólidos, Telecomunicaciones e Innovación, Transporte, Planificación Urbana, Agua y Saneamiento, Aguas Residuales, Vivienda) y un total de 100 indicadores distribuidos en cada uno de ellos. Y la segunda edición publicada en el año 2018 denominada Norma ISO 37120:2018, que contempla 19 grupos temáticos, mostrados en la figura 2.3, incorporando los siguientes grupos temáticos Población y Condiciones Sociales, Cultura y Deporte, Agricultura Urbana-Local y Seguridad Alimentaria, a los ya 17 existentes, teniendo en cuenta que el grupo temático de “Respuesta ante Incendios y Emergencias” fue incluido en el grupo de Seguridad.

Figura 2.3.

Grupos temáticos Norma ISO 37120:2018.



Fuente: Elaboración Propia

Cabe resaltar que las normas ISO son estándares de buenas prácticas que deben ser adaptados a la realidad de cada empresa u organización, puesto que cada uno tiene diferentes características y niveles de madurez.

Según la Organización Internacional de Estandarización ISO (2018) cada grupo temático cuenta con indicadores clasificados en:

- **Indicador central (C):** Indicadores que se requieren para demostrar el desempeño en la prestación de servicios de la ciudad y la calidad de vida.
- **Indicador de perfil (P):** Indicadores que se recomiendan para proporcionar estadísticas básicas e información de fondo para ayudar a las ciudades a determinar qué ciudades son de interés.
- **Indicador de apoyo (A):** Indicadores que se recomiendan para demostrar el desempeño en la prestación de servicios de la ciudad y la calidad de vida. Estos indicadores se pueden seleccionar de acuerdo con los objetivos de la ciudad.

Los grupos temáticos e indicadores alineados a la igualdad de género y a la lucha de violencia contra la mujer son:

2.2.1. Educación

- Porcentaje de la población femenina en edad escolar matriculada en las escuelas (C)

2.2.2. Gobierno

- Porcentaje total de mujeres elegidas para cargos en el estado (C)

2.2.3. Población y Condiciones Sociales

- Porcentaje de la población que son nuevos inmigrantes (P)

2.2.4. Seguridad

- Número de delitos violentos contra la mujer por cada 100.000 habitantes (A)

2.3. Analytics

Según Power et.al., (2018) existen múltiples definiciones que contempla el término analytics, entre ellas se encuentra:

1. La definición propuesta por Davenport y Harris (como se citó en Power et.al., 2018) es “Uso extensivo de datos, análisis estadístico y cuantitativo, modelos exploratorios y predictivos y gestión basada en hechos para impulsar decisiones y acciones”.
2. Por otro lado, la definición planteada por Wilder y Ozgur (como se citó en Power et.al. ,2018) es la siguiente, “La aplicación de procesos y técnicas que transforman los datos en bruto en información significativa para mejorar la toma de decisiones”.

3. Finalmente, INFORMS recomienda la definición: “La analítica es el proceso científico de transformar datos en información para tomar mejores decisiones” mencionada en Boyd (como se citó en Power et.al., 2018)

De acuerdo con Intel (2017), la analítica se clasifica en tradicional y avanzada, las cuales contienen los 5 tipos de analítica (ver figura 2.4) que pueden ser desarrolladas por las organizaciones para generar valor a través de uno de sus activos más valiosos, sus datos.

A continuación, se describe los tipos de analítica tradicional

2.3.1 Analítica descriptiva

Responde a la pregunta ¿Qué sucedió?, con el fin de identificar valores atípicos utilizando data histórica de diferentes tipos de fuentes que genera la organización, tales como, archivos planos, BD relacionales y no relacionales, entre otros.

2.3.2 Analítica de diagnóstico

Responde a las preguntas ¿Qué sucedió? y ¿Por qué sucedió?, tomando como base el análisis descriptivo, donde los especialistas encuentran el motivo por el cual se genera valores atípicos dentro de la organización, decidiendo así el lugar y momento correcto para utilizar la información obtenida.

Finalmente, se describe los tipos de analítica avanzada

2.3.3 Analítica predictiva

Responde a las preguntas ¿Qué sucederá?, ¿Cuándo sucederá? y ¿Por qué sucederá?, a través de la identificación de patrones que permiten realizar pronósticos de eventos y tendencias para una mejor toma de decisiones dentro de una organización.

2.3.4 Analítica prescriptiva

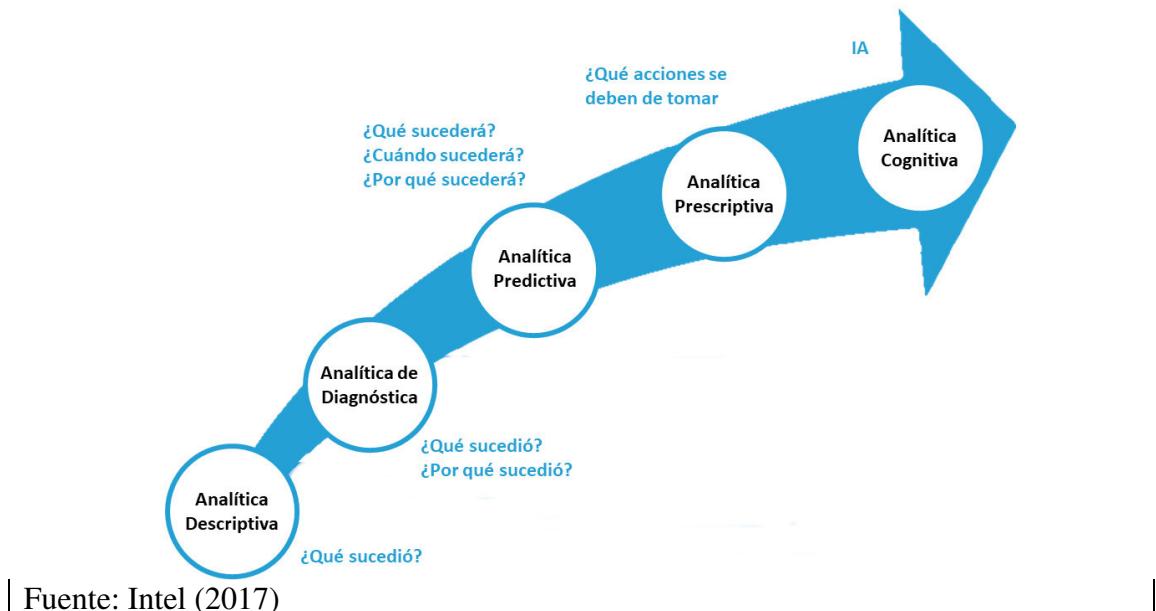
Responde a la pregunta ¿Qué acciones se deben de tomar?, en base a los pronósticos realizados en la analítica predictiva, con el fin de sugerir a la organización acciones que debe de aplicar para lograr cumplir sus objetivos.

2.3.5 Analítica cognitiva

Desarrolla un análisis de datos de alto rendimiento, usando por ejemplo tecnologías de inteligencia artificial para mejorar la eficiencia de la toma de decisiones dentro de una organización.

Figura 2.4.

Tipos de analítica.



2.4. Algoritmos predictivos supervisados

De acuerdo con Ruiz (2019), los algoritmos predictivos supervisados utilizan dos grupos de datos, uno de ellos para realizar el entrenamiento y el otro para la prueba. Para que el algoritmo realice una predicción correcta se debe de entrenar al modelo con el primer grupo de datos, y realizar los ajustes necesarios a los parámetros. Con el segundo grupo de datos se valida la precisión del modelo predictivo.

A continuación, se lista algunos algoritmos supervisados:

2.4.1 Bosques Aleatorios (Random Forest)

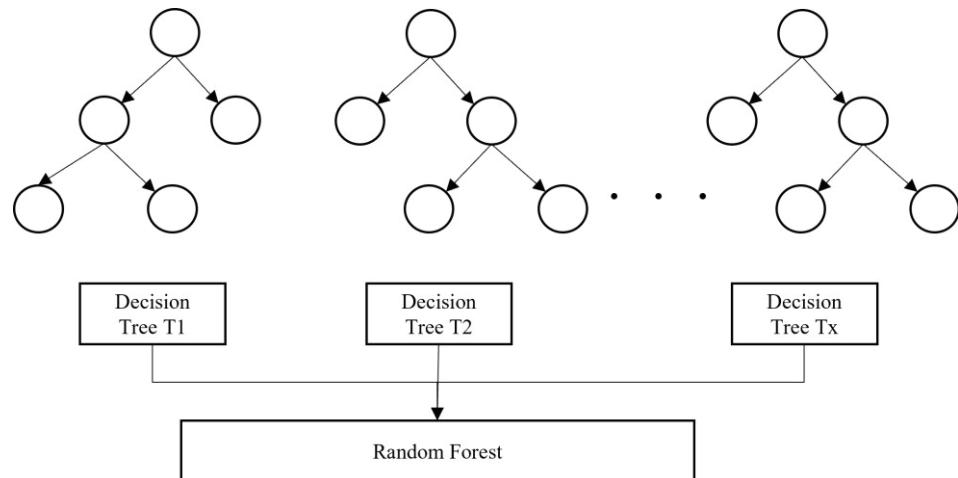
El algoritmo de Random Forest está basado en el uso de cierta cantidad de árboles de decisión ramificados en atributos (nodos de cada árbol) como se evidencia en la figura 2.5, los cuales permiten tener una mejor precisión predictiva y un control al sobreajuste de los datos (Scikit learn, 2022a).

Según Aguilar y Camargo, (2021), a pesar de las ventajas y la potencia de los árboles de decisión, el algoritmo de Random Forest se convierte en una alternativa para

disminuir la tasa de error y la inestabilidad de los árboles. Como se indica el párrafo anterior, el algoritmo de Random Forest, combina predictores del tipo de árboles de decisión, convirtiendo al bosque aleatorio en un modelo más robusto, con una mejor precisión y estabilidad en la estimación.

Figura 2.5.

Random Forest.

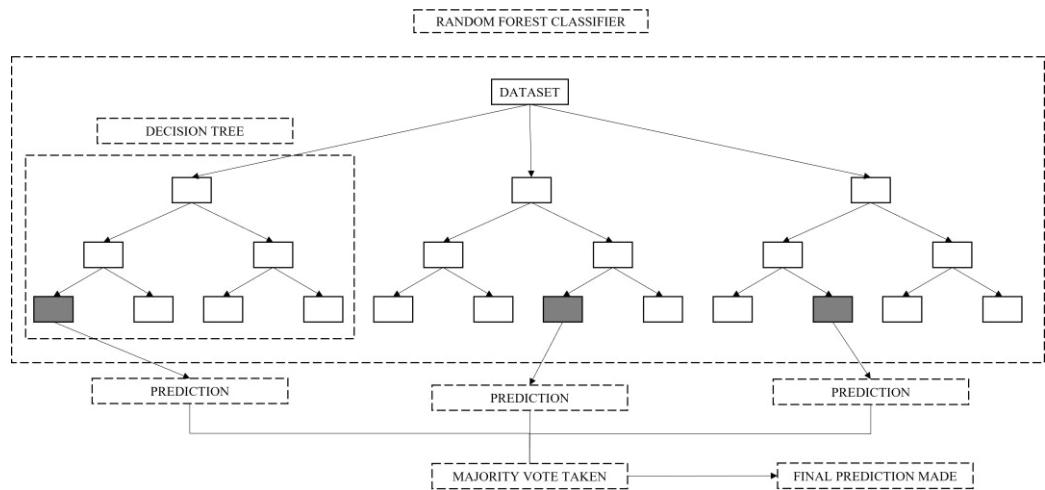


Fuente: Aguilar y Camargo (2021)

En la figura 2.6, se muestra el esquema realizado por Kashyap (2019), donde se detalla el proceso interno desarrollado por el algoritmo Random Forest al momento de realizar una predicción.

Figura 2.6.

Esquema del proceso de funcionamiento del Random Forest.



Fuente: Kashyap (2019)

Dentro de la documentación de Scikit learn (2022a), la biblioteca usada para Random Forest es `sklearn.ensemble.RandomForestClassifier`, la cual tiene los siguientes parámetros detallados en la figura 2.7.

Figura 2.7.

Parámetros de `sklearn.ensemble.RandomForestClassifier`.

sklearn.ensemble.RandomForestClassifier

```
class sklearn.ensemble.RandomForestClassifier(n_estimators=100, *, criterion='gini', max_depth=None,
min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='sqrt', max_leaf_nodes=None,
min_impurity_decrease=0.0, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False,
class_weight=None, ccp_alpha=0.0, max_samples=None)
[source]
```

Fuente: Scikit learn (2022a)

2.4.2 Regresión Logística

Según Bernachea et al., (2021), la regresión logística es usado como un método estadístico para realizar análisis multivariado, así como también poder ser usado como una técnica predictiva.

Siendo la regresión logística, una técnica estadística se utiliza la siguiente fórmula:

$$\text{Logit}(P) = \log \log (P) / (1 - p) = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

Siendo:

P : Probabilidad de evento de interés

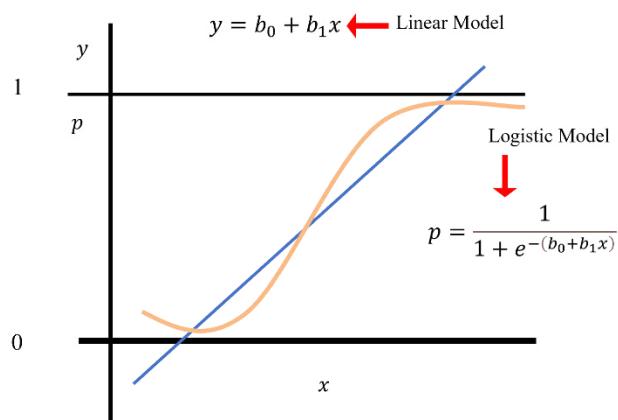
$b_0, b_1, b_2, \dots, b_n$: Los parámetros

$X_0, X_1, X_2, \dots, X_n$: Variables independientes

Despejando el valor de P se tendría una ecuación de 2 variables (regresión logística binaria), el cual es representado gráficamente en el plano xy , mostrado en la figura 2.8.

Figura 2.8.

Regresión Logística para 2 variables.



Fuente: Quiza (2018)

“La regresión logística de naranja viene diferenciada de la regresión lineal en azul, para diferenciar el llamado límite de decisión, la cual indica que el valor menor a 0.5 será igual 0, mientras que todo valor por encima de 0.5 será igual a 1.” (Bernachea et al., 2021).

Dentro de la documentación de Scikit learn (2022b), la biblioteca usada para regresión logística es `sklearn.linear_model.LogisticRegression`, la cual tiene los siguientes parámetros detallados en la figura 2.9.

Figura 2.9.

Parámetros de `sklearn.linear_model.LogisticRegression`.

`sklearn.linear_model.LogisticRegression`

```
class sklearn.linear_model.LogisticRegression(penalty='l2', *, dual=False, tol=0.0001, C=1.0, fit_intercept=True,  
intercept_scaling=1, class_weight=None, random_state=None, solver='lbfgs', max_iter=100, multi_class='auto', verbose=0,  
warm_start=False, n_jobs=None, l1_ratio=None)
```

[\[source\]](#)

Fuente: Scikit learn (2022b)

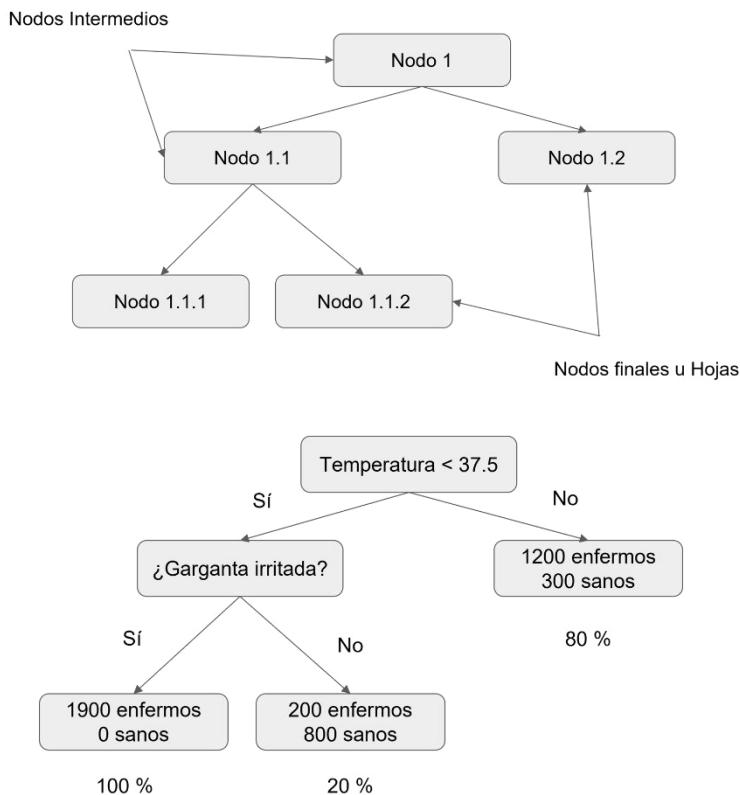
2.4.3 Árbol de decisión

Según Scikit learn (2022c), los árboles de decisión son algoritmos predictivos supervisados que pueden ser utilizados para la regresión o clasificación. Este algoritmo busca como finalidad la creación de modelos predictivos para una variable de destino, a través de reglas de decisión simples obtenidas de las características de los datos. Sus resultados son muy simples de interpretar.

Los nodos se clasifican en intermedios y finales u hojas (ver figura 2.10), estos últimos pueden ser puros (contienen elementos de una sola clase) o impuros (presentan ambas clases)

Figura 2.10.

Árbol de decisión.



Fuente: Elaboración Propia

Según Scikit learn (2022c), los árboles de decisión se dividen en:

- Árboles de decisión de clasificación: Las hojas del árbol son representadas mediante una variable categórica (clasificación). Donde la salida del modelo consiste en la proporción de cada una de las clases de cada hoja.
- Árboles de decisión de regresión: Las hojas del árbol son representadas mediante una variable numérica (regresión). Donde la salida del modelo consiste en la media de todos los valores de cada hoja.

Dentro de la documentación de Scikit learn (2022d), la biblioteca usada para Árbol de decisión es `sklearn.tree.DecisionTreeClassifier`, la cual tiene los siguientes parámetros detallados en la figura 2.11.

Figura 2.11.

Parámetros de `sklearn.tree.DecisionTreeClassifier`.

sklearn.tree.DecisionTreeClassifier

```
class sklearn.tree.DecisionTreeClassifier(*, criterion='gini', splitter='best', max_depth=None, min_samples_split=2,
min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, class_weight=None, ccp_alpha=0.0)
```

[source]

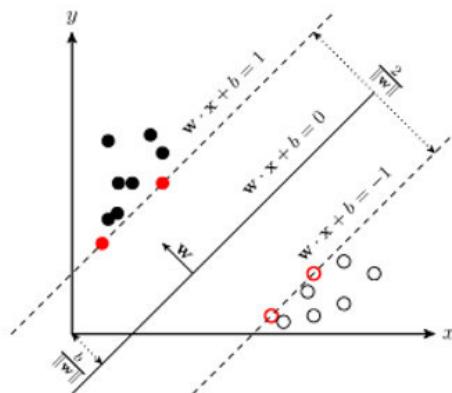
Fuente: Scikit learn (2022d)

2.4.4 Máquinas de Soporte Vectorial (SVM)

Según Aguilar y Camargo, (2021), las máquinas de soporte vectorial (SVM) fueron propuestas por Vapnik en 1992, para la resolución de problemas de clasificación de grandes volúmenes de datos complejos y ruidosos. La SVM tiene como objetivo principal establecer un hiperplano que separe los datos de entrada en subgrupos, de tal manera que se separe su distancia lo máximo posible de los valores de entrada más cercanos, los cuales son llamados vectores de soporte, ver figura 2.12.

Figura 2.12.

Máquina de Soporte Vectorial.



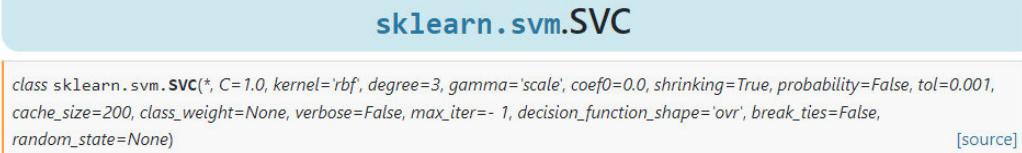
Fuente: Aguilar y Camargo (2021)

Maximizar la separación de los valores de entrada y el hiperplano lo convierte en un problema de programación cuadrática, el cual se resuelve mediante los multiplicadores de Lagrange para posteriormente hallar la solución para su problema dual.

Dentro de la documentación de Scikit learn (2022e), la biblioteca usada para Support Vector Machine es `sklearn.svm.SVC`, la cual tiene los siguientes parámetros detallados en la figura 2.13.

Figura 2.13.

Parámetros de `sklearn.svm.SVC`.



The screenshot shows the Python code definition for the `sklearn.svm.SVC` class. The code is as follows:

```
class sklearn.svm.SVC(*, C=1.0, kernel='rbf', degree=3, gamma='scale', coef0=0.0, shrinking=True, probability=False, tol=0.001, cache_size=200, class_weight=None, verbose=False, max_iter=-1, decision_function_shape='ovr', break_ties=False, random_state=None)
```

At the bottom right of the code block, there is a link labeled "[source]".

Fuente: Scikit learn (2022e)

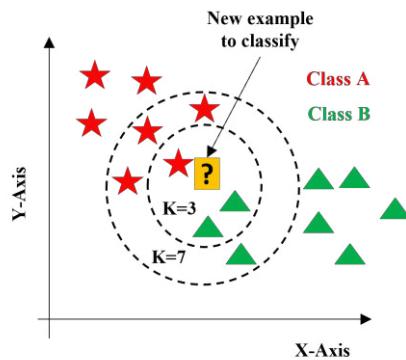
2.4.5 KNN

Según Bernachea et al., (2021), este algoritmo busca predecir a partir de la agrupación de los valores más cercanos de una muestra, además realiza una clasificación de acuerdo con el tipo de datos que lo rodean.

Como se muestra en la figura 2.14, el algoritmo realiza un cálculo a los elementos más cercanos al elemento a analizar y los etiqueta con el objetivo de predecir a qué grupo pertenece.

Figura 2.14.

Representación del algoritmo KNN.



