ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

Aplicación de modelos de aprendizaje automático para pronosticar un ratio de daños ocasionados por inundaciones.

PROYECTO INTEGRADOR

Previo a la obtención del Título de:

Ingeniero Estadístico

Presentado por:

Sebastian Francisco Poveda Sandoval

Santiago Gerardo Salazar Ruiz

GUAYAQUIL - ECUADOR

Año: 2024

DEDICATORIA

Le dedico el resultado de este trabajo a toda mi familia, en especial a mis hermanas Paula y María de los Ángeles quienes han sido una fuente constante de apoyo y fortaleza.

Deseo de todo corazón que encuentren en cada página de esta tesis una motivación para seguir adelante.

Quiero finalizar esta dedicatoria con una promesa; siempre estaré allí para ustedes, celebrando cada pequeño paso que den y alentándolas en cada desafío que se enfrenten.

Con todo mi amor,

Sebastian Poveda S.

DEDICATORIA

A mi querida familia,

En las páginas de este trabajo se refleja más que el esfuerzo académico y la dedicación a la investigación; se plasman las horas de apoyo incondicional, los momentos de aliento y la fortaleza que me han brindado a lo largo de este viaje. Cada palabra escrita lleva consigo una parte de las enseñanzas y valores que me han transmitido, siendo el cimiento sobre el cual he construido mis sueños y aspiraciones.

A mi madre, luz y guía de mis días, a ti te debo el coraje para enfrentar los desafíos y la inspiración para perseguir mis metas. Tu amor y sacrificio han sido la fuente de motivación que me ha impulsado a continuar, incluso cuando el camino parecía incierto. Tu fe inquebrantable en mí ha sido el viento bajo mis alas, permitiéndome volar más alto de lo que jamás imaginé.

Santiago Salazar R.

AGRADECIMIENTOS

Mi agradecimiento más profundo a Dios en primer lugar, por haberme dado la fortaleza y la salud para terminar esta memorable etapa de mi vida. A mis amados padres, Francisco y María Augusta por ser ese faro que ha iluminado cada instancia de mi vida, su amor y llenarlos de orgullo ha sido mi motor más grande.

Agradezco también a mi padre Ricardo su apoyo incondicional y por compartir conmigo los momentos más cruciales de mi vida. A Paula y Allison, con quienes inicié esta etapa universitaria, gracias por permanecer conmigo en todo este proceso.

Y finalmente un reconocimiento especial a mis hijos Milagros y Camilo por su fiel compañía y por estar a mi lado en cada desafío.

Sebastian Poveda S.

AGRADECIMIENTOS

Con profunda gratitud, elevo mi agradecimiento a Dios, cuya guía ha sido mi faro en la oscuridad y mi fortaleza en la duda.

A mi familia, amigos y profesores, su amor y apoyo han sido el refugio y la energía que me impulsaron a seguir adelante. Son parte esencial de este logro.

Un especial reconocimiento a Mariela y Omar, profesores que han dejado una marca indeleble en mi alma y en mi formación. Su enseñanza trasciende el conocimiento; han forjado mi carácter y mi corazón.

Gracias a todos por ser la luz en mi camino.

Con el alma llena de agradecimiento.

Santiago Salazar R.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; Sebastian Francisco Poveda Sandoval y Santiago Gerardo Salazar Ruíz damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"

Sebastian Poveda S.

Santiago Salazar R.

EVALUADORES

Firmado electrónicamente por:
SANDRA LORENA
GARCIA BUSTOS

Ph.D. Sandra García Bustos

PROFESOR DE LA MATERIA

Firmado electrónicamente por:
HEYDI MARIANA ROA
LOPEZ

M.Sc. Heydi Roa López

PROFESOR TUTOR

RESUMEN

El presente proyecto tiene como objetivo la identificación de los sectores de Guayaquil que presentan una mayor propensión a sufrir daños significativos a nivel de extensión promedio, a consecuencia de inundaciones. La relevancia de esta iniciativa reside en que las inundaciones representan una de las principales amenazas naturales que enfrenta Guayaquil, y es crucial abordar este desafío de manera proactiva y estratégica. Se propone la aplicación del método de aprendizaje automático Random Forest para predecir los daños por sector, utilizando variables sociodemográficas proporcionadas por el Centro Internacional del Pacífico para la Reducción del Riesgo de Desastres, variables geográficas suministradas por la Municipalidad de Guayaquil y datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo del Instituto Nacional de Estadística y Censos (INEC). Este estudio adquiere relevancia al vincularse con el Objetivo de Desarrollo Sostenible (ODS) 11, ya que la identificación de zonas de riesgo permite dirigir la atención y asignar recursos de manera específica para prevenir posibles daños futuros. El proceso de entrenamiento del modelo incorporó diversos parámetros y técnicas, culminando en una precisión del 57%. Aunque se reconoce que aún existen oportunidades de mejora, el modelo demuestra la capacidad de realizar estimaciones en metros cuadrados del área total afectada por sector debido a inundaciones, considerando variables sociodemográficas obtenidas del ENEMDU. El trabajo realizado representa un avance significativo en la aplicación de modelos de aprendizaje automático para mejorar la calidad de vida de los habitantes de Guayaquil.