Fuente: Bernachea et al. (2021)

Dentro de la documentación de Scikit learn (2022f), la biblioteca usada para KNeighborsClassifier es `sklearn.neighbors.KNeighborsClassifier`, la cual tiene los siguientes parámetros detallados en la figura 2.15.

Figura 2.15.

Parámetros de `sklearn.neighbors.KNeighborsClassifier`.

`sklearn.neighbors.KNeighborsClassifier`

```
class sklearn.neighbors.KNeighborsClassifier(n_neighbors=5, *, weights='uniform', algorithm='auto', leaf_size=30, p=2,
metric='minkowski', metric_params=None, n_jobs=None)
[source]
```

Fuente: Scikit learn (2022f)

Capítulo III: Estado del Arte

En este capítulo se aborda los artículos de investigación y/o tesis que plantean una solución ante la problemática expuesta anteriormente, detallando los criterios de búsqueda, selección y efectividad de la solución a través de indicadores.

La búsqueda de los artículos fue realizada mediante el uso de palabras claves, listadas a continuación, las cuales apoyaron a una selección más exhaustiva y rigurosa:

- Violencia contra la mujer – Violence against women
- Monitoreo y predicción de crimen y/o violencia
- Tecnologías de la información

Otro de los criterios de búsqueda, fue los años de publicación de los artículos y/o tesis, los cuales se encuentran entre el 2016 y 2021. Además, estos fueron extraídos de repositorios reconocidos por su nivel académico y prestigio, tales como:

- **Science Direct:** Base de datos de información científica sobre diversas áreas de investigación
- **IEEE:** Repositorio digital que ofrece acceso de literatura técnica de ingeniería y tecnología.
- **ACM:** Repositorio digital que alberga los trabajos de investigación exclusivamente del campo de la informática.
- **Repositorio UNMSM:** Repositorio digital de la Universidad Nacional Mayor de San Marcos (CyberTesis-UNMSM) que alberga los trabajos de investigación de la universidad mencionada.

- **Repositorio UNI:** Repositorio digital de la Universidad Nacional de Ingeniería que alberga los trabajos de investigación de la universidad mencionada.
- **Repositorio UPC:** Repositorio digital de la Universidad Peruana de Ciencias Aplicadas que alberga los trabajos de investigación de la universidad mencionada.
- **EBSCO:** Repositorio digital que ofrece acceso de trabajos de investigación de ciencias y humanidades.

3.1 Trabajos previos

La violencia contra la mujer es abordada por diferentes instituciones a nivel mundial, donde la tecnología de la información brinda soluciones para apoyar la atención de los casos y empoderamiento de las mujeres.

En esta sección se detalla las diferentes tecnologías, identificadas dentro la literatura, para evaluar el impacto de sus soluciones alineados a la lucha contra la violencia y el crimen, además, se presenta una comparativa entre estas.

3.1.1 Internet de las Cosas - Monitoreo y Predicción

Kumar et al., (2019)

I. Características:

- Aporta a mejorar la calidad de vida de las personas a través del uso e interacción de dispositivos inteligentes.
- Es una tecnología Open Hardware y Open Software.
- Es de gran utilidad dentro del desarrollo y construcción de ciudades inteligentes.
- Permite realizar seguimiento y prevención en tiempo real a través de la recopilación, identificación y detección de datos.

II. Ventajas:

- Capacidad de interconectar dispositivos a través de internet o Gateway, de forma rápida y en tiempo real.
- Facilidad de captación de datos con el objetivo de realizar un adecuado monitoreo.

III. Desventajas:

- Vulnerabilidad de la seguridad y privacidad de los dispositivos IoT debido a que los gateways pueden ser usados para fines delictivos.
- Dificultad en la construcción o implementación de soluciones IoT debido a las diferentes configuraciones de los dispositivos.

IV. Casos de estudio

A. Woman Safety Device Designed using IoT and Machine Learning

Muskan et al., (2018) proponen el desarrollo de una solución tecnológica que permita detectar si una mujer se encuentra en peligro a partir de la predicción de los niveles de temperatura y frecuencia cardiaca registrados en situaciones que la pueden poner en peligro. Lo que buscan los autores es que el dispositivo no tenga que ser accionado por un humano para que pueda funcionar y alertar, adicionalmente plantean que la solución puede funcionar con o sin Internet.

El sistema propuesto busca alertar a los contactos de emergencia de la mujer víctima cuando se encuentre en peligro, a través de mensajes y llamadas telefónicas. Para poder determinar si la mujer se encuentra en peligro, el sistema posee sensores de temperatura y pulso que constantemente están monitoreando los números registrados, lo cual permite que el algoritmo de predicción pueda aprender de los patrones emitidos por la posible víctima y determinar si se encuentra o no en peligro.

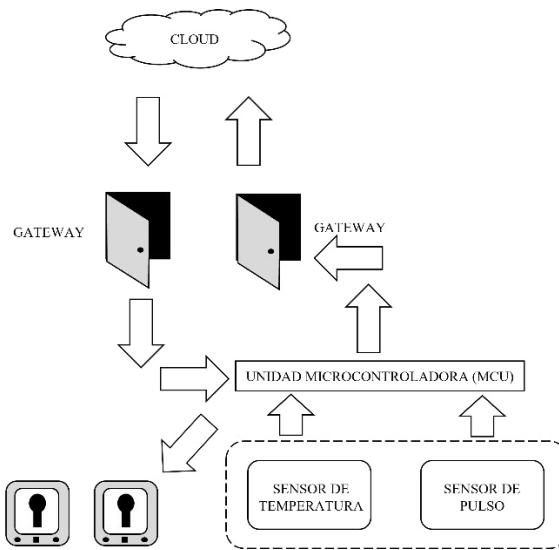
La arquitectura propuesta por los autores se muestra en la figura 3.1, donde se puede observar las diferentes capas de una arquitectura IOT.

- Capa de dispositivos físicos y controladores: En esta capa se encuentran los sensores que permitirán obtener los datos, que podrán ser consultados o controlados de acuerdo con el resultado de la evaluación del algoritmo.
- Conectividad: Esta capa permite la conexión entre la capa física y la entrega de información, para el sistema propuesto por los autores, utilizan el módulo de Zigbee.
- Edge computing: Esta capa permite evaluar y procesar los datos recopilados sin la necesidad de ir a niveles más altos. Lo que apoya al tiempo de respuesta.

- Data Accumulation: Esta capa permite captura y almacenar los datos para que puedan ser usados como consulta en diversas aplicaciones.

Figura 3.1.

Arquitectura del sistema propuesto.



Fuente: Muskan et al. (2018)

Para la parte predictiva, los autores eligen el algoritmo de aprendizaje de regresión logística para poder predecir si la víctima se encuentra en peligro, para poder entrenar y probar su modelo utilizaron 500 registros obtenidos de diferentes aplicaciones móviles, midiendo el pulso y temperatura de una persona después de correr en circunstancias específicas. Logrando así un accuracy de 78.8% para su data de entrenamiento y 74.3% para su data de prueba.

B. An IOT-Based Architecture for Crime Management in Nigeria

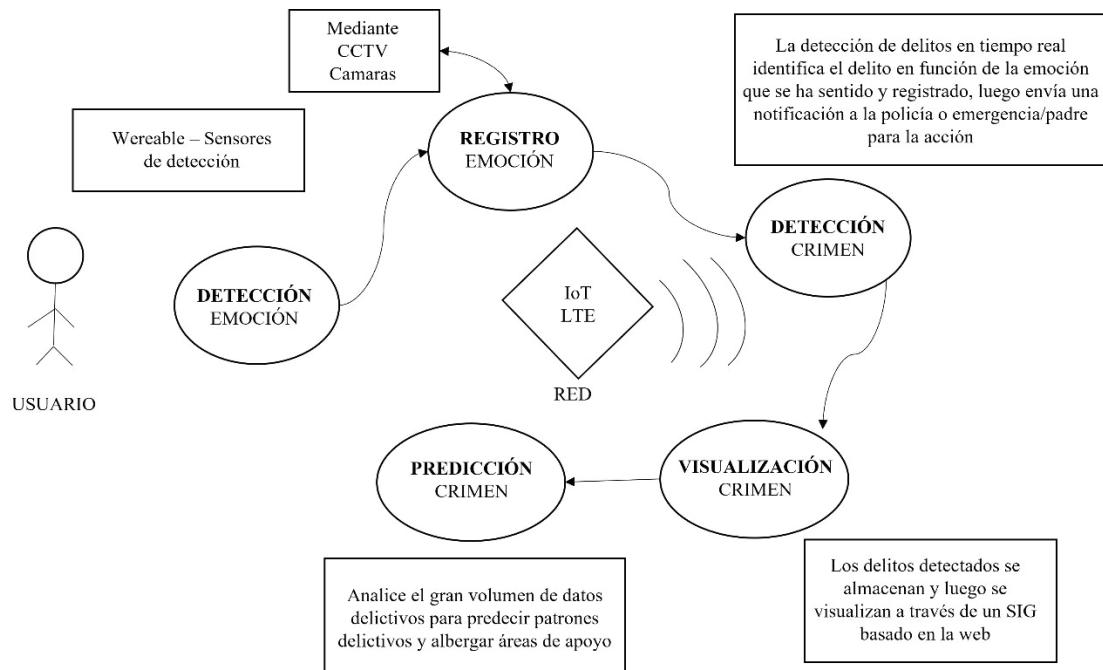
Adesola et al., (2019) desarrollan una solución de TI basado en Internet de las Cosas y Big Data, con el fin de rastrear y monitorear la delincuencia en tiempo real en Nigeria, país que muestra un creciente índice de casos a través de robo con armas, asesinato secuestro o violación.

El modelo conceptual de la solución propuesta se describe a continuación (ver Figura 3.2):

1. Módulo de detección de estado de emoción del usuario, el cual está equipado con un dispositivo portable que contiene un sensor inteligente.
2. Módulo de registro de estado de emoción, almacena el tipo de delito detectado por el sensor.
3. Módulo de detección de delitos, envía alertas a la autoridad policial más cercana o los familiares de la víctima.
4. Módulo de visualización de delitos, presenta los delitos almacenados en una Sistema de Información Geográfica (SIG)
5. Módulo de predicción de delitos a través de un gran volumen de datos almacenados.

Figura 3.2.

Modelo de monitoreo y predicción del crimen propuesto.



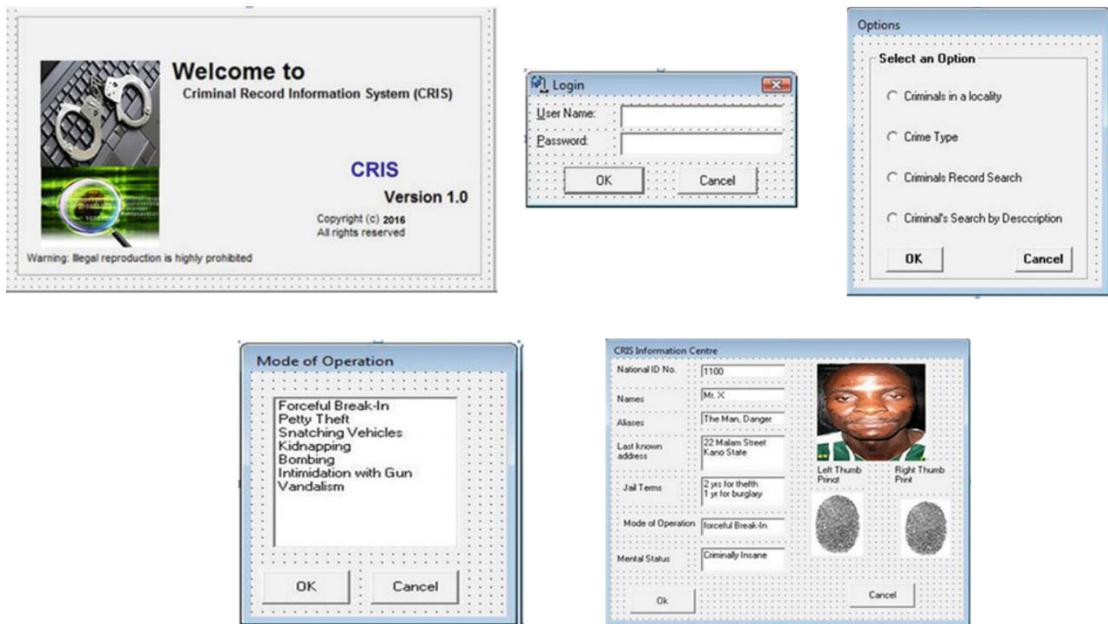
Fuente: Adesola et al. (2019)

La implementación del modelo conceptual presentado se describe a través de una prueba de concepto con las siguientes funcionalidades del sistema (ver Figura 3.3):

1. Pantalla de bienvenida y el login para el usuario
2. Menú de opciones para el administrador del sistema, el cual incluye
 - Criminales en una localidad
 - Tipo de delito
 - Búsqueda de registros de delincuentes
 - Búsqueda de delincuentes por descripción

Figura 3.3.

Pantallas de la implementación del modelo.



Fuente: Adesola et al. (2019)

Concluyendo que su sistema en comparación al existente “Crimen tracker”, cuenta con mejor funcionalidad como el GPS, el SIG para rastrear delincuentes y en cuanto al costo es más barato ya que está elaborada en una plataforma de código abierto. Además, cumple con el objetivo planteado, de detectar y monitorear los delitos en Nigeria.

3.1.2 Analítica aplicada a monitoreo y predicción

Power et.al., (2018)

I. Características:

- Los proyectos de analítica utilizan la data como insumo para desarrollar soluciones, siendo esta uno de los activos más importantes dentro de una organización.

- Generar visualizadores luego del análisis de data histórica que permite a los expertos de una organización identificar problemas y tomar mejores decisiones.
- Uso de data histórica para descubrir e interpretar patrones de comportamiento y predecir posibles escenarios.

II. Ventajas:

- Visibilidad de los problemas, lo que permite poner atención a los puntos de dolor en las organizaciones.
- Su implementación permite gestionar y monitorear procesos para una mejor toma de decisiones.
- Capacidad de adaptación de la solución en un entorno cambiante ocasionado por las preferencias de los usuarios.
- Alineado a la analítica predictiva, Makond y Mayuening (2019), recomienda aplicar modelos de machine learning frente a redes neuronales, para casos similares al planteado en la presente tesis, debido a su practicidad e interpretabilidad.

III. Desventajas:

- Necesidad de constante actualización de herramientas o implementaciones debido a la rápida variabilidad de las tecnologías o tipo de datos de la organización.
- El inadecuado tratamiento y uso de la data analizada puede vulnerar la privacidad de las personas u organizaciones.

- La implementación de los proyectos analíticos se da en su mayoría en grandes empresas debido al costo.

IV. Casos de estudio

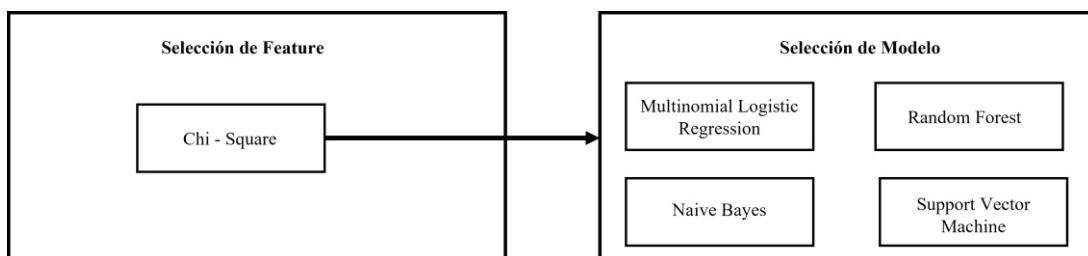
A. Comparison of Classifiers Models for Prediction of Intimate Partner Violence

Guerrero et al., (2021) evidencia que no hay muchas investigaciones en el Perú que hayan realizado experimentos para poder identificar la ocurrencia de un tipo de violencia hacia la mujer. Por tal motivo, los autores proponen un flujo a seguir para poder identificar las variables más significativas y así poder realizar una buena predicción de violencia hacia la mujer.

Cabe resaltar que la data utilizada en la investigación está conformada por 70510 denuncias del Ministerio de la mujer del año 2016 y cada una cuenta con 236 variables. Estas denuncias participarán en un flujo diseñado por los autores como se muestra en la Figura 3.4.

Figura 3.4.

Flujo para predecir el índice de violencia de pareja.



Fuente: Guerrero et al. (2021)

A continuación, se detalla los pasos del flujo presentado en la figura anterior:

- Paso 1: Se realiza el preprocesamiento de datos, eliminando registros que tengan más del 70 % de valores perdidos o nulos, también se eliminó las variables que no tengan que ver con violencia. Se consideró solo casos donde la persona agredida es mujer y el agresor es un hombre.
- Paso 2: Se utiliza Cramer V para eliminar las variables que tienen poca importancia para el modelo.
- Paso 3: Se utilizar chi² para organizar las variables según su importancia.
- Paso 4: Realiza “one hot encoding” para la transformación de variables categóricas en variables nominales.
- Paso 5: División de la data en subgrupos de 6,12,18,24,32 variables (podemos poner las 32 variables).
- Paso 6: Validación cruzada
 - ✓ Aplicación de Smote para balancear la data a partir de clases minoritarios.
 - ✓ Aplicación de los modelos de clasificación.
 - ✓ Validación de los modelos de clasificación. (ver Tabla 3.1)

Tabla 3.1.

Resultados de la aplicación de los métodos predictivos.

		6 variables			
Feature Selection Technique	Classifier	Accuracy	Precision	Recall	F1
X ²	Multinomial Logistic Regression	61.59% +/-2.22%	61.62% +/-2.07%	61.59% +/-2.22%	61.34% +/-2.18%
	Random Forest	61.15% +/-2.66%	61.30% +/-2.27%	61.15% +/-2.66%	60.46% +/-2.63%
	Naive Bayes	58.42% +/-1.58%	58.86% +/-1.51%	58.42% +/-1.58%	58.43% +/-1.56%
	SVM	60.92% +/-2.55%	61.13% +/-2.55%	60.92% +/-2.55%	60.20% +/-2.65%

		12 variables			
Feature Selection Technique	Classifier	Accuracy	Precision	Recall	F1
X ²	Multinomial Logistic Regression	62.67% +/-2.47%	62.68% +/-2.30%	62.67% +/-2.47%	62.33% +/-2.37%
	Random Forest	60.88% +/-8.68%	60.86% +/-8.65%	60.88% +/-8.68%	60.76% +/-8.70%
	Naive Bayes	59.11% +/-1.63%	59.83% +/-1.49%	59.11% +/-1.63%	59.18% +/-1.54%
	SVM	63.07% +/-2.50%	63.36% +/-2.33%	63.07% +/-2.50%	62.34% +/-2.20
		18 variables			
Feature Selection Technique	Classifier	Accuracy	Precision	Recall	F1
2X ²	Multinomial Logistic Regression	63.24% +/-2.53%	63.31% +/-2.31%	63.24% +/-2.53%	62.81% +/-2.43%
	Random Forest	61.84% +/-10.00%	61.78% +/-10.00%	61.84% +/-10.00%	61.68% +/-10.04%
	Naive Bayes	59.84% +/-1.77%	60.71% +/-1.64%	59.84% +/-1.77%	59.90% +/-1.69%
	SVM	63.37% +/-2.51%	63.32% +/-2.35%	63.37% +/-2.51%	62.68% +/-2.35%
		24 variables			
Feature Selection Technique	Classifier	Accuracy	Precision	Recall	F1
X ²	Multinomial Logistic Regression	63.21% +/-2.66%	63.21% +/-2.48%	63.21% +/-2.66%	62.77% +/-2.53%
	Random Forest	62.90% +/-10.26%	62.85% +/-10.26%	62.90% +/-10.26%	62.74% +/-10.31%
	Naive Bayes	59.80% +/-1.73%	60.62% +/-1.66%	59.80% +/-1.73%	59.80% +/-1.64%
	SVM	63.46% +/-2.49%	63.70% +/-2.31%	63.46% +/-2.49%	62.78% +/-2.33%
		32 variables			
Feature Selection Technique	Classifier	Accuracy	Precision	Recall	F1
X ²	Multinomial Logistic Regression	63.22% +/-2.53%	63.30% +/-2.32%	63.22% +/-2.53%	62.78% +/-2.38%
	Random Forest	63.36% +/-10.39%	63.28% +/-10.40%	63.36% +/-10.39%	63.17% +/-10.41%
	Naive Bayes	59.79% +/-1.78%	60.60% +/-1.69%	59.79% +/-1.78%	59.80% +/-1.78%
	SVM	63.33% +/-2.40%	63.57% +/-2.18%	63.33% +/-2.40%	62.68% +/-2.25%

Fuente: Guerrero et al. (2021)

- Paso 7: Comparación de los resultados de los modelos

Se puede concluir que el modelo más preciso es SMV aplicado con 18 variables, obteniendo una precisión del 63,37 % y un F1, que considera la sensibilidad de falsos positivos y negativos, de 62.68 %.

B. Predictive Model Based on Machine Learning for the Detection of Physically Mistreated Women in the Peruvian Scope

Saboya et al., (2019), desarrolla un modelo predictivo en el ámbito peruano, país que cuenta con un alto índice de violencia contra la mujer, con el objetivo de identificar y alertar si una mujer será maltratada en un futuro.

Para el desarrollo del modelo predictivo (ver Figura 3.5), realizaron las siguientes fases:

1. Recopilación de datos

La extracción de los datos estructurados del periodo 2005 - 2018 se realizó desde el portal web del Instituto Nacional de Estadística e Informática del Perú (INEI), específicamente de la base de datos de la Encuesta Demográfica y de Salud Familiar (ENDES) que se encuentran en dicho portal.

2. Preparación de datos

2.1. Comprensión de datos

En esta etapa se analizaron y validaron las variables con el apoyo de especialistas en psicología y sociología con experiencia en atención de casos de violencia contra la mujer.

Siendo las variables seleccionadas, las siguientes:

- Variables que describen el entorno material y cultural que habita el sujeto.
- Variables que denotan el perfil sociodemográfico del individuo.
- Nivel de educación del individuo.
- Situación económica del individuo.
- Variables que informan sobre maltrato físico.
- Variable binaria que identifica si la mujer recibió maltrato físico o no.

2.2. Limpieza de datos

En esta etapa se realizaron integración y procesamiento de datos con el objetivo de obtener información significativa, además desarrollaron transformaciones de las variables y verificaron los datos atípicos dentro del conjunto de datos.

3. Desarrollo de modelos predictivos

Los autores realizaron la distribución aleatoria de los datos de la siguiente manera: 40% para el entrenamiento, 40% para la validación y 20% para las pruebas.

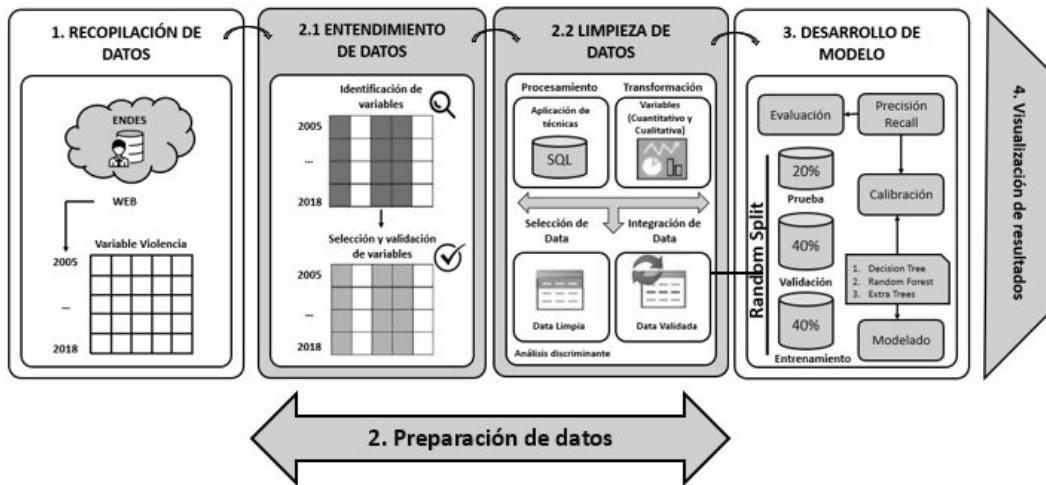
Utilizaron los siguientes algoritmos supervisados de Machine Learning para la construcción del modelo predictivo:

- Random Forest Classifier
- Decision Tree Classifier
- Extra Trees Classifier

Los cuales fueron evaluados en precisión, recall y score, luego de pasar por un proceso de calibración usando la biblioteca GridSearchCV para obtener los mejores parámetros.

Figura 3.5.

Proceso metodológico del modelo predictivo.



Fuente: Saboya et al. (2019)

Finalmente presentaron los siguientes resultados (ver Tabla 3.2) respecto a la evaluación de los tres modelos probados:

Tabla 3.2.

Resumen de resultados de los modelos.

		Precisión	Recall	F1 - Score
Random Forest Classifier Score = 0.6420	NO	0.70	0.79	0.74
	YES	0.51	0.40	0.45
Decision Tree Classifier Score = 0.6217	NO	0.70	0.74	0.72
	YES	0.48	0.44	0.46
Extra Trees Classifier Score = 0.6378	NO	0.70	0.78	0.74
	YES	0.50	0.40	0.44

Fuente: Saboya et al., (2019)

Como se evidencia en la tabla anterior, el mejor modelo de predicción es el Random Forest Classifier con una precisión del 51% y un recall de 40%.

C. Crime Against Women: Analysis and Prediction

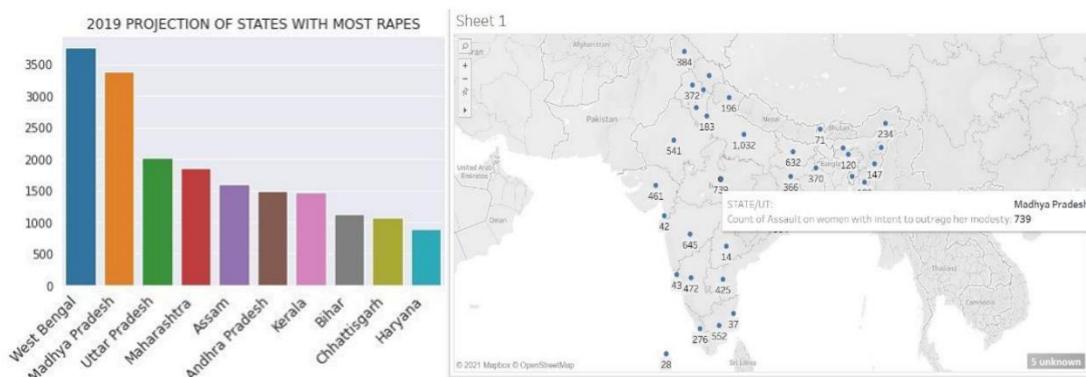
Nair et al., (2021), trabajo de investigación desarrollado en la India, país donde diariamente el número de delitos en contra de las mujeres aumenta desenfrenadamente. Ante este problema social que afecta a las mujeres de la India y del mundo, esta investigación propone una solución tecnológica para analizar y predecir los diferentes tipos de delitos en cada estado de la India usando la Regresión de Huber, mostrando los datos resultantes en forma gráfica y logrando así identificar el/los delito(s) dominante(s) en cada estado.

La solución fue desarrollada a través de las siguientes fases:

1. Conjunto de datos: Datos de los diferentes delitos ocurridos en la India como violación, secuestro, muerte, insulto, crueldad, tráfico inmoral, entre otros, en el periodo de 2007 a 2018, y delitos predichos para el 2019.
2. Conversión de datos de series de tiempo en aprendizaje supervisado sensibles al contexto: Aplican el modelo lineal de regresión de Huber para la obtención de las predicciones de delitos para el 2019, usando el conjunto de datos del 2007 al 2018.
3. Iterando para obtener puntajes y predicciones para cada estado: En base al modelo predictivo encontrado anteriormente, en esta parte iteran para lograr los mejores resultados del modelo.
4. Visualización: Muestran gráficos con los datos obtenidos luego de la predicción de los estados con más violaciones, agresiones, etc. (ver Figura 3.6).

Figura 3.6.

Gráficos de barras con el número de delitos por estado y mapa de calor para representar el mayor número de asaltos en un estado.



Fuente: Nair et al., (2021)

3.1.3 Redes Neuronales - Seguimiento y Predicción

Aggarwal (2018)

I. Características:

- La implementación de la solución es un modelo matemático que permite realizar un procesamiento interconectado, emulando las neuronas.
- Capacidad de recordar y generalizar lo aprendido en el modelo de predicción a partir de los datos ingresados.

II. Ventajas:

- Capacidad de realizar un aprendizaje automático y toma de decisiones ante casuísticas similares.
- Capacidad de realizar análisis de procesamiento en paralelo.
- Tolerancia a fallos frente a una afectación de una parte de la red.

III. Desventajas:

- Mayor dependencia con la capacidad de los procesadores para la implementación de soluciones.
- Dificultad de interpretación del comportamiento de la red al brindar la predicción, debido a que se desconoce los pasos de como llegó a la solución.

IV. Casos de estudio

A. An intelligent system for complex violence pattern analysis and detection

Ullah et al., (2021) proponen en esta investigación un eficiente método de detección de violencia (VD siglas en inglés), con el fin de detectar, monitorear y prevenir las actividades delictivas en entornos interiores como exteriores en tiempo real, ofreciendo, así como consecuencia una ambiente pacífico y seguro.

El método de VD presentado en este artículo de investigación, consta de cuatro principales pasos:

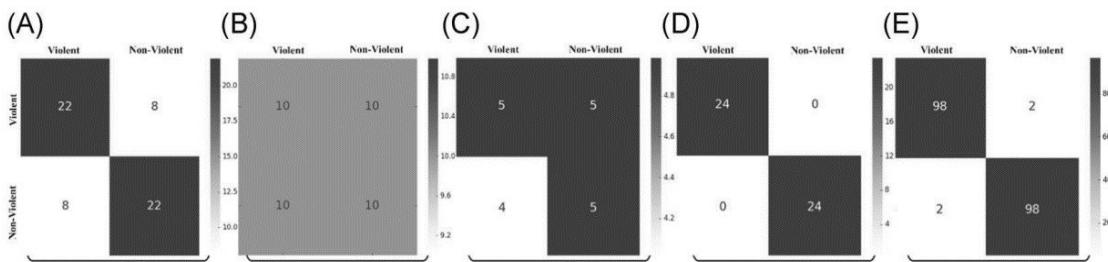
1. Preprocesamiento para la detección de objetos, los cuales para este artículo son las personas y autos debido a su importante relevancia para detección de violencia. Para ello usa la máscara R-CNN que proporciona etiquetas para cada objeto.
2. Aumento de datos antes del entrenamiento del modelo, debido al desequilibrio y falta de datos en las clases, se usa la técnica de (IMGAUG) para el aumento de imágenes.
3. Extracción de features, se realiza a través de dos fases, la primera es el de Flujo Óptico Temporal y la segunda el modelo Darknet - 19, concatenando las dos fases que sirven de input para las redes de memoria a corto plazo largo (LSTM siglas en inglés).

4. Secuencia de aprendizaje mediante LSTM apilado, debido al problema de fuga de la red neuronal recurrente (RNN siglas en inglés), usa una especial que es el de memoria a corto plazo (LSTM)

La implementación del modelo propuesto utilizó un conjunto de datos públicos de VD, los cuales consisten en videos de escenas violentas y no violentas, además de videos con escenas al interior y exterior de Surveillance Fight, teniendo los siguientes resultados (Ver figura 3.7):

Figura 3.7.

Matriz de confusión de cada conjunto de datos.



Nota. Donde (A) Surveillance Fight destaca claramente los valores clasificados correcta e incorrectamente de cada clase mientras que (B) y (C) explican el rendimiento del método propuesto sobre vigilancia interior y exterior. Los conjuntos de datos de Violent Flow y Hockey Fight se visualizan en (D) y (E), respectivamente. Fuente: Ullah et al., (2021)

En el gráfico se puede observar un mapa de calor, donde en (A) para el conjunto de datos de Surveillance Fight muestra una clasificación correcta para la clase violenta y no violenta con un 73,33 % y 73,33 % respectivamente. En (B) y (C) muestra el mapa de calor para el conjunto de datos de vigilancia al interior y exterior. Por último, el conjunto de datos Violent Flow y Hockey Fight se observa en los gráficos (D) y (E),

respectivamente, donde se puede visualizar una clasificación correcta de la clase violenta y no violenta. Concluyendo que estos cuadros resaltan la eficacia del método presentado.

La Tabla 3.3 muestra la comparación del método propuesto con otros métodos avanzados en base al conjunto de datos de Hockey Fight (A) y Violent Flow (B), demostrando una mayor efectividad el método propuesto. En cuanto a la Figura 3.8 presentada el eje vertical representa la precisión en porcentaje, mientras que el eje horizontal lista los métodos avanzados y el método propuesto identificado como P-M.

Tabla 3.3.

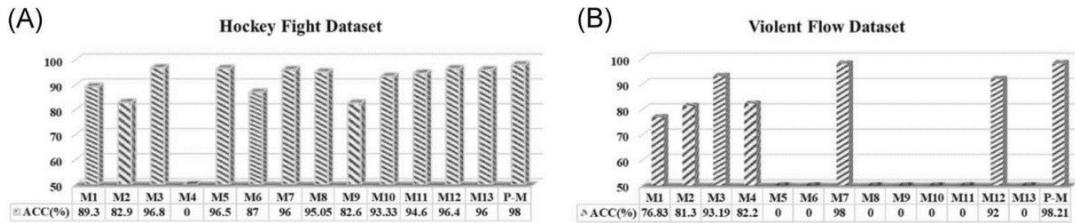
Comparación detallada de los resultados obtenidos en el conjunto de datos Violent Flow, Hockey Fight y Surveillance Flight.

ID	Techniques	Ref.	Method	Data set		
				Violent Flow (%)	Hockey Fight (%)	Surveillance Fight (%)
M1	Traditional Learning	45	HOMO-SVM	76.83	89.3	-
M2		44	ViF	81.3	82.90	-
M3		46	Motion-IWLD	93.19	96.8	-
M4		47	s-HOT	82.2	-	-
M5		48	MoB-SIFT	-	96.5	-
M6	Deep Learning	49	Mobile Net	-	87.0	-
M7		50	3D-CNN-STF	98	96	-
M8		51	HOG3D-BoVW	-	95.05	-
M9		52	Cuboid Trajectories – Hough Forests	-	82.6	-
M10		53	Inception ResNet Model	-	93.33	-
M11		8	Hough 2D – CNN	-	94.6	-
M12		54	CNN	-	96.40	-
M13		42	Fight – CNN + BD – LSTM	-	96	72
P-M			Proposed method	98.21	98	74

Fuente: Ullah et al., (2021)

Figura 3.8.

Análisis comparativo del método propuesto con métodos avanzados.



Nota. Donde el gráfico comparativo en (A) es de Hockey Fight mientras que (B) es de Violent Flow. Fuente: Ullah et al., (2021)

Finalmente, el artículo concluye la importancia de las cámaras de videovigilancia junto a los métodos y/o aplicaciones automáticas para el monitoreo y detección de violencia como importantes para reducir la tasa de criminalidad. Además, indicó el eficiente método de VD desarrollado en esta investigación, en cuanto a la detección en videos de escenas violentas, notificando a los usuarios interesados para tomar alguna acción, y no violentas. En números el método presentado obtuvo una precisión de 98,21 %, 98% y 74% en el conjunto de datos de Violent Flow, Hockey Fight y Surveillance Fight respectivamente; y la precisión para vigilancia interior y exterior es del 53,68 % y 49%.

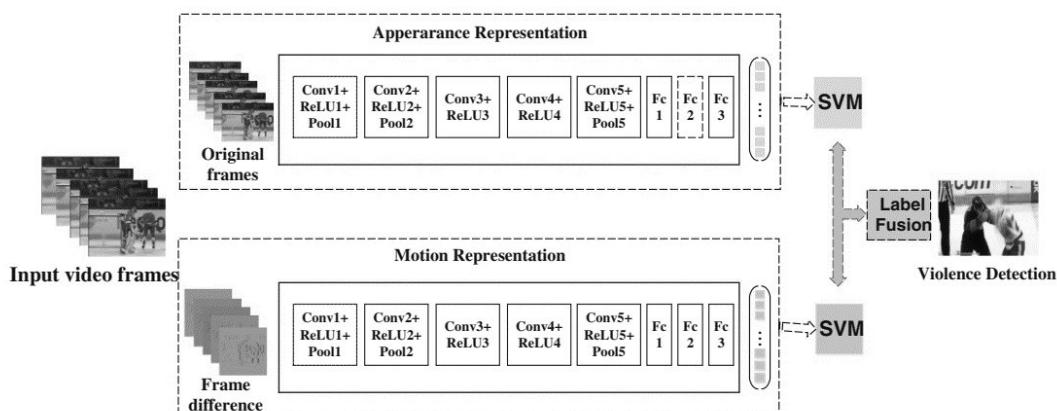
B. Real Time Violence Detection Based on Deep Spatio-Temporal Features

Xia et al., (2018) plantea un modelo predictivo basado en redes neuronales convolucionales (CNN) y SVM, los cuales permitirán detectar agresiones a partir de videos de seguridad, permitiendo así alertar a las personas responsables de mantener el orden en colegios, prisiones o lugares públicos.

Como se muestra en la figura 3.9, los autores proponen extraer las características de movimiento y apariencia a través de fotogramas fijos de videos, para obtener ello utilizan redes neuronales convolucionales profundas, las cuales trabajan bien con ese tipo de escenario, luego de conseguir las características descritas se busca entrenarlas por separado, con el objetivo de reducir la influencia entre ellas, utilizando el algoritmo de clasificación SVM, este modelo predictivo permite obtener 2 etiquetas (de movimiento y de apariencia) las cuales mediante el método de fusión de etiquetas permitirán predecir si es o no violencia.

Figura 3.9.

Descripción general de la red bicanal propuesto.



Fuente: Xia et al., (2018)

Para probar el algoritmo propuesto los autores utilizan imágenes de una pelea de hockey y de una multitud violenta, lo cual permite tener la siguiente precisión (ver tabla 3.4), adicionalmente verifican el rendimiento y estabilidad del modelo predictivo descrito en el párrafo anterior, a través de la validación cruzada como se muestra en la figura 3.10

Tabla 3.4.

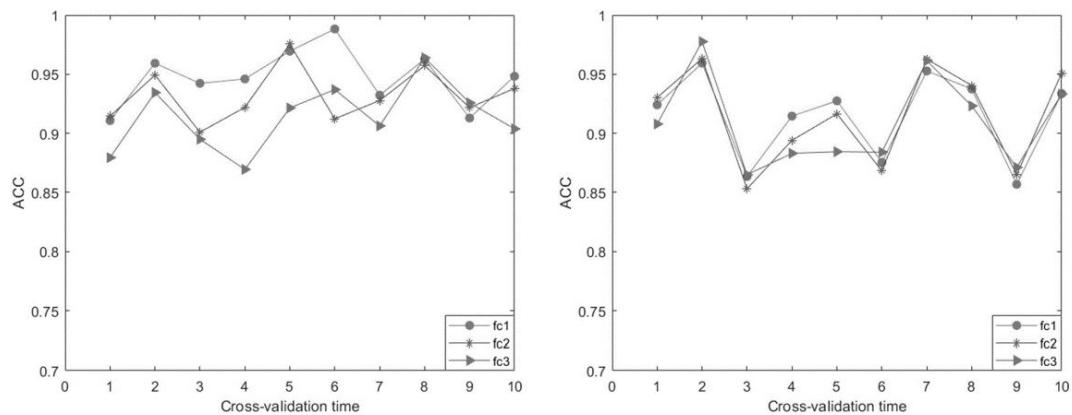
Comparación del rendimiento entre los diferentes conjuntos de datos.

ACC \pm SD(%)	Hockey fight	Violent crowd
Appearance accuracy	94.72 ± 3.6	91.46 ± 2.69
Motion accuracy	84.56 ± 4.85	75.87 ± 4.07
Fusion accuracy	95.90 ± 3.53	93.25 ± 2.34

| Fuente: Xia et al., (2018) |

Figura 3.10.

Validación cruzada del modelo predictivo propuesto



| Fuente: Xia et al., (2018) |

Finalmente, los autores buscan comparar el método propuesto y aplicado sobre imágenes de una pelea de hockey y de una multitud violenta con otros métodos avanzados, como se muestra en la figura 3.11, de la cual se pudo concluir que de acuerdo con las pruebas realizadas, el método propuesto en su investigación supera a los métodos existentes debido a que su propuesta logró un mejor rendimiento en tiempo real para el conjunto de datos de violencia en multitudes, sin embargo para el conjunto de datos de la pelea de hockey, el método ganador fue diferente al método propuesto, debido a que los comportamientos violentos a detectar se presentan en una distancia muy corta, dicha característica es utilizada por el algoritmo ganador para contar con una mejor precisión.

Figura 3.11.

Tablas de comparación del método propuesto vs métodos avanzados para los dos conjuntos de datos.

Tabla 3. Comparación de métodos propuestos y otros en el conjunto de datos de violencia de multitudes.

Métodos	CAC ± DAKOTA DEL SUR(%)
STIP (HoG)	57.43 ± 0.37
STIP (HoF)	58.53 ± 0.32
MoSIFT	83.42 ± 8.03
ViF	81.3 ± 0.21
Dos flujos	91.83 ± 3.34
LaSIFT	93.12 ± 8.77
Dos flujos+IDT	92.5
Propuesto	93.25 ± 2.34

Tabla 4. Comparación de métodos propuestos y otros en el conjunto de datos de Hockey Fight.

Métodos	CAC ± DAKOTA DEL SUR(%)
STIP (HoG)	91.7
STIP (HoF)	88.6
MoSIFT	90,9
ViF	82.90 ± 0.14
Dos flujos	93
LaSIFT	94.42 ± 2.82
Dos flujos+IDT	98
Propuesto	95.90 ± 3.53

Fuente: Xia et al., (2018)

3.2. Determinación de la solución

Para el desarrollo de esta tesis, se definió las siguientes métricas, cuyo objetivo es apoyar a elegir y sustentar que solución es la más idónea para resolver el problema planteado

I. Interpretación de la solución

- A. Descripción:** Capacidad del desarrollador de interpretar el resultado obtenido de una solución TIC.
- B. Prioridad:** Media
- C. Medición:** De 1 a 5, donde un mayor valor indica menor complejidad al interpretar la solución.

II. Usabilidad

- A. Descripción:** Indica el nivel de facilidad de uso e interacción del usuario con cada solución.
- B. Prioridad:** Media
- C. Medición:** De 1 a 5, donde un mayor valor indica mayor usabilidad.

III. Disponibilidad

- A. Descripción:** Indica la capacidad de cada solución para ser usado en todo tiempo, todo lugar y todo medio
- B. Prioridad:** Alta
- C. Medición:** De 1 a 5, donde un mayor valor indica una mayor disponibilidad.

IV. Seguridad

- A. Descripción:** Indica el nivel de seguridad que aplica cada solución para proteger la privacidad de los datos.
- B. Prioridad:** Alta
- C. Medición:** De 1 a 5, donde un mayor valor indica una mayor seguridad.

V. Escalabilidad

- A. Descripción:** Es la capacidad de cada solución para adaptarse al aumento de requerimientos.
- B. Prioridad:** Media
- C. Medición:** De 1 a 5, donde un mayor valor indica una mayor escalabilidad.

A continuación, después de la evaluación de los valores, se muestra la Tabla 3.5 correspondiente a los resultados de la solución más idónea para resolver el problema planteado.

Tabla 3.5.

Resultados de la evaluación.

Criterios	IOT – Monitoreo y predicción	Analítica – Monitoreo y predicción	Red Neuronal – Monitoreo y predicción
Interpretación de la solución	4	4	2
Usabilidad	4	4	4
Disponibilidad	3	4	4
Seguridad	2	3	3
Escalabilidad	3	4	4
Total	16	19	17

Fuente: Elaboración Propia

De acuerdo con la tabla anterior, se puede concluir que la tecnología elegida es analítica, al tener un total de 19 puntos, superando por 2 y 3 puntos a la solución de Red Neuronal y IOT respectivamente.