Palabras Clave: Inundaciones, Calidad de vida, Gestión de Riesgos, Random Forest.

ABSTRACT

This project aims to identify Guayaquil's most flood-prone areas. The relevance of this initiative lies in the fact that flooding is one of the main natural hazards facing Guayaquil, and it is crucial to address this challenge proactively and strategically. Using socio-demographic variables provided by the Pacific International Center for Disaster Risk Reduction, geographical variables provided by the Municipality of Guayaquil, and data from the National Survey of Employment, Unemployment and Underemployment of the National Institute of Statistics and Census (INEC), the application of the Random Forest machine learning method is proposed to predict damages by sector. This study is relevant in the context of Sustainable Development Goal (SDG) 11, as the identification of risk areas allows for the targeting of attention and the allocation of resources in order to prevent possible future damage. The model training process incorporated various parameters and techniques, culminating in an accuracy of 57%. Although it is recognised that there is still room for improvement, the model demonstrates the ability to estimate the total area affected by flooding per sector in square metres, taking into account socio-demographic variables obtained from the ENEMDU. This work represents significant progress in applying machine learning models to improve the quality of life of Guayaquil residents.

Keywords: *Flooding, Sustainable Development Goal (SDG), Risk management, Random Forest.*

ÍNDICE GENERAL

ABREVIATURAS

ESPOL	Escuela Superior Politécnica del Litoral
ENEMDU	Encuesta Nacional de Empleo, Desempleo y Subempleo
INEC	Instituto Nacional de Estadística y Censos
CIP-RRD	Centro Internacional del Pacífico para la Reducción del Riesgo de Desastres
DT	Decision Trees
ID3	Iterative Dichotomiser 3
CART	Classification and Regression Trees
RF	Random Forest

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

CAPÍTULO 1

1. INTRODUCCIÓN

Los riesgos naturales, son acontecimientos extremos y repentinos causados por factores medioambientales. Dentro de estos, podemos encontrar tormentas, inundaciones, sequías, incendios y olas de calor. Una inundación, también conocida como riada, se define como la ocupación por el agua en zonas o áreas que en condiciones normales se encuentran secas (Christina Nuñez, 2010).Puede variar en su nivel de gravedad, alcance e impacto.

Son causadas por factores naturales, como lluvias, oleajes, deshielos, tormentas costeras, nieve, incremento del nivel de los ríos, pero principalmente por la ocurrencia de lluvias intensas durante un corto periodo de tiempo. Las frecuentes precipitaciones en áreas pequeñas pueden generar crecidas, lo que los hace màs vulnerables ante el peligro de inundación, estos resultados se muestran mediante la lectura de los hidrogramas, gráficos utilizados para mostrar la variación en el tiempo de alguna información hidrológica.

Una combinación de crecimiento económico y demográfico ha provocado la construcción y la modificación de muchos paisajes naturales con fines agrícolas, industriales y urbanos (Geography Revision Edexcel, 2021). Derivándose de esto, actividades como diseños inadecuado de canales y estructuras de drenaje, obstrucción por escombros traídos por las aguas de crecida, la destrucción de ecosistemas, vegetaciones que absorben el agua y evitan fuertes corrientes. Estas mismas acciones, aumentan con frecuencia el riesgo de inundación.

Los fenómenos naturales, como las inundaciones, pueden causar daños devastadores. Por ejemplo, Aragón-Durand, 2014 informó que Argentina sufrió el mayor daño económico por inundaciones en América Latina y el Caribe entre 1900 y 2013, con $9.998.210\$$ en 48 ocurrencias de este tipo de fenómenos, afectando a una población de 14.102.249 personas y resultando en 836 muertes. Además de las consecuencias fatales y económicas de estos eventos, también existen consecuencias socioculturales y ambientales, como el retroceso en la educación, la falta de alimentos, el aumento de enfermedades y la destrucción de los sistemas naturales. Estas consecuencias son más pronunciadas en las zonas urbanas debido a las características demográficas y sociales de las ciudades.

En el contexto de la presente investigación, Guayaquil es una ciudad costera situada en la provincia de Guayas en Ecuador, que se ha visto afectada por fuertes lluvias y eventos de inundaciones en el pasado. La geografía de la ciudad, con su ubicación entre el río Guayas y el Estuario Salado, la hace susceptible a inundaciones durante la temporada de lluvias. Un estudio de José Núñez Ramos y Jordy Bastidas Guerrero 2021 identificó que las características más significativas que hacen vulnerables a inundaciones a ciertos sectores de Guayaquil son el tipo de suelo, sobre todo si es fangoso debido a que retiene el agua con facilidad, si la zona experimenta inundaciones anuales, y si los sectores se ubican en la zona norte de la ciudad. En los últimos años, Guayaquil ha sufrido eventos de inundaciones que han causado daños significativos, como los ocurridos en marzo 2023, donde se reportó dos derrumbes estructurales, árboles caídos, deslizamientos de tierra y el desplegó de ocho embarcaciones inflables en las zonas afectadas. Sin embargo, no existe levantamiento de información de daños desagregados en la ciudad de Guayaquil. Por lo tanto, se propone una investigación para estimar el ratio de

daños ocasionados por inundaciones. Este ratio se estimará mediante la cuantificación de pérdidas materiales promedio por sectores urbanos, utilizando variables socioeconómicas, geográficas y temporales que permitirán una gestión adecuada y eficiente de los recursos disponibles.