Capítulo IV: Aporte Teórico

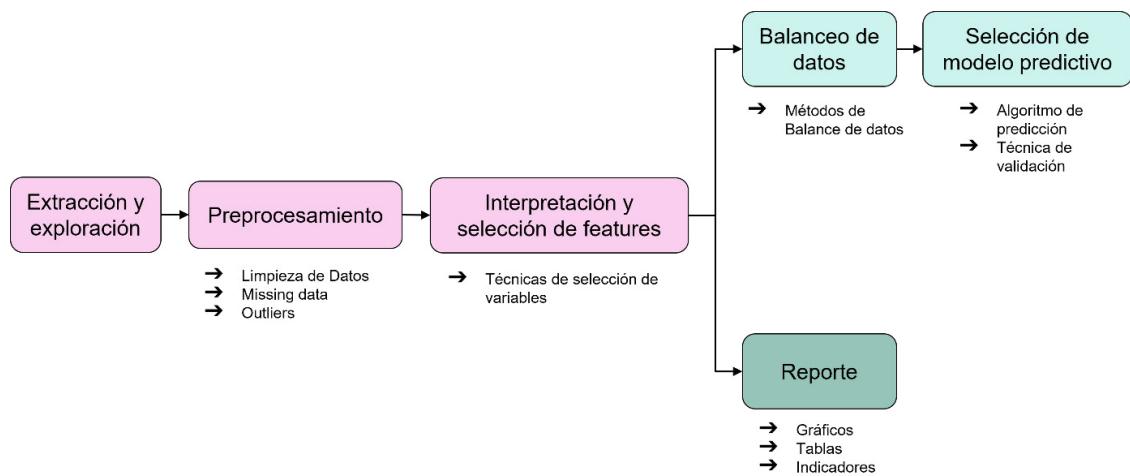
Para abordar la problemática expuesta anteriormente y en base a la investigación realizada en el estado del arte, este capítulo describe las fases utilizadas para el desarrollo de la solución, mediante un diagrama que enlaza cada una de estas.

4.1. Esquema general de la solución

En esta sección se presenta el diagrama general de la propuesta solución alineado a los conceptos de analítica descriptiva y predictiva, como se muestra en la figura 4.1.

Figura 4.1.

Esquema propuesto para el desarrollo de la solución.



Fuente: Elaboración Propia

4.1.1 Fase de extracción y exploración de datos

El proceso de extracción tiene como objetivo obtener los datos de las diferentes fuentes y/o repositorios de información brindadas por las entidades responsables de la lucha contra la violencia a la mujer, que correspondan a periodos históricos necesarios para un mejor análisis. A través de técnicas de extracción usando el lenguaje de programación de Python.

El proceso de exploración tiene como finalidad interpretar cada uno de los campos de las fuentes de datos, identificar los tipos de características (features) (categóricas y numéricas), detectar posibles valores atípicos (outliers) y/o valores perdidos presentes en los registros extraídos.

Figura 4.2.

Esquema de extracción y exploración.



Fuente: Elaboración Propia

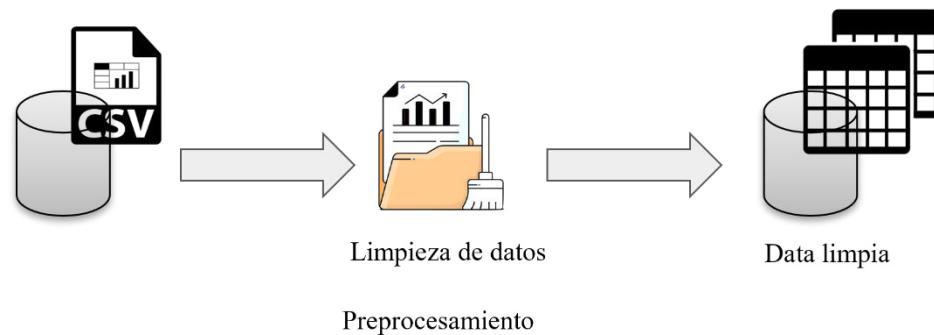
4.1.2 Fase de preprocesamiento

Esta fase se realiza las siguientes actividades:

- **Limpieza de datos:** Se realiza la eliminación de los registros erróneos e inconsistentes de acuerdo con las reglas del escenario de aplicación.
- **Imputación/Eliminación de datos perdidos:** Se aplica la imputación de valores perdidos identificados en los features, y en algunos casos la eliminación de estos.
- **Detección de valores atípicos (Outliers):** Se identifica registros que puedan alterar el análisis y el desarrollo de patrones.

Figura 4.3.

Esquema de preprocesamiento.



Fuente: Elaboración Propia

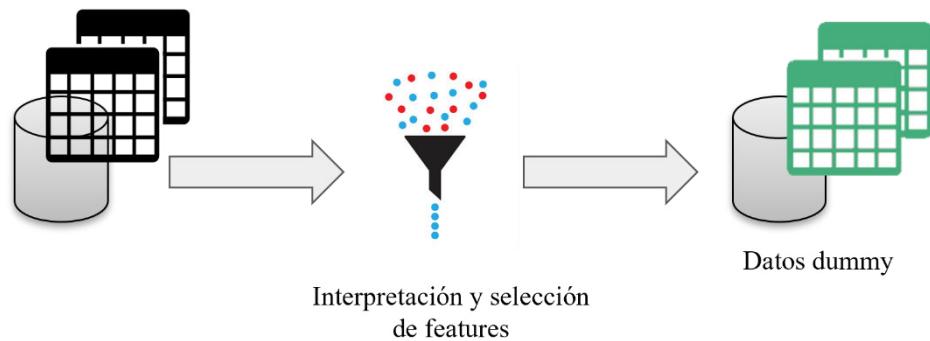
4.1.3 Fase de interpretación y selección de features

Esta fase se orienta a la interpretación de las dependencias existentes entre el variable dependiente (target) y los demás features a través del uso de técnicas como: Chi-cuadrado (cualitativa-cualitativa), pearson (cuantitativa-cuantitativa), spearman (ordinal-ordinal), biserial-puntual (cualitativa-cuantitativa).

Además, los resultados obtenidos, de acuerdo con lo descrito en el párrafo anterior, permiten seleccionar los features con una mejor correlación con el target, obteniendo mejores resultados para la identificación de patrones y predicción en los próximos pasos.

Figura 4.4.

Esquema de interpretación y selección de features.



Fuente: Elaboración Propia

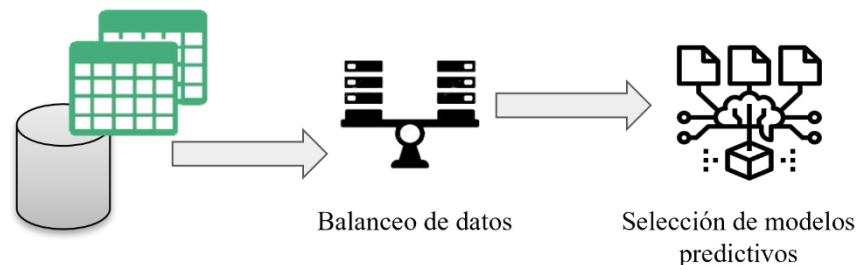
4.1.4 Fases de analítica predictiva

Esta sección se comprende las siguientes fases, como se muestra en la figura 4.5:

Figura 4.5.

Esquema de balance de datos y selección de modelos predictivos.

Analítica predictiva



Fuente: Elaboración Propia

4.1.4.1.Fase de Balance de datos

Esta fase tiene como objetivo equilibrar la cantidad de registros de la clase minoritaria (oversampling) y la clase mayoritaria (undersampling) para evitar el sobreajuste de los datos (sobreentrenamiento del algoritmo para un conjunto de datos específico), esto se logra realizando una creación y/o eliminación de registros respectivamente. Utilizando diversas técnicas, tales como: SMOTE, NearMiss, SMOTEN, como se muestra en la siguiente tabla.

Tabla 4.1

Técnica y algoritmos de balanceo de datos

Técnicas	Algoritmos
Oversampling	SMOTE
Undersampling	NearMiss
Oversampling y Undersampling	SMOTEN

Fuente: Elaboración Propia

4.1.4.2.Fase de selección de modelo predictivo

Esta fase contempla la aplicación de modelos predictivos supervisados, como: Random Forest, Decision Tree, KNN, SVM y Regresión Logística.

Estos modelos pasan por una revisión de métricas (precisión (accuracy), desviación estándar) y de estabilidad (validación cruzada), con el objetivo de seleccionar uno de ellos para la solución del problema planteado.

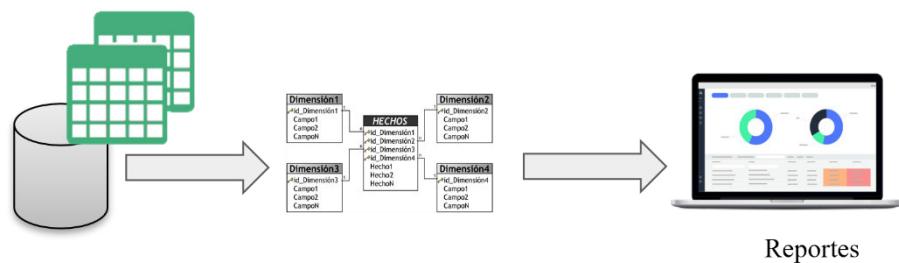
4.1.5 Fase de analítica descriptiva

Esta sección abarca la siguiente fase, como se muestra en la figura 4.6:

Figura 4.6.

Esquema de balance de datos y selección de modelos predictivos.

Analítica descriptiva



Fuente: Elaboración Propia

4.1.5.1.Fase de reporte

Esta fase busca desarrollar un dashboard que muestre indicadores relacionados ISO 37120:2018 y casos de violencia contra la mujer. Para lograr ello, se define un modelo dimensional de copo de nieve y medidas que permitan al usuario final tomar decisiones eficientemente.

Capítulo V: Aporte Práctico

El presente capítulo se centra en describir el desarrollo de la solución que contribuirá en la prevención y monitoreo de los casos de violencia contra la mujer, a través de la metodología propuesta en el capítulo anterior, así como las herramientas de hardware y software necesarias para su construcción.

5.1. Selección de tecnologías basadas en hardware y software para la solución.

En esta sección, se detalla las herramientas necesarias para la implementación de la solución, como se lista a continuación:

5.1.1. Herramientas de Hardware

A. Laptop

Esta herramienta soporta la ejecución del software necesario para el desarrollo de la solución, a continuación, se listan sus características:

- Procesador Intel i7
- 12 GB de RAM
- 1 TB Disco Duro

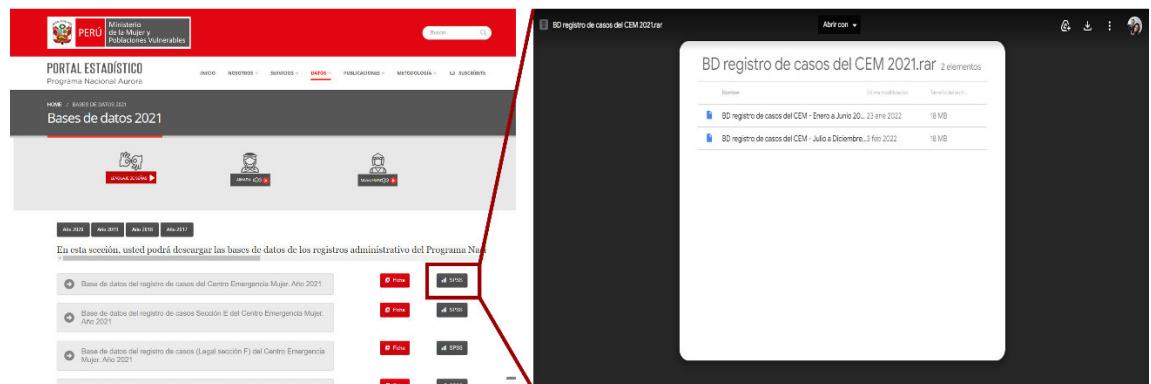
5.1.2. Herramientas de Software

A continuación, se listan las herramientas utilizadas para la construcción y despliegue de la solución:

A. Google Drive: Es un gestor de almacenamiento de archivos donde se encuentra alojadas las fuentes de datos publicadas por PEPNA (2021b). Para la presente tesis se usa la data relacionada a la “Base de datos del registro de casos del Centro Emergencia Mujer”, como se muestra en la figura 5.1

Figura 5.1.

Base de datos del registro de casos del Centro Emergencia Mujer.



Fuente: PEPNA (2021b)

Además, es usado para almacenar los códigos fuente en python, desarrollados para la solución planteada, así como, los archivos generados en la fase de preprocesamientos, balanceo de datos y modelos predictivos.

B. Python: Es el lenguaje de programación seleccionado para el desarrollo de la solución, debido a que es flexible, fácil de usar, intuitivo, código abierto y posee una variedad de librerías.

En la presente tesis se usa Google Colab como entorno de programación para la implementación de la fase de extracción y exploración, preprocesamiento, interpretación y selección de features, balanceo de datos y modelos predictivos.

C. Power BI: Es un visualizador que se usa en la presente tesis para el desarrollo de la fase de reporte (analítica descriptiva), el cual presenta los indicadores alineados a la norma ISO 37120 y una sección para el seguimiento de los casos reportados en el CEM.

5.1.3. Servicios en la nube

Dada las consideraciones de software detalladas en el párrafo anterior se ha seleccionado a Google Workspace para desarrollar las fases de extracción y exploración, preprocesamiento, interpretación y selección de features, balanceo de datos y selección del modelo predictivo, debido a que los servicios de Google Drive y Google Colab se adaptan a la data fuente y el lenguaje de programación seleccionado para la presente tesis.

5.2. Desarrollo e implementación de la metodología propuesta

En esta sección, se detalla la implementación de la solución, en base a las fases propuestas en el capítulo IV.

5.2.1. Fase de extracción y exploración de datos

En esta fase se implementa el código en Python, asociado a la extracción de los datos abiertos de los registros de casos del CEM comprendidos desde la fecha 2017 hasta 2021 de la página del Programa Nacional Aurora – MIMP. Se muestra parte del código en la figura 5.2.

Figura 5.2.

Código Python de extracción de archivos fuente.

```
[ ] # web Scrapping de la data - botón SPSS de Base de datos del registro de casos del Centro Emergencia
lista_url= list()
lista_general = list()

for i in lista_anios:
    url='https://portalestadistico.pe/bases-de-datos-' + i + '/'
    page= requests.get(url)

    soup= BeautifulSoup(page.content,'html.parser')
    if i == '2017' or i == '2018':
        link=soup.find_all('a', class_='vc_general vc_btn3 vc_btn3-size-md vc_btn3-shape-round vc_btn3-style-3d vc_btn3-icon-left vc_btn3-color-primary btn')
        print(link)
        for tag in link:
            lista_url.append(tag.get('href'))
        lista_general.append(lista_url[0])
    else:
        link=soup.find_all('a', class_='vc_general vc_btn3 vc_btn3-size-md vc_btn3-shape-rounded vc_btn3-style-custom vc_btn3-icon-left btn')
        for tag in link:
            lista_url.append(tag.get('href'))
        lista_general.append(lista_url[1])

    lista_url.clear()

[ ] # Descarga de archivos .sav de la pagina del observatorio
for z in range(len(lista_id)):
    gdd.download_file_from_google_drive(file_id=lista_id[z],
                                         dest_path='./drive/MyDrive/Código_Conversión_Python/Registro de Casos del CEM '+ lista_anios[z]+'.sav')

[ ] # Lectura de archivos descargados y conversión de .sav a .csv
for x in range(len(lista_anios)):
    df = pd.read_spss('./drive/MyDrive/Código_Conversión_Python/Registro de Casos del CEM '+ lista_anios[x]+'.sav')
    df.to_csv('./drive/MyDrive/Código_Conversión_Python/Registro de Casos del CEM '+ lista_anios[x]+'.csv')
```

Fuente: Elaboración propia

Luego de la extracción de los datos de los registros de los años mencionados, se realiza la unificación de dichos archivos, estandarizando las columnas que tienen nombres iguales y/o parecidos, quedando así 98 columnas que posteriormente entraran en la fase de procesamiento para convertirse en features; el número de columnas de los archivos de cada año se muestra en la tabla 5.1

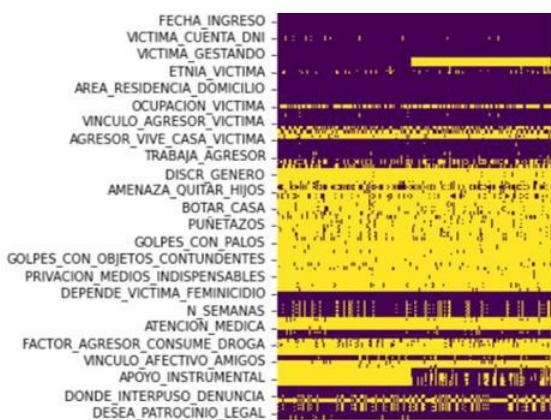
Tabla 5.1.*Número de columnas de los archivos por cada año*

Años	2017	2018	2019	2020	2021
Nº de columnas	143	158	183	158	167

Fuente: Elaboración propia

Finalmente, en la parte de extracción, se considera realizar un filtro de los casos reportados a los CEM de Lima Metropolitana y otro donde los casos cuya víctima es mujer.

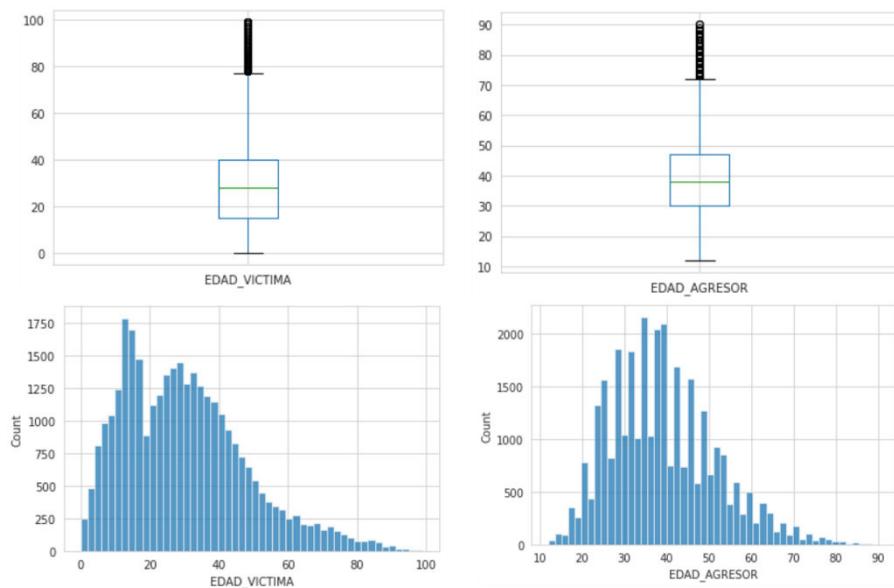
La sección asociada a la exploración de los datos busca identificar valores perdidos, outliers y la distribución de la data de los features (ver figura 5.3, figura 5.4 y figura 5.5 respectivamente)

Figura 5.3.*Identificación de valores perdidos.*

Fuente: Elaboración propia

Figura 5.4.

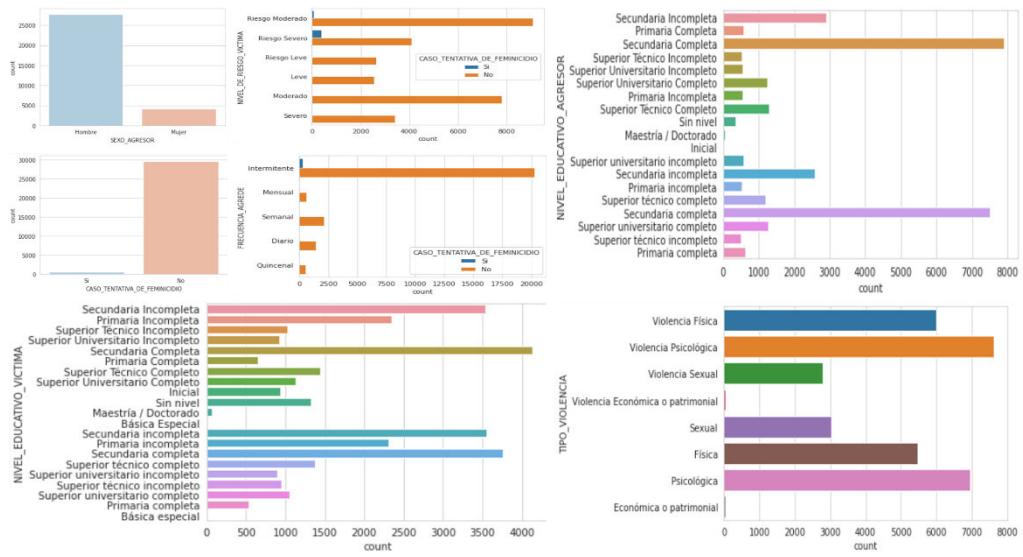
Identificación de outliers.



Fuente: Elaboración propia

Figura 5.5.

Distribución de los datos de algunos features



Fuente: Elaboración propia

5.2.2. Fase de preprocesamiento

En esta fase se implementa el código en Python, relacionado al preprocesamiento de los registros de casos del CEM estandarizados y filtrados en la fase anterior, siendo este archivo fuente para la presente sección.

Como primera validación, se realiza un filtro del feature “Sexo_Agresor” =” Hombre”, debido a que se visualiza en la fase de exploración que dichos registros representan un mayor porcentaje respecto a las mujeres agresoras.

A continuación, se muestra la tabla 5.2 que contiene el resumen del preprocesamiento aplicado a cada uno de los 98 features obtenidos del archivo input; la columna “Imputación de valores” representa la aplicación de funciones y librerías de Python para completar valores nulos o perdidos, adicionalmente se aplicó el concepto de expresar una variable cualitativa en cuantitativa (dummies) a cada una de ellas; mientras que “Eliminación de features” representa la eliminación de la columna por tener un porcentaje >30% de valores perdidos (lo cual puede afectar en el entrenamiento del modelo predictivo) o el feature es reemplazado por uno o más.

Tabla 5.2.

Preprocesamiento aplicado a los features del registro de casos del Centro Emergencia Mujer 2017-2021

Features	Imputación de valores	Eliminación de features	Columna creada o modificada
CEM			CEM
CONDICION			CONDICION
FECHA_INGRESO		X	MES
INFORMANTE	X		INFORMANTE (*)
FORMA_INGRESO		X	-
VICTIMA_PERUANA	X		
VICTIMA CUENTA_DNI		X	VICTIMA_PERUANA
VICTIMA_EXTRANJERA		X	

EDAD_VICTIMA	X	EDAD_VICTIMA
SEXO_VICTIMA	X	
VICTIMA_GESTANDO	X	VICTIMA_GESTA NDO (*)
HIJAS_VIVAS	X	
HIJOS_VIVOS	X	
LENGUA_MATERNA_VICTIM A	X	LENGUA_MATER NA_VICTIMA
ETNIA_VICTIMA	X	ETNIA_VICTIMA
DPTO_DOMICILIO	X	
PROV_DOMICILIO	X	
DIST_DOMICILIO	X	
AREA_RESIDENCIA_DOMICI LIO	X	
ESTADO_CIVIL_VICTIMA		ESTADO_CIVIL_V ICTIMA
NIVEL_EDUCATIVO_VICTIM A	X	NIVEL_EDUCATI VO_VICTIMA
TRABAJA_VICTIMA	X	TRABAJA_VICTI MA
OCUPACION_VICTIMA	X	
AGRESOR_PERUANO	X	X
AGRESOR CUENTA_DNI		X
AGRESOR_EXTRANJERO	X	X
VINCULO_AGRESOR_VICTIM A		VINCULO_AGRES OR_VICTIMA
VINCULO_PAREJA	X	X
VINCULO_FAMILIAR	X	X
SIN_VINCULO	X	X
AGRESOR_VIVE_CASA_VICT IMA	X	AGRESOR_VIVE_ CASA_VICTIMA
EDAD_AGRESOR	X	X
SEXO_AGRESOR		X
NIVEL_EDUCATIVO_AGRESO R	X	NIVEL_EDUCATI VO_AGRESOR
TRABAJA_AGRESOR	X	TRABAJA_AGRES OR
OCUPACION_AGRESOR		X
GRITOS_INSULTOS	X	GRITOS_INSULTO S
DISCR_ORIENTACION_SEXU AL	X	DISCR_ORIENTAC ION_SEXUAL
DISCR_GENERO	X	DISCR_GENERO
DISCR_IDENTIDAD_GENERO	X	DISCR_IDENTIDA D_GENERO
RECHAZO	X	RECHAZO
DESVALORIZACION_HUMIL LACION	X	DESVALORIZAC ION_HUMILLACIO N
AMENAZA_QUITAR_HIJOS	X	AMENAZA_QUIT AR_HIJOS
OTRAS_AMENAZAS		X

PROHIBE_ESTUDIAR_TRABA JAR_SALIR	X	PROHIBE_ESTUDI AR_TRABAJAR_S ALIR
ROMPE_DESTRUYE_COSAS	X	ROMPE_DESTRU YE_COSAS
BOTAR_CASA	X	BOTAR_CASA
AMENAZA_DE_MUERTE	X	AMENAZA_DE_M UERTE
OTRA_VPSI	X	
PUNTAPIES_PATADAS	X	PUNTAPIES_PATA DAS
PUÑETAZOS	X	PUÑETAZOS
BOFETADAS	X	BOFETADAS
JALONES_CABELLO	X	JALONES_CABEL LO
EMPUJONES	X	EMPUJONES
GOLPES_CON_PA LOS	X	GOLPES_CON_PA LOS
LATIGAZO	X	LATIGAZO
AHORCAMIENTO	X	AHORCAMIENTO
HERIDAS_CON_ARMAS	X	HERIDAS_CON_A RMAS
GOLPES_CON_OBJETOS_CO NTUNDENTES	X	GOLPES_CON_OB JETOS_CONTUND ENTES
NEGLIGENCIA	X	NEGLIGENCIA
OTRA_VFIS	X	
LIMITACION_REC URSOS_ECONOMI COS	X	LIMITACION_REC URSOS_ECONOMI COS
PRIVACION_MEDI OS_INDISPENSAB LES	X	PRIVACION_MEDI OS_INDISPENSAB LES
CONTROL_DE_INGRESOS	X	CONTROL_DE_IN GRESOS
VIOLACION	X	VIOLACION
ACTOS CONTRA EL PUDOR	X	ACTOS CONTRA_ EL_PUDOR
DEPEND_VICTIMA_FEMINI CIDIO	X	DEPEND_VICTI MA_FEMINICIDIO
N_AÑOS	X	
N_MESES	X	
N_SEMANAS	X	
PRIMERA_VEZ AGREDE	X	X
FRECUENCIA AGREDE	X	X
TRATAMIENTO_PSICOLOGI CO	X	
TRATAMIENTO_PSIQUIATRI CO	X	
ATENCION_MEDICA	X	
NINGUN TRATAMIENTO	X	
CONTINUA_RECIBIENDO TR ATAMIENTO	X	
FACTOR_AGRESOR_CONSU MO_ALCOHOL	X	
FACTOR_AGRESOR_CONSU ME_DROGA	X	

FACTOR_VICTIMA_ABUSO_C ONSUMO_ALCOHOL	X	X	
FACTOR_VICTIMA_CONSUM E_DROGAS	X	X	ADICCION_VICTI MA
VINCULO_AFECTIVO_FAMILY		X	
VINCULO_AFECTIVO_AMIG OS		X	VINCULO_SOPOR TE_VICTIMA
VINCULO_AFECTIVO_NINGU NO	X	X	
APOYO_EMOCIONAL		X	
APOYO_COGNITIVO		X	
APOYO_INSTRUMENTAL		X	
APOYO_MATERIAL		X	
NIVEL_DE_RIESGO_VICTIMA			NIVEL_DE_RIESG O_VICTIMA
INTERPUSO_DENUNCIA	X	X	
DONDE_INTERPUSO_DENUN CIA	X	X	INTERPUSO_LA_ DENUNCIA
DESEA_DENUNCIAR		X	
CUENTA_MEDIDAS_PROTEC CION	X		CUENTA_MEDID AS_PROTECCION
CUENTA_MEDIDAS_CAUTEL ARES	X		CUENTA_MEDID AS_CAUTELARES
DESEA_PATROCINIO_LEGAL			DESEA_PATROCI NIO_LEGAL
CASO_TENTATIVA_DE_FEMI NICIDIO	X		CASO_TENTATIV A_DE_FEMINICID IO
SITUACION_AGRESOR		X	-
TIPO_VIOLENCIA		X	TIPO_VIOLENCIA

Nota: (*) Solo se aplicó el concepto de variables Dummies.

Fuente: Elaboración propia

Luego de realizar el preprocesamiento descrito en el párrafo anterior, se obtiene un total de 59 features, a los cuales se les debe de aplicar el concepto de variables dummies utilizando principalmente la librería math de python. Siendo este un paso a realizar a los siguientes features que aún están expresados con valores cualitativos, como se muestra en la tabla 5.3.

Tabla 5.3.

Concepto de variables dummies aplicado a los features faltantes del registro de casos del Centro Emergencia Mujer 2017-2021

Columna modificada
CEM
CONDICION
LENGUA_MATERNA_VICTIMA
ETNIA_VICTIMA
ESTADO_CIVIL_VICTIMA
NIVEL_EDUCATIVO_VICTIMA
VINCULO_AGRESOR_VICTIMA
VINCULO_VIC AGRE
NIVEL_EDUCATIVO_AGRESOR
FRECUENCIA_AGRESION
NIVEL_DE_RIESGO_VICTIMA
TIPO_VIOLENCIA

Fuente: Elaboración propia

Finalmente, en esta fase se obtiene 2 archivos (25,380 registros cada uno), uno de ellos (relacionado a la tabla 5.2) servirá como input para la fase de interpretación y selección de features relevantes para el modelo predictivo y el otro (relacionado a la tabla 5.3) servirá para la fase de analítica predictiva.

5.2.3. Fase de interpretación y selección de features

En esta fase se evalúa la relación de dependencia entre el target (“CASO_TENTATIVA_DE_FEMINICIDIO”) con los demás features. Para ello se eligió el método de coeficiente de correlación chi-cuadrado debido a que nuestras variables dependientes e independiente son variables cualitativas.

A continuación, la tabla 5.4 muestra la relación de significancia de cada feature con el target a través de la aceptación o el rechazo de la hipótesis nula (la variable x e y

son independientes por ende la dependencia entre las variables no es significativa), siendo ésta representada por la columna “Hipótesis chi2”. Mientras que la columna “p_values” representa el grado de significancia entre las variables; de acuerdo con el método de chi cuadrado $p \leq 0.05$ es el valor más adecuado para rechazar la hipótesis nula, concluyendo así que los features que cumplen dicha característica tienen una adecuada significancia con respecto al target definido.

Tabla 5.4.

Concepto de variables dummies aplicado a los features faltantes del registro de casos del Centro Emergencia Mujer 2017-2021

Nombre de Features	Hipótesis Chi2	p_values
HERIDAS_CON_ARMAS	Reject Null Hypothesis	0
AHORCAMIENTO	Reject Null Hypothesis	0
NIVEL_DE_RIESGO_VICTIMA	Reject Null Hypothesis	0
AMENAZA_DE_MUERTE	Reject Null Hypothesis	0
TIPO_VIOLENCIA	Reject Null Hypothesis	0
CEM	Reject Null Hypothesis	0
ETNIA_VICTIMA	Reject Null Hypothesis	0
PUÑETAZOS	Reject Null Hypothesis	0
PUNTAPIES_PATADAS	Reject Null Hypothesis	0
JALONES_CABELLO	Reject Null Hypothesis	0
GOLPES_CON_OBJETOS_CONTUNDENTES	Reject Null Hypothesis	0
EMPUJONES	Reject Null Hypothesis	0
VINCULO_AGRESOR_VICTIMA	Reject Null Hypothesis	0
VINCULO_VIC AGRE	Reject Null Hypothesis	0
RECIBE_TRATAMIENTO	Reject Null Hypothesis	0
BOFETADAS	Reject Null Hypothesis	0
NIVEL_EDUCATIVO_VICTIMA	Reject Null Hypothesis	0
ROMPE_DESTRUYE_COSAS	Reject Null Hypothesis	0
FRECUENCIA_AGRECION	Reject Null Hypothesis	0
LENGUA_MATERNA_VICTIMA	Reject Null Hypothesis	0
ADICCION_AGRESOR	Reject Null Hypothesis	0

INFORMANTE	Reject Null Hypothesis	0
PROHIBE_ESTUDIAR_TRABAJAR_SALIR	Reject Null Hypothesis	0
INTERPUSO_LA_DENUNCIA	Reject Null Hypothesis	0
GRITOS_INSULTOS	Reject Null Hypothesis	0
ACTOS_CONTRA_EL_PUDOR	Reject Null Hypothesis	0.000015
CONDICION	Reject Null Hypothesis	0.00002
TRABAJA_VICTIMA	Reject Null Hypothesis	0.000022
DISCR_GENERO	Reject Null Hypothesis	0.000081
GOLPES_CON_PALOS	Reject Null Hypothesis	0.000096
ADICCION_VICTIMA	Reject Null Hypothesis	0.000215
DEPENDE_VICTIMA_FEMINICIDIO	Reject Null Hypothesis	0.000223
VICTIMA_PERUANA	Reject Null Hypothesis	0.001013
MES	Reject Null Hypothesis	0.001512
AGRESOR_NACION	Reject Null Hypothesis	0.001942
NEGLIGENCIA	Reject Null Hypothesis	0.002615
DISCR_IDENTIDAD_GENERO	Reject Null Hypothesis	0.008383
TRABAJA_AGRESOR	Reject Null Hypothesis	0.008862
NIVEL_EDUCATIVO_AGRESOR	Reject Null Hypothesis	0.013541
AGRESOR_VIVE_CASA_VICTIMA	Reject Null Hypothesis	0.038759
VIOLACION	Acept Null Hypothesis	0.078901
DESEA_PATROCINIO_LEGAL	Acept Null Hypothesis	0.121116
DISCR_ORIENTACION_SEXUAL	Acept Null Hypothesis	0.138477
ESTADO_CIVIL_VICTIMA	Acept Null Hypothesis	0.335876
BOTAR_CASA	Acept Null Hypothesis	0.398603
LATIGAZO	Acept Null Hypothesis	0.412845
DESVALORIZACION_HUMILLACION	Acept Null Hypothesis	0.53506
VICTIMA_GESTANDO	Acept Null Hypothesis	0.610873
CUENTA_MEDIDAS_PROTECCION	Acept Null Hypothesis	0.612321
AMENAZA_QUITAR_HIJOS	Acept Null Hypothesis	0.624438
PRIVACION_MEDIOS_INDISPENSABLES	Acept Null Hypothesis	0.736081
LIMITACION_RECURSOS_ECONOMICOS	Acept Null Hypothesis	0.802157
CONTROL_DE_INGRESOS	Acept Null Hypothesis	0.823536
CUENTA_MEDIDAS_CAUTELARES	Acept Null Hypothesis	0.918413
VINCULO_SOPORTE_VICTIMA	Acept Null Hypothesis	0.924335
RECHAZO	Acept Null Hypothesis	0.999716

Fuente: Elaboración propia

5.2.4. Fases de analítica predictiva

5.2.4.1. Fase de Balanceo de datos

En esta fase se evalúa la efectividad de la aplicación de las técnicas de balanceo mencionadas en el capítulo IV junto con algoritmos de predicción (KNN, Regresión logística, SVM, Random Forest, Árbol de decisión), con la finalidad de seleccionar la técnica de balanceo adecuada para el entrenamiento del modelo predictivo, para ello se define 2 consideraciones: Los algoritmos de predicción deben ser probados con los parámetros por defecto y que el accuracy sea mayor 95%.

En la tabla 5.5 se presenta los resultados de cada técnica de balanceo por cada algoritmo de predicción.

Tabla 5.5.

Resultado de la evaluación de los algoritmos de predicción con las técnicas de balanceo (accuracy)

Algoritmo	SMOTE	Nearmiss	SMOTEEEN
KNN	95	80	98
Regresión Logística	92	89	91
SVM	1	1	1
Random Forest	1	1	1
Árbol de decisión	1	1	1

Nota: El archivo input para la evaluación es el obtenido en la fase de preprocesamiento. Fuente: Elaboración propia

De acuerdo con los resultados detallados en la tabla 5.5, los algoritmos que alcanzaron una exactitud de 1 con las tres técnicas de balanceo no están siendo tomadas en cuenta debido al sobreajuste, ya que como consideración se tomaron valores por

defecto para su entrenamiento, concluyendo así que la mejor técnica de balanceo a aplicar en relación con la problemática presentada junto a sus datos es SMOTEEN.

5.2.4.2. Fase de selección de modelo predictivo

En esta fase se implementa el código en Python usando librerías propias del lenguaje, entre las más resaltantes scikit learn, con el objetivo de realizar el entrenamiento de los modelos predictivos aplicando 5 algoritmos (KNN, Regresión logística, SVM, Random Forest, Árbol de decisión), selección de uno de ellos de acuerdo con métricas que evidencian la precisión del modelo, y la aplicación de validación cruzada para medir su estabilidad.

A continuación, se lista la evaluación del modelo predictivo de cada algoritmo mencionado. En cada uno de ellos se realiza el balanceo de los datos con la técnica seleccionada (SMOTEEN), además la división de registros en data de entrenamiento (train) y de prueba (test) tiene una proporción de 70-30 respectivamente.

1. KNN

Para la implementación de este algoritmo se invoca la biblioteca sklearn.neighbors.KNeighborsClassifier donde, para la presente tesis, se usa como parámetro de evaluación del modelo **n_neighbors**, como se muestra a continuación.

```
from sklearn.neighbors import KNeighborsClassifier
model = KNeighborsClassifier(n_neighbors=5)
model.fit(x_train, y_train)

y_predict = model.predict(x_test)
x_predict = model.predict(x_train)
```

El valor del parámetro definido que brinde un mejor resultado a nivel de accuracy, es seleccionado para el entrenamiento del modelo, como se presenta en la tabla 5.6.

Tabla 5.6.

Resultado de la evaluación del parámetro n_neighbors en KNeighborsClassifier

Nº	n_neighbors	Accuracy train	Accuracy test
1	5	0.98	0.97
2	6	0.98	0.97
3	7	0.97	0.96
4	8	0.97	0.96
5	9	0.96	0.95
6	10	0.96	0.95
7	11	0.95	0.94
8	12	0.95	0.95
9	13	0.94	0.93
10	14	0.95	0.94
11	15	94	93
12	50	0.88	0.88
13	100	0.85	0.85
14	200	0.82	0.81
15	300	0.80	0.80
16	400	0.79	0.78
17	500	0.78	0.78

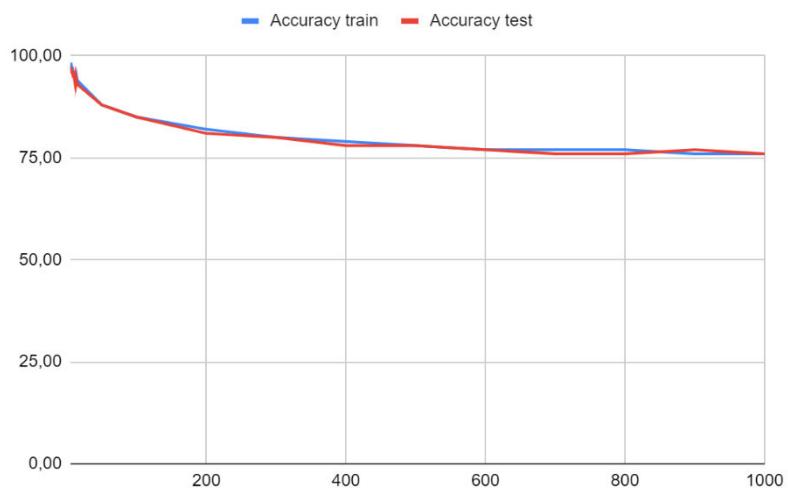
18	600	0.77	0.77
19	700	0.77	0.76
20	800	0.77	0.76
21	900	0.76	0.77
22	1000	0.76	0.76

Fuente: Elaboración propia

Además, en la figura 5.6 se muestra el comportamiento del parámetro n_neighbors de acuerdo con el accuracy del train y test, donde se puede visualizar que a mayor valor del parámetro menor accuracy.

Figura 5.6.

Comportamiento del valor del parámetro n_neighbors de acuerdo con el accuracy obtenido.



Fuente: Elaboración propia

De acuerdo con la tabla 5.6, los mejores valores para el parámetro n_neighbors en relación con el accuracy son n_neighbors=5, n_neighbors=6, n_neighbors=7 y n_neighbors=8.

Para la selección de uno de ellos se realiza la validación cruzada con el objetivo de validar la estabilidad del modelo, siendo $n_{neighbors}=8$ la que presenta menor desviación estándar entre los valores obtenidos, mostrados a continuación en la figura 5.7.

Figura 5.7.

Validación cruzada y desviación estándar del modelo predictivo con KNN.

Validación cruzada: n_neighbors=5 $\sigma = 0.003427323$

```
[21] print("Metricas cross_validation", scores)
```

Metricas cross_validation [0.96364757 0.96210721 0.96672828 0.96118299 0.95841035 0.95902649
0.96548536 0.96024653 0.96579353 0.96271186]

Validación cruzada: n_neighbors=6 $\sigma = 0.003399904$

```
[61] print("Metricas cross_validation", scores)

Metricas cross_validation [0.96303142 0.96241528 0.9642527 0.96733436 0.95716487 0.9614792
 0.96024653 0.96240837 0.96918336 0.96209553]
```

Validación cruzada: n_neighbors=7 $\sigma = 0.003593019$

```
[84] print("Metricas cross_validation", scores)

Metricas cross_validation [0.96580407 0.96056685 0.97042514 0.96272335 0.95995071 0.96178737
 0.95932203 0.95808937 0.9633282 0.96117103]
```

Validación cruzada: n_neighbors=8 $\sigma = 0.003111392$

```
[107] print("Metricas cross_validation", scores)
Metricas cross_validation [0.96520936 0.96305419 0.96644089 0.95935961 0.96366995 0.96735448
0.96550662 0.96365876 0.95873114 0.95934709]
```

Fuente: Elaboración propia

2. Regresión logística

Para la implementación de este algoritmo se invoca la librería `sklearn.linear_model.LogisticRegression` donde, para la presente tesis, se usan como parámetros de evaluación del modelo **max_iter** y **class_weight**, como se muestra a continuación.

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression(max_iter = 150,
class_weight='balanced')
model.fit(x_train, y_train)

y_predict = model.predict(x_test)
y_predict_train = model.predict(x_train)
```

Al realizar la evaluación del modelo con los valores de los parámetros definidos no se obtuvo un accuracy mayor a 95%, como se presenta en la tabla 5.7.

Tabla 5.7.

Resultado de la evaluación de los parámetros max_iter y class_weight en LogisticRegression.

Nº	max_iter	class_weight	Accuracy train	Accuracy test
1	2000	balanced	91	92
2	1500	balanced	92	91
3	1400	balanced	92	91
4	1300	balanced	91	91
5	1200	balanced	92	91
6	1100	balanced	91	91
7	1000	balanced	92	91
8	900	balanced	92	92
9	800	balanced	91	91

10	700	balanced	91	92
11	600	balanced	92	91
12	500	balanced	91	91
13	400	balanced	91	91
14	300	balanced	91	91
15	250	balanced	91	91
16	200	balanced	91	91
17	150	balanced	91	91
18	100	balanced	90	90
19	50	balanced	89	89
20	25	balanced	87	87
21	2	balanced	70	71

Fuente: Elaboración propia

De acuerdo con la tabla 5.7, y lo especificado en el párrafo anterior, este algoritmo de predicción no cumple el mínimo valor de accuracy definido para su elección en la presente tesis.

3. SVM

Para la implementación de este algoritmo se invoca la librería sklearn.svm donde, para la presente tesis, se usan como parámetros de evaluación del modelo a **gamma** y **C**, como se muestra a continuación.

```
from sklearn.svm import SVC
svc = SVC(kernel = 'rbf', gamma = 1, C = 0.1,probability=True)

scaler = MinMaxScaler()
scaler.fit(x_train)
X_train_scaled = scaler.transform(x_train)
```

```

x_test_scaled = scaler.transform(x_test)

svc.fit(X_train_scaled, y_train)

y_train_pred_svm = svc.predict(X_train_scaled)
y_test_pred_svm = svc.predict(X_test_scaled)

```

Al realizar la evaluación del modelo con los valores de los parámetros definidos se obtuvo un registro con un accuracy mayor a 95%, como se presenta en la tabla 5.8.

Tabla 5.8.

Resultado de la evaluación de los parámetros gamma y C en SVM.

Nº	kernel	gamma	c	probability	Accuracy train	Accuracy test
1	rbf	1	1	True	1	0,98
2	rbf	1	10	True	1	0,99
3	rbf	1	100	True	1	0,99
4	rbf	1	1000	True	1	0,98
5	rbf	10	1	True	1	0,98
6	rbf	100	1	True	1	0,95
7	rbf	1	0,1	True	0,97	0,97
8	rbf	0,1	0,1	True	0,93	0,93

Fuente: Elaboración propia

De acuerdo con la tabla 5.8, los mejores valores para los parámetros gamma y C en relación con el accuracy son 1 y 0,1 respectivamente.

Para poder validar que los valores de los parámetros elegidos kernel =rbf, gamma=1, c=0,1 y probability=True son correctos para el conjunto de datos, se procede a realizar una validación del modelo frente a casos nuevos y

diferentes a la data de entrenamiento y prueba, dichos resultados se muestran en la figura 5.8

Figura 5.8.

Validación de modelo elegido SVM frente a casos nuevo.