Snehil y Ruchi Goel (2020), concluyeron que Random Forest es una de las técnicas más utilizadas y que ha dado óptimos resultados frente a otros modelos para pronosticar el ratio de daños ocasionados por inundaciones. Por lo tanto, se propone aplicar un modelo de Random Forest para estimar el ratio promedio de daños materiales provocados por inundaciones, lo que servirá como variable dependiente, inspirada en la investigación de Bidadian, Maxwell y Strager (2023). La implementación de modelos de aprendizaje automático para pronosticar ratios de daños en inundaciones es crucial para minimizar los daños materiales, mejorar la precisión y eficiencia, y proporcionar herramientas para un diseño de sistemas de alerta temprana que permitan una mejor gestión de riesgo. La inteligencia artificial, en particular el aprendizaje automático, desempeña un papel fundamental en la reducción de riesgos de desastre, ya que puede analizar grandes cantidades de datos y generar alertas tempranas precisas, lo que puede ayudar a tomar decisiones asertivas. Los sistemas de alerta temprana basados en la comunidad también son de gran ayuda para alertar y dirigir acciones apropiadas en caso de desastres, reduciendo pérdidas y daños.

1.1 Descripción del problema

Guayaquil, una metrópolis de Ecuador, es altamente vulnerable a inundaciones recurrentes, que se intensifican por fenómenos climáticos como El Niño. A pesar de la alta exposición a estas amenazas naturales, el país carece de capacidad suficiente para gestionar

eficazmente las consecuencias de las inundaciones, particularmente en las zonas urbanas y rurales más susceptibles, donde la devastación causada por las inundaciones puede tener un impacto catastrófico. El desafío que se aborda en esta investigación es la identificación precisa de las zonas más propensas a inundaciones en Guayaquil y la proyección efectiva del radio de daño causado por estos eventos, aspecto que no ha sido investigado de manera exhaustiva en esta ciudad. Los requisitos para este proyecto incluyen la implementación del modelo de aprendizaje automático Random Forest, el método de análisis de importancia de permutación condicional (CPI) y métodos multivariados. Entre las limitaciones se encuentra la necesidad de minimizar los recursos invertidos por el municipio de Guayaquil para enfrentar este fenómeno.

Desde el punto de vista metodológico, se utilizará únicamente Random Forest como algoritmo de aprendizaje estadístico, y nos centraremos geográficamente en la ciudad de Guayaquil. Las variables de interés más relevantes en el estudio son las zonas más vulnerables a inundaciones en Guayaquil, la proporción de daños causados por inundaciones en ciertos sectores de la ciudad, las características biofísicas intrínsecas de la tierra y el agua en Guayaquil, los factores socioeconómicos de las comunidades urbanas, características ambientales y características físicas de las áreas vulnerables. La relevancia de este proyecto radica en que las inundaciones representan una de las principales amenazas naturales que enfrenta Guayaquil, y es imperativo abordar este desafío de manera proactiva y estratégica. Este es un problema actual que es susceptible de observación, medición y análisis. Nuestro cliente es el Centro Internacional del Pacífico para la Reducción del Riesgo de Desastres, una reconocida organización afiliada a la ESPOL. La Dra. María del Pilar Cornejo, nuestra contacto en la organización, ha manifestado interés en este proyecto por su potencial para minimizar los

gastos producidos por los daños a infraestructuras por inundaciones y salvar vidas. Este proyecto busca responder a la siguiente pregunta: ¿Cómo podemos utilizar modelos de aprendizaje automático para identificar las zonas más propensas a inundaciones y pronosticar el radio de daño causado por este tipo de fenómenos?.

1.2 Justificación del problema

La relevancia de este proyecto radica en su enfoque en la identificación y cuantificación del daño en diversas zonas urbanas de Guayaquil a través de la aplicación de modelos de aprendizaje automático, como el Random Forest y sus variantes. La iniciativa busca abordar un problema de gran trascendencia, vinculado a la evaluación de la eficacia de las medidas implementadas para la mitigación de los daños causados por inundaciones. Estas medidas, financiadas con fondos públicos, han sido adoptadas con el propósito de reducir el impacto de las inundaciones en las áreas urbanas. Sin embargo, en muchas ocasiones, la eficacia de dichas contramedidas no ha sido adecuadamente evaluada en el contexto de eventos climáticos extremos, lo que ha resultado en pérdidas significativas, tanto en términos económicos como en términos de calidad de vida para la población afectada.