```
mi_modelito_prod.predict(df)
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:444: UserWarning: X has feature names, but SVC was fitted without feature names
  "X has feature names, but %s was fitted without"
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

Fuente: Elaboración propia

De acuerdo con los resultados de figura 5.8, se puede concluir que, si bien el modelo presenta un accuracy del 97%, al utilizarlo frente a nuevos casos no logra identificar correctamente si el caso es una tentativa de feminicidio o no, lo que nos indica de que existe un sobreajuste del modelo para con el conjunto de datos de entrenamiento.

4. Random Forest

Para la implementación de este algoritmo se invoca la librería `sklearn.ensemble.RandomForestClassifier` donde, para la presente tesis, se usa como parámetros de evaluación del modelo **n_estimators**, **max_depth** y **max_samples** como se muestra a continuación.

```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators=500,
max_depth = 13, max_samples=0.2,
random_state=0)
rf.fit(x_train, y_train)

y_train_pred_rf = rf.predict(x_train)
y_test_pred_rf = rf.predict(x_test)
```

El valor del parámetro definido que brinde un mejor resultado a nivel de accuracy, es seleccionado para el entrenamiento del modelo, como se presenta en la tabla 5.9.

Tabla 5.9.

Resultado de la evaluación de los parámetros n_estimators (A), max_depth (B), max_samples (C) y random_state en RandomForestClassifier

Nº	A	B	C	Accuracy train	Accuracy test
1	500	30	0,7	1	0,99
2	500	30	0,6	1	0,99
3	500	30	0,5	1	0,99
4	500	30	0,4	1	0,99
5	500	30	0,3	1	0,99
6	500	30	0,2	0,99	0,98
7	400	30	0,2	0,99	0,98
8	300	30	0,2	0,99	0,98
9	200	30	0,2	0,99	0,98
10	200	30	0,2	0,99	0,98
11	500	25	0,2	0,99	0,98
12	500	20	0,2	0,99	0,98
13	500	19	0,2	0,99	0,98
14	500	18	0,2	0,99	0,98
15	500	17	0,2	0,98	0,98
16	500	16	0,2	0,98	0,98
17	500	15	0,2	0,98	0,97

18	500	14	0,2	0,98	0,97
19	500	13	0,2	0,97	0,97
20	500	12	0,2	0,97	0,97
21	500	11	0,2	0,96	0,96
22	500	10	0,2	0,96	0,95

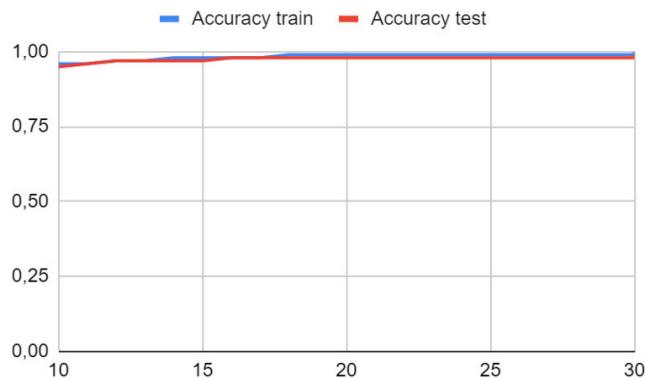
Fuente: Elaboración propia

Además, en la figura 5.9 se muestra el comportamiento de los parámetros Max_depth y Max_Samples de acuerdo con el accuracy del train y test, donde se puede visualizar que a mayor valor del parámetro tiende a ser un sobreajuste.

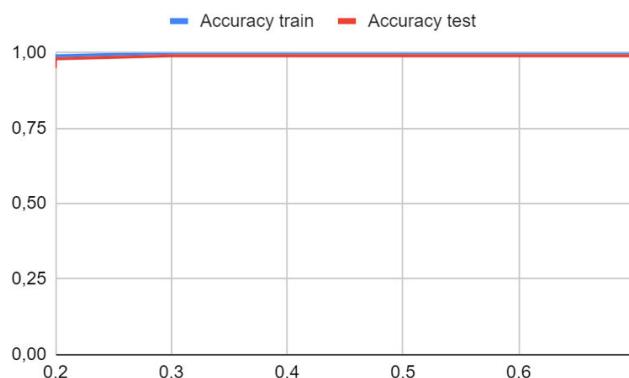
Figura 5.9.

Comportamiento del valor del parámetro Max_depth y Max_Samples de acuerdo con el accuracy obtenido.

Max_depth



Max_Samples



Fuente: Elaboración propia

De acuerdo con la tabla 5.9, los mejores valores para los parámetros n_estimators, max_depth y max_depth en relación con el accuracy son los mostrados en la tabla 5.10.

Tabla 5.10.

Resultado de los mejores valores de los parámetros n_estimators (A), max_depth (B), max_samples (C) y random_state en relación con el accuracy

Nº	A	B	C	Accuracy train	Accuracy test
1	500	17	0,2	0,98	0,98
2	500	16	0,2	0,98	0,98
3	500	15	0,2	0,98	0,97
4	500	14	0,2	0,98	0,97
5	500	13	0,2	0,97	0,97
6	500	12	0,2	0,97	0,97

Fuente: Elaboración propia

Para la selección de uno de ellos se realiza la validación cruzada con el objetivo de validar la estabilidad del modelo, siendo n_estimators=500, max_depth = 13, max_samples=0.2 los que presentan la menor desviación estándar entre los valores obtenidos, mostrados a continuación en la figura 5.10.

Figura 5.10.

Validación cruzada y desviación estándar del modelo predictivo con Random Forest.

```
n_estimators=500, max_depth = 17, max_samples=0.2          σ = 0.002074849
print("Metricas cross_validation", scores)

Metricas cross_validation [0.97264064 0.9747925  0.97632954 0.97602213 0.97694436 0.98032585
 0.97448509 0.97632954 0.97755918 0.9747925 ]

n_estimators=500, max_depth = 16, max_samples=0.2          σ = 0.001839035
print("Metricas cross_validation", scores)

Metricas cross_validation [0.97171841 0.97356287 0.97417768 0.97602213 0.97448509 0.97786658
 0.97202582 0.97417768 0.97540732 0.97325546]

n_estimators=500, max_depth = 15, max_samples=0.2          σ = 0.001339571
print("Metricas cross_validation", scores)

Metricas cross_validation [0.97141101 0.97233323 0.97264064 0.97202582 0.97387027 0.97509991
 0.97294885 0.97325546 0.97233323 0.97018137]

n_estimators=500, max_depth = 14, max_samples=0.2          σ = 0.001521251
print("Metricas cross_validation", scores)

Metricas cross_validation [0.96526283 0.96895174 0.96710729 0.96864433 0.96618506 0.96987396
 0.96557024 0.96741469 0.96710729 0.96618506]

n_estimators=500, max_depth = 13, max_samples=0.2          σ = 0.00117863
print("Metricas cross_validation", scores)

Metricas cross_validation [0.96802951 0.97018137 0.97079619 0.97079619 0.97079619 0.97233323
 0.97079619 0.97171841 0.9711036 0.96956655]

n_estimators=500, max_depth = 12, max_samples=0.2          σ = 0.002049646
print("Metricas cross_validation", scores)

Metricas cross_validation [0.9603443  0.96218875 0.96188134 0.96526283 0.96065171 0.96679988
 0.96249616 0.96126652 0.96157393 0.96341838]
```

Fuente: Elaboración propia

5. Árbol de decisión

Para la implementación de este algoritmo se invoca la librería `sklearn.tree.DecisionTreeClassifier` donde, para la presente tesis, se usa como parámetros de evaluación del modelo **criterion**, **max_depth** y **random_state** como se muestra a continuación.

```
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier(criterion='gini',
max_depth=15, random_state=0)

dt.fit(x_train, y_train)

y_train_pred_tree = dt.predict(x_train)
y_test_pred_tree = dt.predict(x_test)
```

El valor del parámetro definido que brinde un mejor resultado a nivel de accuracy, es seleccionado para el entrenamiento del modelo, como se presenta en la tabla 5.11.

Tabla 5.11.

Resultado de la evaluación de los parámetros criterion (A), max_depth (B), max_samples (C) y random_state en DecisionTreeClassifier

Nº	A	B	C	Accuracy train	Accuracy test
1	entropy	24	0	1	0,98
2	entropy	23	0	1	0,97
3	entropy	22	0	1	0,97
4	entropy	17	0	0,99	0,97
5	entropy	16	0	0,99	0,97
6	entropy	15	0	0,99	0,96
7	entropy	14	0	0,98	0,96
8	entropy	13	0	0,97	0,95

9	entropy	12	0	0,97	0,95
10	entropy	11	0	0,95	0,94
11	entropy	10	0	0,94	0,93
12	gini	20	0	0,99	0,97
13	gini	19	0	0,99	0,97
14	gini	18	0	0,98	0,96
15	gini	17	0	0,98	0,96
16	gini	16	0	0,98	0,96
17	gini	15	0	0,98	0,96
18	gini	14	0	0,98	0,96
19	gini	13	0	0,97	0,95
20	gini	12	0	0,97	0,95
21	gini	11	0	0,96	0,94
22	gini	10	0	0,95	0,94

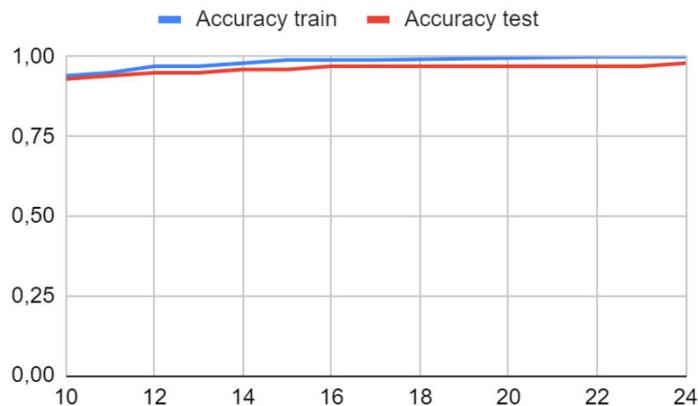
| Fuente: Elaboración propia |

Además, en la figura 5.11 se muestra el comportamiento de los parámetros criterion y max_depth de acuerdo con el accuracy del train y test, donde se puede visualizar que a mayor valor del parámetro tiende a ser un sobreajuste.

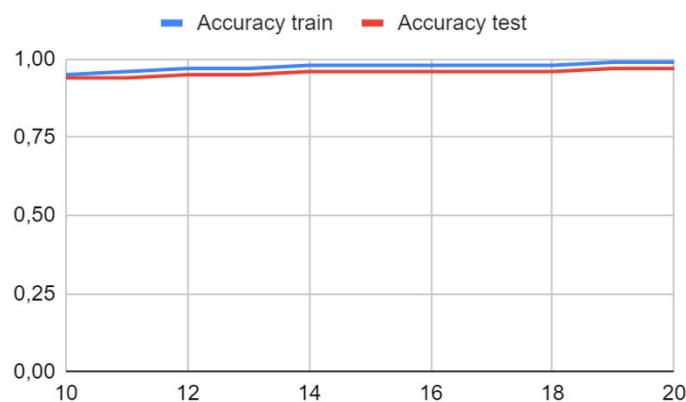
Figura 5.11.

Comportamiento del valor del parámetro criterion y max_depth de acuerdo con el accuracy obtenido.

Max_Depth y Criterion = Entropy



Max_Depth y Criterion = Gini



Fuente: Elaboración propia

De acuerdo con la tabla 5.11, los mejores valores para los parámetros criterion, max_depth y random_state en relación con el accuracy son los mostrados en la tabla 5.12.

Tabla 5.12.

Resultado de los mejores valores de los parámetros criterion (A), max_depth (B), max_samples (C) y random_state) en relación con el accuracy

Nº	A	B	C	Accuracy train	Accuracy test
	14	Entrop	0,98	0,96	14
1		y			
2	18	Gini	0,98	0,96	18
3	17	Gini	0,98	0,96	17
4	16	Gini	0,98	0,96	16
5	15	Gini	0,98	0,96	15
6	14	Gini	0,98	0,96	14

Fuente: Elaboración propia

Para la selección de uno de ellos se realiza la validación cruzada con el objetivo de validar la estabilidad del modelo, siendo criterion='gini', max_depth=15, random_state=0 los que presentan la menor desviación estándar entre los valores obtenidos, mostrados a continuación en la figura 5.12.

Figura 5.12.

Validación cruzada y desviación estándar del modelo predictivo con Decision Tree.

```
criterion=entropy, max_depth=14, random_state=0 σ = 0.004188572
print("Metricas cross_validation", scores)
Metricas cross_validation [0.9585126 0.95604058 0.96864433 0.96249616 0.96526283 0.96557024
0.96065171 0.96003689 0.9566554 0.95819244]

criterion=gini, max_depth=18, random_state=0 σ = 0.003472089
print("Metricas cross_validation", scores)
Metricas cross_validation [0.96127843 0.95849985 0.96741469 0.96833692 0.96157393 0.96802951
0.96526283 0.96741469 0.96372579 0.96157393]

criterion=gini, max_depth=17, random_state=0 σ = 0.00336362
print("Metricas cross_validation", scores)
Metricas cross_validation [0.96004917 0.95757762 0.96741469 0.96495543 0.95972948 0.96618506
0.96157393 0.96526283 0.96003689 0.96003689]

criterion=gini, max_depth=16, random_state=0 σ = 0.002739947
print("Metricas cross_validation", scores)
Metricas cross_validation [0.95789797 0.95849985 0.96495543 0.9640332 0.9603443 0.96618506
0.96372579 0.96218875 0.96126652 0.96065171]

criterion=gini, max_depth=15, random_state=0 σ = 0.001963657
print("Metricas cross_validation", scores)
Metricas cross_validation [0.95728334 0.95788503 0.96280357 0.96188134 0.95819244 0.96218875
0.95911466 0.96157393 0.95972948 0.95972948]

criterion=gini, max_depth=14, random_state=0 σ = 0.002866333
print("Metricas cross_validation", scores)
Metricas cross_validation [0.95697603 0.95511835 0.96249616 0.96095911 0.95542576 0.96065171
0.9566554 0.96249616 0.95634799 0.95819244]
```

Fuente: Elaboración propia

De acuerdo con el análisis predictivo detallado para cada algoritmo, se concluye que Random Forest es el algoritmo elegido para la presente tesis debido a que obtiene un accuracy de 0.97 para la data de prueba como se muestra en la tabla 5.13.

Tabla 5.13.

Resumen de resultados de los algoritmos de predicción de acuerdo con su respectivo accuracy

Algoritmo	Accuracy Train	Accuracy Test
KNN	0.97	0.96
Regresión Logística	0.92	0.92
Random Forest	0.97	0.97
Árbol de decisión	0.98	0.96

Nota: SVM no se considera en el cuadro resumen debido a que se verificó que para la presente tesis es susceptible a los datos. Fuente: Elaboración propia

Para validar la efectividad de acuerdo con la precisión obtenida por el modelo predictivo de Random Forest, se procede a probar nuevos registros de casos para conocer si el caso reportado es una tentativa de feminicidio o no. Dicha evaluación será revisada en el apartado de “Validación de modelo predictivo”.

5.2.5. Fase de analítica descriptiva

La fase descriptiva de la metodología propuesta se basa en la Norma ISO 37120-2018, para ello se han seleccionado indicadores de tres ámbitos correspondientes a (ISO, 2018) que tengan relación a las mujeres o violencia contra ellas, alineado a ellos en la presente tesis se elaboran indicadores para los casos de violencia contra la mujer reportados a los CEM de Lima-Perú, lo cuales son detallados en la tabla 5.14.

Tabla 5.14.*Indicadores propuestos en base a la Norma ISO 37120-2018*

Indicador Propuesto	Ámbitos / Indicadores ISO 37120 - 2018				Condiciones Sociales y población	Seguridad
	Educación		6.6 Número de títulos de educación superior por cada 100 000 habitantes (indicador de apoyo)	13.4.4 Porcentaje de población que son nuevos inmigrantes (indicador de profile)		
6.1 Porcentaje de población femenina en edad escolar matriculada en escuelas (indicador core)	6.2 Porcentaje de estudiantes que finalizan la educación primaria: tasa de supervivencia (indicador core)	6.3 Porcentaje de estudiantes que completan la educación secundaria: tasa de supervivencia (indicador core)	6.6 Número de títulos de educación superior por cada 100 000 habitantes (indicador de apoyo)	13.4.4 Porcentaje de población que son nuevos inmigrantes (indicador de profile)	15.10 Número de delitos violentos contra la mujer por cada 100 000 habitantes (indicador de apoyo)	
Número y porcentaje de casos de violencia contra la mujer reportados al CEM según el tipo de violencia recibida						✓
Top 10 de los CEM según el número de casos de violencia contra la mujer atendidos						✓
Número de casos nuevos reportados al CEM, feminicidios y tentativas de feminicidios por año						✓

Número de casos de violencia contra la mujer reportados al CEM según nacionalidad				✓
Número de casos de violencia contra la mujer reportados al CEM por grado de instrucción que recibieron algún tipo de violencia	✓	✓	✓	✓
Número de casos de violencia contra la mujer reportados al CEM respecto a su empleabilidad		✓	✓	✓
Número de casos de violencia contra la mujer reportados al CEM que llegan con alguna medida de protección				✓
Número de casos de violencia contra la mujer reportados al CEM según el nivel de riesgo				✓

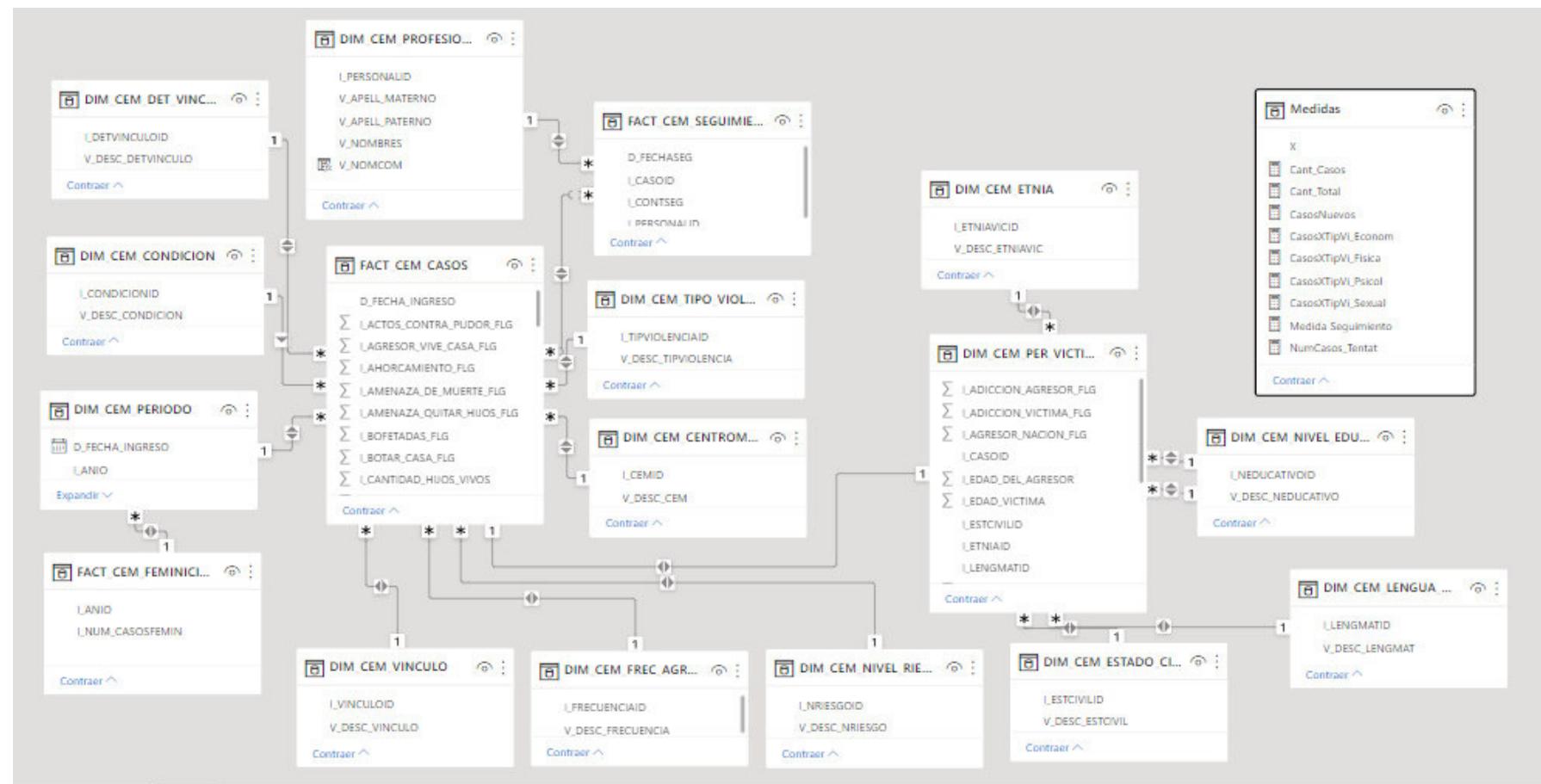
Fuente: Elaboración propia

5.2.5.1.Fase de reporte

Con el apoyo de la herramienta Power BI y la data obtenida de la fase de preprocesamiento, se define un modelo de datos copo de nieve para los registros de los casos de violencia contra la mujer reportados a los CEM de Lima Metropolitana 2017-2021, como se visualiza en la figura 5.13.

Figura 5.13.