Mediante la implementación de modelos de aprendizaje automático, se puede llevar a cabo una evaluación más precisa y objetiva del daño provocado por inundaciones en distintas zonas urbanas de Guayaquil. Este enfoque permitirá determinar en qué dimensión las medidas de mitigación existentes han logrado reducir los daños y, en consecuencia, cuantificar su efectividad. Además, se podrán identificar aquellas áreas que siguen siendo vulnerables a las inundaciones, a pesar de las inversiones en contramedidas.

Uno de los aspectos más significativos de este proyecto es su capacidad para proporcionar

una visión económica más clara de la gestión de riesgo por inundaciones en Guayaquil. Al evaluar la eficacia de las medidas de mitigación, se podrá tomar decisiones más informadas sobre la asignación de recursos públicos, lo que conducirá a un mejor manejo de los fondos disponibles. Esto, a su vez, se traducirá en un mayor apoyo y protección para la ciudadanía en el período post inundaciones, permitiendo una respuesta más eficiente a las necesidades de las comunidades afectadas.

1.3 Objetivos

1.3.1 Objetivo General

Construir un modelo estadístico para la cuantificación de daño mediante técnicas de aprendizaje automático que permita una mejora para la mitigación del daño por inundaciones.

1.3.2 Objetivos Específicos

- Establecer un marco de predictores integrando aspectos socioeconómicos, geográficos y temporales, para optimizar la base de datos destinada a la modelización predictiva.
- Entrenar un modelo de aprendizaje automático Random Forest, utilizando el conjunto de predictores para predecir con alta precisión el ratio de daños materiales ocacionado por inundaciones en los sectores urbanos.
- Clasificar zonas urbanas de Guayaquil por su susceptibilidad a inundaciones mediante técnicas de clusterización, facilitando la gestión de riesgos y la planificación urbana.

1.4 Marco teórico

Aprendizaje automático

El aprendizaje automático o Machine Learning (ML), es una rama de la inteligencia artificial (IA) que tiene por propósito desarrollar métodos que permitan formar algoritmos que sean capaces de generalizar comportamientos y reconocer patrones desde una información entregada a manera de bosquejos. Así, este es un campo empleado para conocer patrones de forma automática e intuitiva en un conglomerado de datos sin requerir de una programación explicita. Para Mosavi et al., 2018 el aprendizaje automático presenta una amplia facilidad de resolución de problemas complejos con un coste computacional disminuido, aquello acompañado de un ágil entrenamiento, validación, prueba y evaluación, así como un alto rendimiento en comparación con los modelos físicos. Siguiendo el mismo criterio, Wagennar (2020) concluye que el aprendizaje automático representa una complejidad relativamente menor.

Por su lado, Caparrini, 2017 indica que desde una perspectiva básica, una de las misión del aprendizaje automático es intentar extraer conocimiento sobre ciertas propiedades no observadas de un objeto, fundamentándose en las propiedades que sí han sido investigadas o de las que sí se tiene registro, es decir, este tipo de aprendizaje implica predecir un comportamiento futuro a partir de la información que anteriormente ha sido proporcionada.

Ahora bien, dentro de las formas de aprendizaje automático, la mayoría de los problemas de aprendizaje estadísticos se engloban en una de las dos siguientes categorías: el aprendizaje supervisado, y el aprendizaje no supervisado.

Aprendizaje supervisado

A partir del aprendizaje supervisado se puede implementar lo aprendido con anterioridad a los nuevos datos percibidos, utilizando ejemplos previamente etiquetados para predecir acontecimientos futuros. Es decir, el fin es concretar un modelo que vincule la respuesta con los

predictores y pronostique con precisión la respuesta para futuras observaciones (entiéndase “predicción”) o comprenda mejor la relación entre la respuesta y los predictores (denominado “inferencia”). Así, desde el análisis de un grupo de datos de prueba conocido, el algoritmo de aprendizaje produce una función inferida para realizar predicciones sobre los resultados (Casella, 2006). Con esto, el sistema de aprendizaje supervisado puede disponer resultados para cualquier nueva información posterior de un periodo de entrenamiento suficiente. Consecuentemente, el algoritmo de aprendizaje también puede comparar su resultado con el resultado correcto y detectar errores para modificar así el modelo.

Cadena Lema (2020), explica que el aprendizaje supervisado necesita de un agente externo denominado tutor, esto significa que, el proceso de aprendizaje se ejecuta mediante un entrenamiento que está bajo control de un agente externo, el cual establece la respuesta que debería generar la red a partir de una entrada determinada. Respecto del párrafo precedente, para los autores Llanos y Romero (2013), el supervisor o tutor se encarga de controlar la salida de la red y en caso de que no coincida con la esperada se procede a modificar los pesos de las conexiones, con el objetivo de lograr que la salida alcanzada se aproxime a la esperada, para esto, se hace uso de un conjunto de datos previos , a los que se les denomina datos de entrenamiento; los cuales incluyen datos de entrada y valores resultantes, de todo ese conjunto de datos se utiliza el 70% como datos de entrenamiento y el 30% restante como datos de prueba para validar el correcto funcionamiento del algoritmo.

Cabe mencionar que varios métodos clásicos de aprendizaje estadístico que operan en el ámbito de aprendizaje supervisado, tal como la regresión lineal y la regresión logística, al igual que enfoques más modernos, como el boosting y las Máquinas de Vectores de Soporte.

Por otro lado, según MathWorks (2017), inmerso en el aprendizaje supervisado se encuentran las categorías de clasificación y regresión:

- Los algoritmos de clasificación son utilizados en valores de respuesta categóricos, donde los datos se pueden segmentar en clases definidas. Esta clase de algoritmo busca inducir un modelo que pueda prever un comportamiento una vez que se otorgan los valores de los atributos.
- Según Sucar (2018), los algoritmos de regresión son empleados para valores de respuesta continua, siendo su finalidad el inducir un modelo que pueda predecir el valor de una clase dados los valores de los atributos.

Cabe mencionar que el modelo de los árboles de decisión se encuentra tanto en los algoritmos de clasificación como en los algoritmos de regresión.

Aprendizaje no supervisado

Por otro lado, el aprendizaje no supervisado se refiere a una situación en la que cada observación tiene predictores sin valor de respuesta asociado que pueden utilizarse para supervisar el análisis. En este caso, para Casella (2006) el modelo trabaja, hasta cierto punto, a ciegas, debido a la falta de etiquetas proporcionadas al algoritmo de aprendizaje, por lo que éste debe encontrar por sí mismo la estructura de los datos de entrada.

Pues bien, acorde Zambrano, 2018, el aprendizaje no supervisado utiliza la función de agrupación, orillando al algoritmo a clasificar los datos por similitud y que de esta forma se creen los grupos, sin poseer la capacidad de delimitar cómo es cada individualidad de cada uno de los datos que comprenden los grupos. Con lo anterior, se sostiene que el aprendizaje no supervisado puede utilizarse para descubrir patrones ocultos en los datos y comprender las relaciones entre

las variables o las observaciones.

Una herramienta de aprendizaje estadístico, que se utiliza a menudo, es el análisis de conglomerados. Además, para Recuerdo de los Santos (2017), los tipos de algoritmos mayormente empleados en el aprendizaje no supervisado son: algoritmos de clustering, aprendizaje hebbiano, y aprendizaje competitivo y comparativo.

Árboles de decisión

En primer lugar, un árbol de decisión es un algoritmo de aprendizaje supervisado no paramétrico, que se emplea para tareas de clasificación y de regresión. Un árbol de decisión posee una estructura jerárquica, la cual cuenta con un nodo raíz, ramas, nodos internos y nodos hojas.

Es decir, comienza con un nodo raíz que no tiene ramas entrantes. Así, las ramas salientes del nodo raíz nutren los nodos internos, mismos que son llamados nodos de decisión. Ahora bien, ambos tipos de nodos desarrollan evaluaciones para conformar subconjuntos homogéneos, los cuales se indican mediante nodos hoja o nodos terminales. Los nodos hojas simbolizan la totalidad de resultados posibles en un conjunto de datos.

A pesar de las grandes ventajas que implica la utilización de árboles de decisión, existe una corriente que sostiene que los árboles de decisión deben sumar complejidad solo si es estrictamente necesario, puesto que la explicación más simple recurrentemente es la mejor. Esto va de la mano con que varios estudiosos coinciden en concluir que a medida que un árbol aumenta su tamaño es más complicado mantener su pureza, siendo común que cuando aquello ocurre, se produzca una fragmentación de datos que puede devengar en sobreajustes.

Ahora bien, los árboles de decisión (DT) engloban uno de los contribuidores en la

modelización predictiva con una extensa aplicación en la simulación de inundaciones. Este modelo usa un árbol de decisiones desde las ramas hasta los valores objetivo de las hojas. Cabe resaltar que en los árboles de clasificación, las variantes finales en un árbol de decisión conllevan un conjunto discreto de valores en los que las hojas representan etiquetas de clase y de ramas que representan conjunciones de etiquetas de características. De esta manera, cuando la variable objetivo de un árbol de decisión tiene valores continuos, tratándose de un conjunto de árboles, se le llama árbol de regresión.

Así, como se menciona en párrafos precedentes, los árboles de regresión y clasificación comparten ciertas similitudes así como conservan sus diferencias. Por lo que, al clasificarse los árboles de decisión como algoritmos rápidos, estos adquirieron popularidad en conjuntos para modelar y predecir inundaciones. De igual forma sucede con el método de bosques aleatorios, que es otro método de árboles de decisión famoso para la predicción de inundaciones, este método incluye varios árboles de predicción, pero cada árbol dispone de un conjunto de valores predictores de respuesta relacionada a un conjunto de valores independientes. (Bonafilia, Tellman, Anderson, & Issenberg, 2020) Por otro lado, el algoritmo de Hunt desarrollado en la década de 1960 para configurar el aprendizaje humano en psicología, compone el fundamento de muchos algoritmos de árboles de decisión conocidos, tales como:

• **ID3:** Su desarrollo es atribuido a Ross Quinlan, su abreviatura significa “Iterative Dichotomiser 3”. El algoritmo aprovecha la entropía y la ganancia de información como métricas para evaluar las divisiones de candidatos.