Modelo de datos copo de nieve

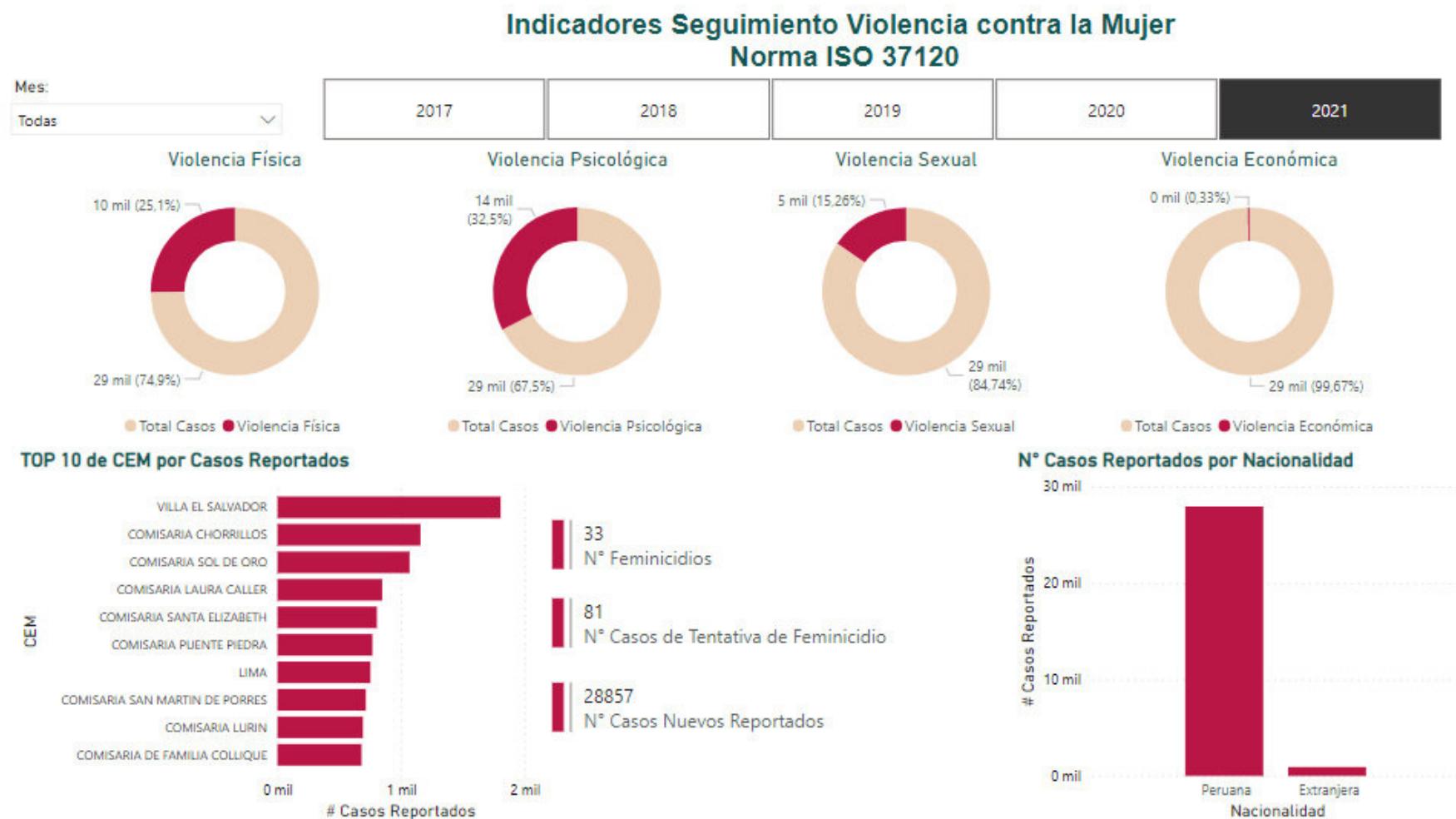


Fuente: Elaboración propia

A partir del modelado de datos y en base a los indicadores propuestos en la tabla 5.14, se elabora los siguientes paneles utilizando la herramienta Power BI, los cuales se muestran a continuación en la figura 5.14.

Figura 5.14.

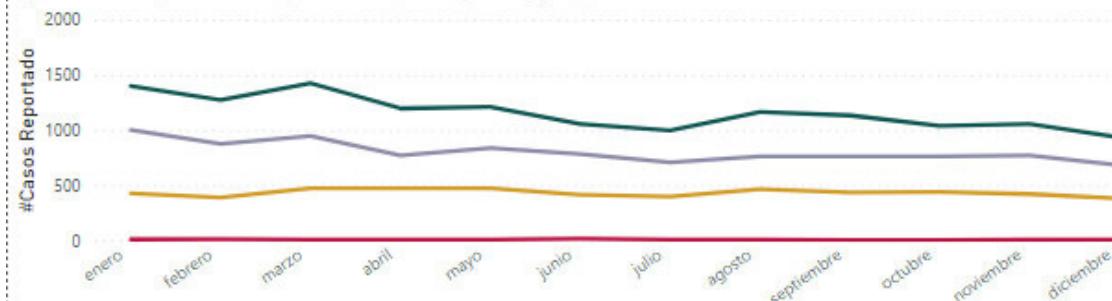
Indicadores propuestos basados en la ISO 37120:2018 plasmados en gráficos de la herramienta Power BI



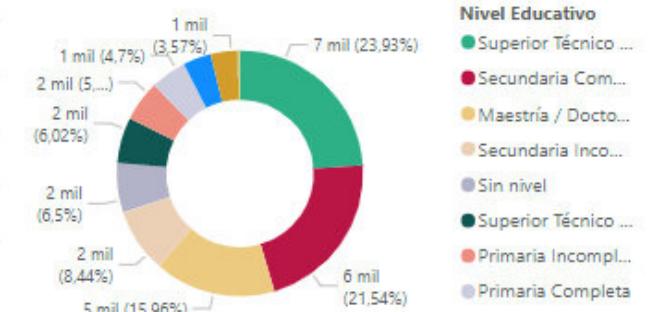
CEM	Mes:	Año:	Tipo de Violencia
Todas	Todas	2021	Todas

Evolución de Casos Reportados de Violencia contra la Mujer

Tipo Violencia: ● Económica o patrimonial ● Física ● Psicológica ● Sexual

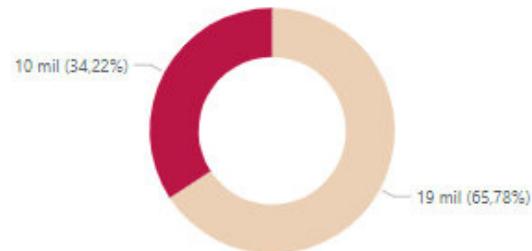


Nº Casos Reportados por Nivel Educativo



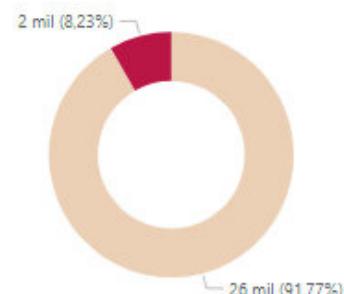
Nº Casos Reportados con Empleabilidad Víctima

Trabaja Víctima ● No ● Sí



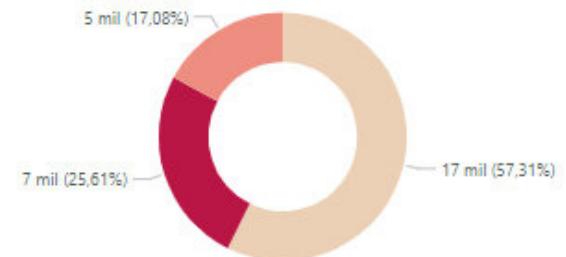
Nº Casos Reportados con Medidas Protección

M. Protección ● No ● Sí



Nº Casos Reportados por Nivel de Riesgo

Nivel de Riesgo ● Moderado ● Severo ● Leve



Fuente: Elaboración propia

En apoyo a la toma de medidas correctivas y preventivas en la lucha contra la violencia hacia la mujer, los paneles de la figura 5.14 muestran indicadores dirigidos a la alta dirección del programa Aurora.

Otro de los paneles desarrollados para la presente tesis se muestra en la figura 5.15.

Figura 5.15.

Panel de seguimiento de los casos de violencia contra la mujer reportados en Lima Metropolitana.

CEM: COMISARIA SANTA ELIZABETH | Mes: diciembre | Año: 2021

Identificador Profesional: 70558864 | Bienvenido(a):

Nº Ficha	Fecha de Ingreso	Nivel de Riesgo	Frecuencia Agresión	Tipo de Violencia	CEM	Vinculo con el Agresor	¿Caso Tentativa?	Última Fecha Seguimiento	Alerta
1-2021	01/12/2021	Moderado	Intermitente	Psicológica	COMISARIA SANTA ELIZABETH	Ex conviviente	No	08/01/2022	🔴
2-2021	01/12/2021	Severo	Intermitente	Física	COMISARIA SANTA ELIZABETH	Ex conviviente	No	09/01/2022	🔴
3-2021	01/12/2021	Moderado	Intermitente	Física	COMISARIA SANTA ELIZABETH	Conviviente	No	08/01/2022	🔴
14-2021	05/12/2021	Severo	Intermitente	Física	COMISARIA SANTA ELIZABETH	Ex conviviente	No	09/01/2022	🟡
16-2021	06/12/2021	Moderado	Primera vez	Física	COMISARIA SANTA ELIZABETH	Conviviente	Sí	09/01/2022	🟢

Nuevo caso de Violencia contra la mujer:

No es un posible Caso de Tentativa de Feminicidio

Ingresar Datos

| Fuente: Elaboración propia |

La tabla mostrada en la figura 5.15, forma parte de la propuesta solución a la problemática, debido a que permite dar un seguimiento a los casos de violencia registrados en los CEMs de acuerdo con las responsabilidades del equipo de atención definidas en (MIMP, 2018).

Como medida preventiva el panel de la figura 5.15 permite predecir si un nuevo caso reportado puede ser una tentativa de feminicidio, para ello se debe de llenar un formulario con los features necesarios para realizar la predicción con el modelo seleccionado en la fase predictiva; dicho formulario se accede luego de dar clic en el botón “Ingresar Datos”, como se muestra en la figura 5.16.

Figura 5.16.

Formulario de apoyo para la predicción de nuevos casos de tentativa de feminicidio.

The screenshot shows a web-based form titled "Predicción de Caso de Tentativa". At the top left is an email input field with "jose.20.mayte.01@gmail.com (no compartidos)" and a "Cambiar de cuenta" link. A note below says "*Obligatorio". The first section asks if the user is the informant, with two radio button options: "Sí" and "No". The second section asks if the victim is Peruvian, also with two radio button options: "Sí" and "No". The third section asks for the victim's age, with a text input field labeled "Tu respuesta".

Nota: El formulario completo se muestra en el Anexo 1. Fuente: Elaboración propia

Luego del llenado del formulario, se actualiza el panel de Power BI para evidencia el resultado del modelo predictivo, en el recuadro de color rojo mostrado en la figura 5.15, con respecto a si es o no un posible caso de tentativa de feminicidio.

5.3. Validación de modelo predictivo

En esta sección se detalla la validación del modelo predictivo seleccionado (Random Forest) para evaluar su precisión frente a casos nuevos y diferentes a la data de entrenamiento y prueba, utilizados en la fase predictiva, para ello se selecciona 68 casos de violencia contra la mujer reportados en los CEMs de los departamentos de la región costera del Perú, los cuales se encuentran en el universo de registros obtenidos en la fase de extracción y exploración de datos.

Estos registros nuevos se sometieron al modelo predictivo del algoritmo seleccionado, con el cual se obtiene los resultados detallados en la tabla 5.15, donde se presenta también los valores reales del target (“CASO_TENTATIVA_DE_FEMINICIDIO”) con el objetivo de comparar la precisión del modelo.

Tabla 5.15

Tabla de validación de los resultados del valor predicho por modelo predictivo vs el valor real de los nuevos casos de violencia de tentativa y no tentativa de feminicidio

Nº Caso	Valor predicho	Valor real	Acierto
1	1	1	VERDADERO
2	0	0	VERDADERO
3	0	0	VERDADERO
4	0	0	VERDADERO
5	0	0	VERDADERO
6	1	1	VERDADERO
7	0	0	VERDADERO
8	0	0	VERDADERO
9	0	0	VERDADERO
10	0	0	VERDADERO
11	0	0	VERDADERO
12	0	0	VERDADERO
13	0	0	VERDADERO
14	0	0	VERDADERO
15	0	0	VERDADERO
16	1	1	VERDADERO
17	1	1	VERDADERO
18	1	1	VERDADERO
19	1	1	VERDADERO
20	1	1	VERDADERO
21	0	0	VERDADERO

22	1	1	VERDADERO
23	1	1	VERDADERO
24	1	1	VERDADERO
25	1	1	VERDADERO
26	1	1	VERDADERO
27	1	0	FALSO
28	0	0	VERDADERO
29	0	0	VERDADERO
30	1	1	VERDADERO
31	1	1	VERDADERO
32	0	0	VERDADERO
33	0	0	VERDADERO
34	1	1	VERDADERO
35	1	1	VERDADERO
36	1	1	VERDADERO
37	1	1	VERDADERO
38	1	1	VERDADERO
39	1	1	VERDADERO
40	1	1	VERDADERO
41	1	1	VERDADERO
42	1	1	VERDADERO
43	1	1	VERDADERO
44	1	1	VERDADERO
45	1	1	VERDADERO
46	1	1	VERDADERO

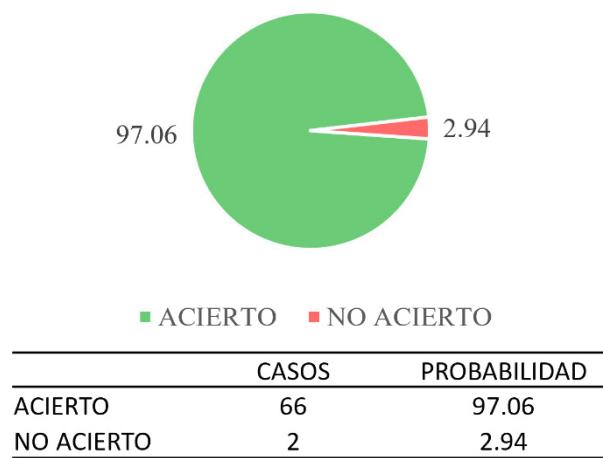
47	1	1	VERDADERO
48	1	1	VERDADERO
49	0	0	VERDADERO
50	1	1	VERDADERO
51	1	1	VERDADERO
52	0	0	VERDADERO
53	0	0	VERDADERO
54	0	0	VERDADERO
55	1	1	VERDADERO
56	1	1	VERDADERO
57	1	1	VERDADERO
58	1	1	VERDADERO
59	0	0	VERDADERO
60	1	1	VERDADERO
61	1	1	VERDADERO
62	1	1	VERDADERO
63	0	0	VERDADERO
64	1	1	VERDADERO
65	1	1	VERDADERO
66	1	1	VERDADERO
67	0	1	FALSO
68	1	1	VERDADERO

| Fuente: Elaboración propia |

En resumen, la figura 5.17 muestra la validación de la precisión del modelo predictivo seleccionado a 97% como se detalla en la fase de predictiva de la presente tesis.

Figura 5.17.

Porcentaje de precisión del modelo predictivo (Random Forest) de los nuevos casos de violencia.



Fuente: Elaboración propia

5.4. Discusión de resultados

Makond y Mayuening (2019), recomienda los modelos de machine learning frente a redes neuronales, debido a su practicidad e interpretabilidad. Alineado a ello y de acuerdo con lo revisado en el estado del arte, la solución tecnológica elegida para abordar la problemática propuesta es la analítica, la cual es desarrollada desde el enfoque descriptivo y predictivo.

Para la parte predictiva, se realizó un análisis y discusión de los resultados obtenidos de 5 modelos de machine learning, como se presenta a continuación:

Tabla 5.16

Tabla de validación de los resultados de algoritmos predictivos

Algoritmo	Parámetros	precision	recall	f1-score
KNN	n_neighbors=8	0 1	0.91 0.93	0.95 0.96
		accuracy		0.96
Regresión Logística	max_iter = 900, class_weight='balanced'	0 1	0.91 0.91	0.90 0.92
		accuracy		0.92
SVM	kernel =rbf, gamma=1, c=0,1 y probability=True	0 1	0.96 0.97	0.96 0.97
		accuracy		0.96
Árbol de Decisión	criterion='gini', max_depth=15, random_state=0	0 1	0.98 0.94	0.93 0.99
		accuracy		0.96
Random Forest	n_estimators=500, max_depth = 13, max_samples=0.2	0 1	0.99 0.95	0.94 0.99
		accuracy		0.97

Fuente: Elaboración propia

Los modelos de predicción de Regresión Logística y SVM, no son considerados para la selección del modelo, debido a que: Para el caso de Regresión Logística, las métricas utilizadas para medir el rendimiento no son mayores al 95% (porcentaje mínimo definido en la presente tesis para la aceptación del modelo) y el SVM, frente a nuevos casos no logra identificar correctamente si es una tentativa de feminicidio, lo que nos indica de que existe un sobreajuste a los datos de entrenamiento. Como se encuentran detallados en el apartado “5.2.4. Fases de analítica predictiva” de la presente tesis.

Continuando con la evaluación de los algoritmos de predicción, como se muestra en la tabla 5.16, los modelos de predicción basado en los algoritmos KNN, Árbol de decisión y Random Forest muestran un desempeño alto en las métricas de precisión, recall y F1-Score (detalle de interpretación mostrado en la tabla 5.17), lo que nos permite concluir que los modelos predictivos mencionados predicen correctamente, sin embargo, el que tiene mayor F1-Score es Random Forest, el cual fue el modelo seleccionado.

Tabla 5.17

Tabla con la interpretación de las métricas que permiten medir el rendimiento del modelo frente a casos de tentativa de feminicidio.

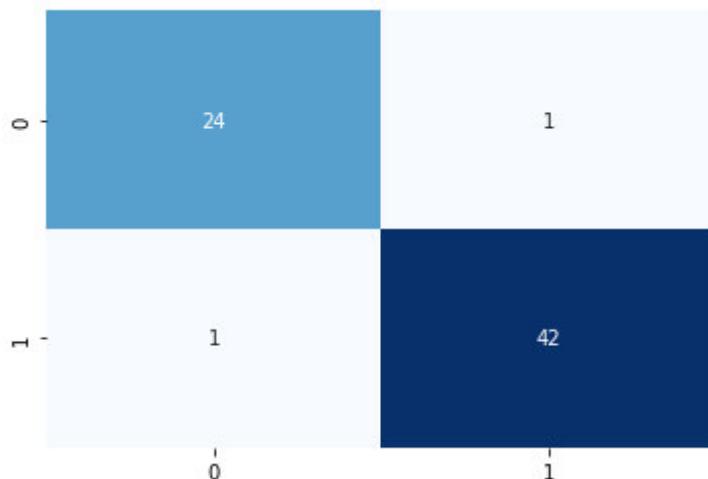
Métrica	Representación
Precisión	De los que se predijo que era tentativa de feminicidio, ¿Qué porcentaje efectivamente era tentativa?
Recall	De todos los que son tentativa de feminicidio, ¿Qué porcentaje identificó el modelo?
F1 Score	Media armónica entre Precisión y Recall

Fuente: Elaboración propia

Para validar el modelo predictivo elegido, Random Forest, se probó con 68 registros de nuevos casos de violencia contra la mujer, consiguiendo resultados satisfactorios, 97% de accuracy, similar a la métrica obtenida de las ejecuciones realizadas a la data de train y test, confirmando la estabilidad del modelo frente a nueva data. En la figura 5.18, muestra la matriz de confusión de los resultados obtenidos, de los cuales 24 registros con valor real 0 y 42 con valor real 1 fueron predichos correctamente, representando el 97% de precisión. Los resultados a más detalle se encuentran en la sección “5.3 validación de modelo predictivo” de la presente tesis.

Figura 5.18.

Matriz de confusión frente a nuevos casos de violencia contra la mujer.



Fuente: Elaboración propia

5.5. Resumen de los resultados obtenidos

Por todo lo detallado en la presente tesis, se puede concluir que se acepta la hipótesis planteada, debido a que la propuesta solución brinda un sistema informático que permitirá apoyar al personal encargado dentro del Centro de Emergencia Mujer en el seguimiento y prevención de los casos de violencia contra la mujer, mediante el uso de una plataforma de visualización que presenta indicadores a través de gráficos y/o tablas, así como la detección de las posibles tentativas de feminicidio a través de un modelo predictivo usando el algoritmo Random Forest.

Capítulo VI: Conclusiones y recomendaciones

6.1. Conclusiones

1. Se implementó un sistema informático basado en analítica predictiva y descriptiva, siendo esta última alineada a indicadores propuestos en base a la norma ISO 37120:2018, en apoyo a la prevención y monitoreo frente a los casos de violencia contra la mujer registrados en los Centro de Emergencia Mujer de Lima Metropolitana.
2. Se logró desarrollar un proceso de Extracción (Registros de casos de violencia contra la mujer desde el Portal Estadístico Programa Nacional Aurora), Transformación (Usando funciones y bibliotecas en Python) y Carga de datos usados en las fases de la analítica predictiva y descriptiva, detallada en el capítulo IV y V.
3. Luego de un análisis de los indicadores de la norma ISO 37120 - 2018 enfocados a la igualdad de género y lucha contra la violencia, se realizó una propuesta de 8 indicadores usados para el seguimiento de los casos de violencia contra la mujer reportados en los Centros de Emergencia Mujer de Lima Metropolitana,
4. Se implementó un modelo predictivo con el algoritmo Random Forest que permite predecir si un caso de violencia contra la mujer es tentativa de feminicidio o no. Logrando una precisión del 97%, el cual fue validado con registros distintos a los de entrenamiento y prueba obteniendo el mismo resultado de precisión.
5. Se implementó en la herramienta Power BI paneles de visualización que apoyan al monitoreo y prevención (junto al modelo predictivo) de los casos registrados de violencia contra la mujer en los Centro de Mujer de Lima Metropolitana.

6.2. Recomendaciones

- Se recomienda realizar una minuciosa validación de la coherencia de los datos obtenidos del Portal Estadístico Programa Nacional Aurora u otros portales de instituciones del Estado que contengan registros de denuncias de diferentes problemas sociales que enfrenta el país.
- Se recomienda aplicar el tema de la presente investigación a otros problemas sociales que aquejan al país en seguridad ciudadana, salud, ambiental, etc, los cuales también se encuentran abordados en los ámbitos de la norma ISO 37120:2018.
- Se recomienda alinear las tecnologías de las instituciones del estado con la propuesta de solución para la obtención de los repositorios internos.
- Se recomienda realizar la predicción de los casos de tentativas de feminicidio centrado en la variación de los niveles de riesgo en una serie de tiempo mediante un seguimiento de los casos.
- Se recomienda, como trabajo futuro, ampliar el ámbito de aplicación de la presente tesis a las demás regiones del Perú considerando el patrón de los casos de violencia contra la mujer, en cada uno de ellos.

ANEXO 1: Formulario de nuevos casos de tentativa

Predicción de Caso de Tentativa

***Obligatorio**

1. ¿La persona usuaria es la informante? *

Marca solo un óvalo.

Sí

No

2. ¿La víctima es peruana? *

Marca solo un óvalo.

Sí

No

3. ¿Cuál es la edad de la víctima? *

4. ¿La víctima está gestando? *

Marca solo un óvalo.

Sí

No

5. ¿La víctima está trabajando? *

Marca solo un óvalo.

Sí

No

6. ¿El agresor vive en casa de víctima? *

Marca solo un óvalo.

Sí

No

7. ¿El agresor está trabajando? *

Marca solo un óvalo.

Sí

No

8. ¿La víctima recibió gritos o insultos? *

Marca solo un óvalo.

Sí

No

9. ¿La víctima fue discriminada por su orientación sexual? *

Marca solo un óvalo.

Sí

No

10. ¿La víctima fue discriminada por su género? *

Marca solo un óvalo.

Sí

No

11. ¿La víctima fue discriminada por su identidad de género? *

Marca solo un óvalo.

Sí

No

12. ¿La víctima sufrió rechazo? *

Marca solo un óvalo.

Sí

No

13. ¿La víctima fue desvalorizada o humillada? *

Marca solo un óvalo.

Sí

No

14. ¿La víctima fue amenazada por quitar a sus hijos? *

Marca solo un óvalo.

Sí

No

15. ¿La víctima fue prohibida de estudiar, trabajar o salir? *

Marca solo un óvalo.

Sí

No

16. ¿El agresor rompe o destruye cosas? *

Marca solo un óvalo.

Sí

No

17. ¿Botaron a la víctima de su casa? *

Marca solo un óvalo.

Sí

No

18. ¿La víctima fue amenazada de muerte? *

Marca solo un óvalo.

Sí

No

19. ¿La víctima fue agredida con puntapiés y/o patadas? *

Marca solo un óvalo.

Sí

No

20. ¿La víctima fue agredida con puñetazos? *

Marca solo un óvalo.

Sí

No

21. ¿La víctima fue agredida con bofetadas? *

Marca solo un óvalo.

Sí

No

22. ¿La víctima fue agredida con jalones de cabello? *

Marca solo un óvalo.

Sí

No

23. ¿La víctima fue agredida con empujones? *

Marca solo un óvalo.

Sí

No

24. ¿La víctima fue agredida de golpes con palos? *

Marca solo un óvalo.

Sí

No

25. ¿La víctima fue agredida con latigazo? *

Marca solo un óvalo.

Sí

No

26. ¿La víctima fue agredida con ahorcamiento? *

Marca solo un óvalo.

 Sí No

27. ¿La víctima fue herida con armas? *

Marca solo un óvalo.

 Sí No

28. ¿La víctima fue golpeada con objetos contundentes? *

Marca solo un óvalo.

 Sí No

29. ¿La víctima fue agredida con negligencia? *

Marca solo un óvalo.

 Sí No

30. ¿Limitaron a la víctima con sus recursos económicos? *

Marca solo un óvalo.

 Sí No

31. ¿La víctima fue privada de medios indispensables? *

Marca solo un óvalo.

Sí

No

32. ¿La víctima fue controlada en sus ingresos? *

Marca solo un óvalo.

Sí

No

33. ¿La víctima sufrió violación? *

Marca solo un óvalo.

Sí

No

34. ¿La víctima sufrió actos contra el pudor? *

Marca solo un óvalo.

Sí

No

35. ¿La víctima depende de feminicidio? *

Marca solo un óvalo.

Sí

No

36. ¿La víctima cuenta con medidas protección? *

Marca solo un óvalo.

Sí

No

37. ¿La víctima cuenta con medidas cautelares? *

Marca solo un óvalo.

Sí

No

38. ¿La víctima desea patrocinio legal? *

Marca solo un óvalo.

Sí

No

39. ¿Cuál es la edad del agresor? *

40. ¿El agresor es peruano? *

Marca solo un óvalo.

Sí

No

41. ¿El agresor sufre de alguna adicción? *

Marca solo un óvalo.

Sí

No

42. ¿La víctima sufre de alguna adicción? *

Marca solo un óvalo.

Sí

No

43. ¿La víctima cuenta con un vínculo de soporte? *

Marca solo un óvalo.

Sí

No

44. ¿La víctima interpuso denuncia? *

Marca solo un óvalo.

Sí

No

45. ¿La víctima recibe tratamiento? *

Marca solo un óvalo.

Sí

No

46. ¿Cuál fue el mes de la agresión? *

Marca solo un óvalo.

- Enero
- Febrero
- Marzo
- Abril
- Mayo
- Junio
- Julio
- Agosto
- Setiembre
- Octubre
- Noviembre
- Diciembre

47. Seleccione el CEM donde se registro el caso *

Marca solo un óvalo.

- CALLAO
- CHOSICA
- CIENEGUILA
- COMISARIA ALFONSO UGARTE
- COMISARIA ANCON
- COMISARIA APOLO
- COMISARIA CAJA DE AGUA
- COMISARIA CARMEN DE LA LEGUA REYNOSO
- COMISARIA CHORRILLOS
- COMISARIA CONDEVILLA
- COMISARIA DE FAMILIA CANTO REY
- COMISARIA DE FAMILIA COLLIQUE
- COMISARIA DE FAMILIA EL AGUSTINO
- COMISARIA HUAYCAN
- COMISARIA JUAN INGUNZA VALDIVIA
- COMISARIA LAS PRADERAS
- COMISARIA LAURA CALLER
- COMISARIA LINCE
- COMISARIA LURIN
- COMISARIA PAMPLONA II
- COMISARIA PLAYA RIMAC
- COMISARIA PUEBLO LIBRE
- COMISARIA PUENTE PIEDRA
- COMISARIA RAMON CASTILLA
- COMISARIA SAGITARIO
- COMISARIA SAN ISIDRO
- COMISARIA SAN MIGUEL
- COMISARIA SANTA ANITA
- COMISARIA SANTA CLARA
- COMISARIA SANTA ELIZABETH
- COMISARIA SOL DE ORO
- COMISARIA VILLA ALEJANDRO

- COMISARIA VILLA MARIA DEL TRIUNFO
- COMISARIA VITARTE
- EL AGUSTINO
- HUAYCAN
- LA MOLINA
- LA VICTORIA
- LIMA
- LOS OLIVOS
- MANCHAY
- PUENTE PIEDRA
- RIMAC
- SAN JUAN DE LURIGANCHO
- SAN JUAN DE MIRAFLORES
- SAN LUIS
- SURCO
- VENTANILLA
- VILLA EL SALVADOR
- CHACLACAYO
- COMAS
- COMISARIA BARRANCO
- COMISARIA BAYOVAR
- COMISARIA CIUDAD Y CAMPO
- COMISARIA INDEPENDENCIA
- COMISARIA MARISCAL CACERES
- COMISARIA PALOMINO
- COMISARIA SAN BORJA
- COMISARIA SAN MARTIN DE PORRES
- SAN MIGUEL - LIMA
- SURQUILLO
- VILLA MARIA DEL TRIUNFO
- CARABAYLLO
- COMISARIA MARANGA
- COMISARIA SAN PEDRO DE CARABAYLLO
- MAGDALENA DEL MAR
- PACHACUTEC

SANTA ROSA

48. ¿Cuál es la condición del caso presentado? *

Marca solo un óvalo.

- Derivado
- Nuevo
- Reincidente
- Reingreso
- Continuador

49. ¿Cuál es la lengua materna de la víctima? *

Marca solo un óvalo.

- Castellano
- Quechua
- Awajún
- Ingles
- Portugués
- Asháninka
- No escucha y/o no habla
- Aimara
- Otra lengua extranjera
- Shipibo
- Lengua de señas

50. ¿Cuál es la etnia de la víctima? *

Marca solo un óvalo.

- Mestizo
- Quechua
- Negro, moreno, zambo, mulato/pueblo afroperuano
- Blanco
- Nativo o indígena de la amazonía
- No sabe
- No especifica
- Otro
- Perteneciente de otro pueblo indígena u originario
- Aimara
- Indígena u originario de la Amazonía

51. ¿Cuál es el estado civil de la víctima? *

Marca solo un óvalo.

- Soltero/a
- Casado/a
- Viudo/a
- Divorciado/a

52. ¿Cuál es el nivel educativo de la víctima? *

Marca solo un óvalo.

- Secundaria Incompleta
- Primaria Incompleta
- Superior Universitario Incompleto
- Secundaria Completa
- Primaria Completa
- Superior Técnico Incompleto
- Superior Técnico Completo
- Superior Universitario Completo
- Inicial
- Sin nivel
- Maestría / Doctorado
- Básica Especial

53. ¿Cuál es el vínculo entre el agresor y la víctima? *

Marca solo un óvalo.

- Vínculo Relacional Familiar
- Vínculo Relacional de Pareja
- Sin Vínculo Relacional de Pareja Ni Familiar

54. ¿Cuál es el nivel educativo del agresor? *

Marca solo un óvalo.

- Secundaria Incompleta
- Primaria Completa
- Secundaria Completa
- Sin nivel
- Superior Técnico Incompleto
- Superior Universitario Incompleto
- Superior Universitario Completo
- Primaria Incompleta
- Superior Técnico Completo
- Maestría / Doctorado
- Inicial

55. ¿Cuál es el nivel de riesgo del caso? *

Marca solo un óvalo.

- Moderado
- Severo
- Leve

56. ¿Cuál es la relación entre víctima y agresor? *

Marca solo un óvalo.

- Primo/a
- Ex conviviente
- Conviviente
- Cónyuge
- Desconocido/a
- Otro
- Hijo/a
- Padrastro/Madrastra
- Enamorado/a
- Ex enamorado/a
- Tío/a
- Progenitor/a de su hijo/a (sin convivencia con la pareja)
- Sobrino/a
- Hermano/a
- Vecino/a
- Cuñado/a
- Padre/Madre
- Ex cónyuge
- Hijastro/a
- Abuelo/a
- Compañero/a de trabajo
- Tío/a-abuelo/a
- Empleador/a de trabajo
- Yerno/Nuera
- Docente
- Nieto/a
- Otro familiar
- Compañero/a de estudios
- Concuñado/a
- Sobrino/a-nieto/a
- Empleado/a de trabajo
- Hermanastro/a

- Habita en el mismo hogar
- Ex novio/a
- Suegro/a
- Novio/a
- Bisabuelo/a
- Bisnieto/a
- Otro tipo de relación sexo-afectiva

57. ¿Cuál es la frecuencia de agresión de la víctima? *

Marca solo un óvalo.

- Primera vez
- Intermitente
- Mensual
- Semanal
- Diario
- Quincenal

58. ¿Qué tipo de violencia sufrió la mujer? *

Marca solo un óvalo.

- Física
 - Psicológica
 - Sexual
 - Económica o patrimonial
-

REFERENCIAS BIBLIOGRÁFICAS

Adesola, F., Misra, S., Omoregbe, N., Damasevicius, R. y Maskeliunas, R. . (2019), An IOT-Based Architecture for Crime Management in Nigeria. *Data, Engineering and Applications. Springer, Singapore*, 245-254. https://doi.org/10.1007/978-981-13-6351-1_19.

Aggarwal, C. (2018) Neural Networks and Deep Learning. Springer Cham (1st ed.).

<https://doi.org/10.1007/978-3-319-94463-0>

Aguilar, D., Camargo, D. (2021). Sistema inteligente basado en redes neuronales, máquina de soporte vectorial y random forest para la predicción de deserción de clientes en microcréditos de bancos. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos]. Cybertesis, Repositorio de Tesis Digitales UNMSM.

<https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16390>

Amagasa, T., Komamizu, T., Nakazawa, J., Kitagawa H. y Tokuda, H. (2017). Analytical toolbox for smart city applications: Garbage collection log use case. *3rd International Workshop on Smart Cities: People, Technology, and Data, co-located with IEEE Big Data 2017*, 4105–4110. doi:10.1109 / BigData.2017.8258429.

Bernachea, c., Chilet, E., Guzmán, P., Inche, V., Leon, J. (2021). Técnica de Machine Learning para el cálculo de la probabilidad de fuga de los clientes de la empresa Bitel. [Trabajo de suficiencia, Universidad ESAN]. Repositorio de la Universidad ESAN.

<https://repositorio.esan.edu.pe/handle/20.500.12640/2929>

Defensoría del Pueblo [DP]. (2018). *Segunda supervisión a nivel de Nacional 2018, p 80.*

https://www.defensoria.gob.pe/wpcontent/uploads/2019/07/INFORME-CEM_OCT-Y-NOV-2018_vf2-1.pdf (Recuperado 7 marzo del 2021)

Defensoría del Pueblo [DP]. (2020). *Violencia contra las mujeres se mantiene latente e invisibilizada durante estado de emergencia.* <https://www.defensoria.gob.pe/violencia-contra-las-mujeres-se-mantiene-latente-e-invisibilizada-durante-estado-de-emergencia/>

Guerrero, A., Cárdenas, J., Romero, V. y Ayma V. (2021). Comparison of Classifiers Models for Prediction of Intimate Partner Violence. *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 2. Advances in Intelligent Systems and Computing,* 1289. Springer, Cham, 469 – 488, https://doi.org/10.1007/978-3-030-63089-8_30

Instituto Nacional de Estadística e Informática [INEI]. (2019). *Encuesta Demográfica y de Salud Familia - 2019.*

https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1736/Libro.pdf

ISO (2018). Organización Internacional de Normalización. Sustainable cities and communities - Indicators for city services and quality of life (ISO 37120).

<https://www.iso.org/standard/68498.html>

Intel (2017). *Getting Started with Advanced Analytics.*

<https://www.intel.com/content/www/us/en/analytics/getting-started-advanced-analytics-planning-guide.html>

Kashyap, K. (2019) Random Forest Classifier. <https://medium.com/analytics-vidhya/machine-learning-decision-trees-and-random-forest-classifiers-81422887a544>

Kmuar, A., Tiwar, N., Gupta, P. y Rajan, S.N (2017). Women Crime Prediction. *International Research Journal of Engineering and Technology (IRJET)*, 4(4), 2395-2396.

<https://www.irjet.net/archives/V4/i4/IRJET-V4I4491.pdf>

Kumar, S., Tiwari, P. y Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: a review. *Journal of Big Data*, 6(1).

<https://doi.org/10.1186/s40537-019-0268-2>

Lockett, K., Naden, C., Oclarin, R. (2020). Creamos Ciudades Sostenibles. *Revista de la Organización Internacional de Normalización (ISOfocus)*.

[https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISOfocus%20\(2013-NOW\)/sp/ISOfocus_140_sp.pdf](https://www.iso.org/files/live/sites/isoorg/files/news/magazine/ISOfocus%20(2013-NOW)/sp/ISOfocus_140_sp.pdf)

Makond, B., Mayuening, P. (2019). Predictive Models for Classifying the Outcomes of Violence: Case Study for Thailand's Deep South. *Advances in Decision Sciences*, 23(3), 59-92.

<https://doi.org/10.47654/v23y2019i3p56-92>

Ministerio de Justicia y Derechos Humanos [MINJUS]. (s.f.). *Sistema de Protección de DD.HH.* <https://www.minjus.gob.pe/sistema-de-proteccion-de-dd-hh/#secpreg>

Ministerio de la Mujer y Poblaciones Vulnerables [MIMP]. (2019). *Protocolo Base de Actuación*

Conjunta en el ámbito de la atención integral y protección frente a la violencia contra las mujeres y los integrantes del grupo familiar.

https://cdn.www.gob.pe/uploads/document/file/313206/DS_012-2019-

[MIMP APRUEBAN PROTOCOLO DE BASE CONJUNTA ... CONTRA LA VIOLENIA ANEXO.pdf](#)

Ministerio de la Mujer y Poblaciones Vulnerables [MIMP]. (2018). *Protocolo de Actuación*

Conjunta de los Centros de Emergencia Mujer y Comisarías o Comisarías Especializadas en materia de protección contra la Violencia Familiar de la Policía Nacional del Perú.

<https://observatorioviolencia.pe/wp-content/uploads/2018/08/GUIAds-006-2018-mimp-protocolo-actuacion-conjunta-cem-comisarias.pdf>

Moschen, S., Macke, J., Bebber, S. y Benetti, M. (2019). Sustainable development of communities: ISO 37120 and UN goals. International Journal of Sustainability in Higher Education, 20(5), 887-900. <https://doi.org/10.1108/IJSHE-01-2019-0020>

Muskan, Khandelwal, T., Khandelwal, M. y Pandey, P. (2018). Women Safety Device Designed Using IoT and Machine Learning. 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI, pp. 1204-1210. 10.1109/SmartWorld.2018.00210.

Nair A., Prasad P. y Dr. Godfrey S. (2021). Crime Against Women: Analysis And Prediction.

International Journal of Engineering Research & Technology (IJERT), 10.

<https://www.ijert.org/crime-against-women-analysis-and-prediction>

Nuraeni, F., Nelis, N., Listiani, L. y Rahmawati, E. (2019). Implementation of K-Means Algorithm with Distance of Euclidean Proximity in Clustering Cases of Violence Against Women and Children. *International Conference on Cybernetics and Intelligent System (ICORIS)*, 162-167. 10.1109/ICORIS.2019.8874883

Observatorio Nacional de la Violencia Contra las Mujeres y los Integrantes del Grupo Familiar [ONVCMIGF]. (2017). *Estadísticas del Poder Judicial*.

<https://observatorioviolecia.pe/estadisticas-del-poder-judicial-2/>

Organismo Mundial de la Salud [OMS]. (2021). *Violencia contra la mujer*.

<https://www.who.int/es/news-room/fact-sheets/detail/violence-against-women>

Pagliaro, F., Mattoni, B., Gugliermenti, F., Bisegna, F., Azzaro, B., Tomei, F. y Catucci, S. (2016). A roadmap toward the development of Sapienza Smart Campus. *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*. 10.1109/eeeic.2016.7555573

Plataforma digital única del Estado Peruano [PDUEP]. (2021). *Centro de Emergencia Mujer - CEM*. <https://www.gob.pe/480-denunciar-violencia-familiar-y-sexual-centros-de-emergencia-mujer-cem>

Portal estadístico Programa Nacional AURORA MIMP [PEPNA]. (2020). *Reporte estadístico de casos con características de feminicidio atendidos por los servicios del Programa Nacional AURORA*. <https://portalestadistico.aurora.gob.pe/cartillas-estadisticas/>

Portal estadístico Programa Nacional AURORA MIMP [PEPNA]. (2021a). *Cartilla estadística, cifras de violencia contra las mujeres*.

https://www.mimp.gob.pe/files/programas_nacionales/pncvfs/estadistica/ResEstad_Femicidio_Tentativas2020_7.xlsx

Portal estadístico Programa Nacional AURORA MIMP [PEPNA]. (2021b). *Base de datos*.

<https://portalestadistico.aurora.gob.pe/bases-de-datos-2021/>

Power, D., Heavin, C., McDermott, J. y M. Daly (2018). Defining business analytics: an empirical approach. *Journal of Business Analytics*, 1(1), 40-53.
10.1080/2573234X.2018.1507605

Programa de las Naciones Unidas para el Desarrollo [PNUD]. (2021), *Objetivo 5: Igualdad de Género*. <https://www.undp.org/es/sustainable-development-goals#igualdad-genero>

Quiza, J. (2018). Regresión Logística con Pytorch. <https://medium.com/datos-y-ciencia/regresi%C3%B3n-log%C3%ADstica-conpytorch-d2212b912b6b>

Ruiz, G. (2019). Modelo de análisis de datos utilizando técnicas de aprendizaje supervisado y no supervisado, para identificar patrones en la información generada por los pacientes,

sometidos a juegos diseñados como un instrumento de apoyo terapéutico. [Tesis Magister, Universidad Jorge Tadeo Lozano]. Repositorio institucional de Universidad Jorge Tadeo Lozano. Recuperado de:

<https://expeditiorepositorio.utadeo.edu.co/bitstream/handle/20.500.12010/8502/Trabajo%20de%20grado.pdf?sequence=1&isAllowed=y>

Saboya, N., Sullon, A. y Loaiza, O. (2019). Predictive Model Based on Machine Learning for the Detection of Physically Mistreated Women in the Peruvian Scope. *3rd International Conference on Advances in Artificial Intelligence Association for Computing Machinery, New York, NY, USA*, 18-23, <https://doi.org/10.1145/3369114.3369143>

Scikit learn. (2022a). Random Forest Classifier. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

Scikit learn (2022b). Regresión Logística. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Scikit learn (2022c). Árboles de decisión. <https://scikit-learn.org/stable/modules/tree.html>

Scikit learn (2022d). Árboles de decisión Clasificación. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier>

Scikit learn (2022e). Máquinas de Soporte Vectorial Clasificación. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC>

Scikit learn (2022f). Algoritmo KNN. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html#sklearn.neighbors.NearestNeighbors>

Sivanagaleela, B. y Rajesh, S. (2019). Crime Analysis and Prediction Using Fuzzy C-Means Algorithm. *Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019)*. IEEE. doi: 10.1109/ICOEI.2019.8862691

United Nations [UN]. (2015), *Sustainable development goals- Goal 5: Achieve gender equality and empower all women and girls.*

<https://www.un.org/sustainabledevelopment/es/gender-equality/>

Ullah, F., Obaida, M., Muhammad, K., Ullah, A., Baik, S., Cuzzolin, F. y Albuquerque V. (2021). An intelligent system for complex violence pattern analysis and detection.

International Journal of Intelligent Systems. <https://doi.org/10.1002/int.22537>

Viviano, T. (2019). *Se implementa el registro integrado de feminicidio.* <https://observatorioviolecia.pe/se-implementa-el-registro-integrado-de-feminicidio/>

Viviano, T. (2018). *Estado de la cuestión y necesidad de estandarizar e integrar los registros administrativos sobre violencia hacia las mujeres y los integrantes del grupo familiar.* <https://observatorioviolecia.pe/wp-content/uploads/2018/10/taller-estandarizacion-teresa1-1.pdf>

Xia, Q., Zhang, P., Wang, J., Tian, M. y Fei, C. (2018). Real Time Violence Detection Based on Deep Spatio-Temporal Features. *Biometric Recognition. CCBR 2018. Lecture Notes in Computer Science()*, Volume 10996. Springer, Cham, https://doi.org/10.1007/978-3-319-97909-0_17