• **CART:** El término CART fue introducido por Leo Breiman y es una abreviatura de “árboles de clasificación y regresión” (classification and regression trees). El algoritmo con frecuencia usa

la impureza de Gini para identificar el atributo adecuado para la división. La impureza de Gini mide la frecuencia con la que se clasifica incorrectamente un tributo elegido al azar.

1.5 Estado del arte

Se realizaron revisiones exhaustivas de una amplia variedad de artículos científicos, siendo el primero de ellos titulado ”Application of machine learning for integrated flood risk assessment: Case study of Hurricane Harvey in Houston, Texas.” Este artículo, elaborado por (Bidadian, Maxwell, & Strager, 2023), se centra en un caso específico de daños por inundaciones derivados de un desastre natural.

Los autores proponen el algoritmo Random Forest, destacando su amplio uso en estudios relacionados con inundaciones y su idoneidad para extraer regresiones, correlaciones y abordar el problema del sobreajuste. El sobreajuste representa una preocupación seria al entrenar algoritmos potentes, ya que puede incorporar ruido o patrones irrelevantes al aprender de manera precisa los datos de entrenamiento, resultando en un rendimiento deficiente en datos no vistos.

En contraste con otros algoritmos, Random Forest aborda eficazmente este problema mediante la utilización de múltiples árboles de decisión, cada uno entrenado de manera independiente en un subconjunto aleatorio de los datos de entrenamiento. Además, los autores destacan la capacidad del algoritmo para estimar la contribución de cada variable utilizada en el modelo como un beneficio adicional. En resumen, los resultados presentados en este artículo respaldan a Random Forest como un algoritmo ideal para abordar los daños causados por inundaciones, sugiriendo que este método produce resultados óptimos al utilizar variables similares a las del caso de estudio.

Otro artículo revisado, titulado ”Flood Damage Analysis Using Machine Learning Techniques,” aborda la comparación de diferentes algoritmos de aprendizaje automático para medir el daño por inundaciones causado por fuertes lluvias en tres estados de la India: Bihar, Uttar Pradesh y Kerala. Escrito por Snehil y Ruchi Goel en 2020, este artículo se enfoca en evaluar la precisión de varios algoritmos de aprendizaje supervisado, identificando a Random Forest y K vecinos más cercanos como los que produjeron los mejores resultados.

Los autores señalan desafíos relacionados con las condiciones topológicas y climáticas variables en India, así como la disponibilidad limitada de datos precisos sobre los daños causados por inundaciones. Nuevamente, los resultados respaldan la eficacia de Random Forest como método de aprendizaje supervisado ideal para este análisis, abordando el problema del sobreajuste y siendo adecuado para problemas en los que el entrenamiento se realiza en datos similares a un árbol de decisiones, con la salida en forma de predicción media (damage ratio).

Como recomendación para futuros análisis, Snehil y Ruchi Goel sugieren la incorporación de variables más precisas para mejorar el rendimiento del modelo, como datos de lluvia más específicos de la región estudiada. Esto, sin duda, contribuiría aún más a la precisión del modelo en futuras investigaciones.

CAPÍTULO 2

2. METODOLOGÍA

Esta investigación se enfoca en la cuantificación del área promedio de daños en los sectores de Guayaquil, ocasionados específicamente por eventos de inundaciones. El estudio adopta una metodología cuantitativa y emplea un algoritmo de aprendizaje automático conocido como Random Forest, para predecir y cuantificar el área de daño. Este capítulo detalla el diseño del estudio, incluyendo la recolección y preparación de datos, el entrenamiento del modelo estadístico y las técnicas de validación implementadas para garantizar la precisión y aplicabilidad de los resultados obtenidos.

El diseño del estudio se fundamenta en un enfoque cuantitativo que permite la modelización y predicción de fenómenos complejos a través de datos numéricos y el uso de estadísticas. La elección del algoritmo Random Forest se justifica por su capacidad para manejar un gran número de predictores y su robustez frente a datos no lineales y de alta dimensionalidad. Además, el método de análisis de importancia de permutación condicional (CPI) se incorpora para identificar la relevancia de cada predictor en el modelo.

Guayaquil es conocida por su alta susceptibilidad a inundaciones, particularmente durante la temporada de lluvias y eventos climáticos como El Niño. El área de estudio abarca varios sectores seleccionados en base a la disponibilidad de registros de información suficientes para considerar cada sector como una observación de estudio. Se toman en cuenta variables

socioeconómicas, incluyendo la densidad poblacional, el ingreso promedio, la tasa de desempleo, la tasa de pobreza y el nivel predominante de educación para cada sector en el estudio. Las variables geográficas consideradas incluyen la elevación del terreno y la proximidad a cuerpos de agua. Además, se incluyen variables temporales que abarcan la estacionalidad y la cronología de eventos climáticos extremos.

La recolección de datos es un paso crítico en la investigación, ya que la calidad y la precisión del modelo dependen en gran medida de la integridad de los datos utilizados. Para este estudio, se han recopilado datos históricos de inundaciones de Centro Internacional del Pacífico para la Reducción del Riesgo de Desastres, abarcando registros por sectores desde 2012 hasta 2023. Los mapas de los sectores de Guayaquil se han generado utilizando el software ARGIS, y los datos de sensibilidad, exposición y vulnerabilidad de los sectores se han obtenido de la investigación (PDF investigación). Adicionalmente, se han empleado variables socioeconómicas recopiladas de la encuesta de empleo y desempleo en Guayaquil (ENEMDU) realizada por el INEC(2017). Se ha llevado a cabo un riguroso proceso de limpieza de datos para corregir inconsistencias, tratar valores faltantes y eliminar duplicados. Asimismo, se han implementado estrategias para validar la calidad de los datos, como la verificación cruzada con múltiples fuentes y la consulta de investigaciones científicas dedicadas al estudio de este fenómeno natural en Guayaquil.

2.1 Preparación de Datos para el Modelo

Además de alimentar los datos al algoritmo de Random Forest, es esencial seleccionar y transformar las variables predictoras para mejorar la eficacia del modelo. Se han aplicado técnicas como la normalización para las variables de tipo cuantitativo y la codificación a variables dummy

para variables categóricas, con el fin de preparar los datos con los que se entrenará el algoritmo de forma adecuada. La división de los datos en dos conjuntos, uno de entrenamiento y otro de prueba, se ha realizado en una proporción de 80% para entrenar y 20% para evaluar, dado que es crucial para aplicaciones en algoritmos de aprendizaje automático.

Conforme a los resultados del estudio de (Orozco y Arias, 2019), el algoritmo Random Forest es un método de ensamble que combina múltiples árboles de decisión para mejorar la precisión y controlar el sobreajuste." Cada árbol se construye utilizando una muestra aleatoria de los datos y un subconjunto de las variables predictoras, lo que aumenta la diversidad de las soluciones y la robustez del modelo. El proceso de entrenamiento del modelo implica ajustar los parámetros del Random Forest, que son: el número de árboles de decisión que se van a incluir en el bosque aleatorio, la profundidad máxima de los árboles, el número mínimo de muestras necesarias para dividir un nodo en dos y el número de variables predictoras que se consideran en cada división de un árbol. Para optimizar el rendimiento, se utiliza la validación cruzada para evaluar la generalización del modelo y evitar el sobreajuste.

Después de la etapa de entrenamiento, es esencial evaluar el rendimiento del algoritmo para verificar que las predicciones obtenidas son confiables y pertinentes. En modelos de regresión, se utilizan métricas específicas para evaluar el rendimiento del modelo. Entre las métricas más conocidas y utilizadas se encuentra el error cuadrático medio (MSE), que es una medida comúnmente utilizada para evaluar la precisión de los modelos. El MSE calcula el promedio de la suma de las diferencias cuadradas entre los valores reales y los valores pronosticados por el modelo. Un valor de MSE más bajo indica un mejor rendimiento del modelo. Otros factores importantes para tener en cuenta son el coeficiente de determinación R^2 y el error

absoluto medio (MAE). Estas métricas proporcionan una visión más completa del rendimiento del modelo y permiten una evaluación más precisa de la calidad de las predicciones realizadas por el algoritmo Random Forest en la estimación del área de daños causadas por inundaciones en los sectores de Guayaquil.

La técnica de análisis de importancia de permutación condicional (CPI) juega un papel crucial en la cuantificación de la contribución de cada variable predictora al poder predictivo del modelo. Este análisis se lleva a cabo mediante la permutación de los valores de una variable, manteniendo constantes las demás, para observar el impacto en el rendimiento del algoritmo. Las variables que presentan una mayor importancia son aquellas cuyo intercambio de valores conduce a una disminución significativa en la precisión del modelo, lo que sugiere que son predictores fundamentales en la estimación del daño por metros cuadrados en los sectores de Guayaquil. Este análisis permite una evaluación más precisa de la relevancia de cada variable predictora en el algoritmo.

La interpretación de los resultados obtenidos a través del algoritmo Random Forest y del análisis de permutación condicional (CPI) es esencial para extraer conclusiones significativas y aplicables. Se examina cómo las variables de mayor importancia influyen en la predicción de los metros cuadrados de daño en los distintos sectores de Guayaquil, y se exploran las implicaciones potenciales para la planificación urbana y la gestión de riesgos. Se abordan tanto los hallazgos estadísticamente significativos como las limitaciones del estudio, incluyendo posibles sesgos en los datos y la necesidad de futuras investigaciones para validar y ampliar los resultados. Esta interpretación y discusión permiten una comprensión más profunda de las fortalezas y debilidades del modelo, y guían la formulación de recomendaciones para la implementación y mejora continua.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

En el presente capítulo, se aborda el análisis y los resultados obtenidos en la investigación, centrados en la identificación y cuantificación de predictores significativos para la evaluación del riesgo ambiental y geomorfológico en Guayaquil.

3.1 Identificación de predictores

Los predictores más significativos, obtenidos de fuentes primarias como el Instituto Nacional de Estadísticas y Censos (INEC) y entidades especializadas en Riesgos y Desastres, se han identificado como variables clave para cuantificar factores asociados con el riesgo ambiental y geomorfológico, así como elementos vinculados a la exposición, vulnerabilidad y susceptibilidad en los diversos sectores de Guayaquil. Estos predictores incluyen: el área promedio en metros cuadrados, la probabilidad de afectación, el número de viviendas comprometidas, el área promedio de impacto, la tasa de superficie agrícola por metro cuadrado, la edad media de los habitantes, la tasa de analfabetismo, la tasa de acceso a servicios de agua potable, la tasa de servicios de recolección de residuos, la tasa de cobertura de alcantarillado, el grado de urbanización, la incidencia de pobreza, la prevalencia de desempleo, el ingreso medio mensual y el nivel educativo. Estas variables han sido meticulosamente cuantificadas a nivel sectorial.

Figura 3.1.

Nota: *Elaborado en ArcGis, representa el Área promedio de daño en metros cuadrados a nivel sectorial.*

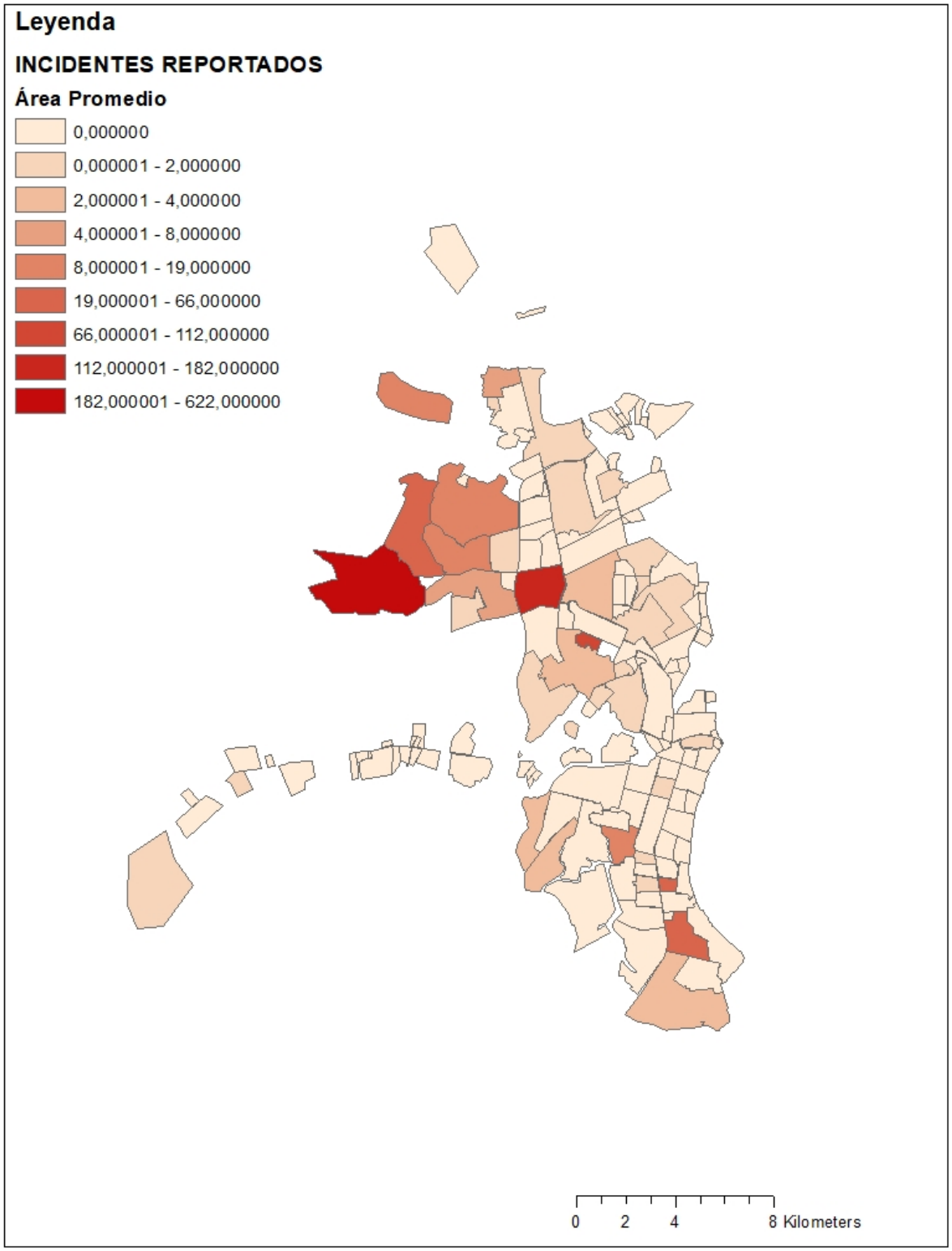
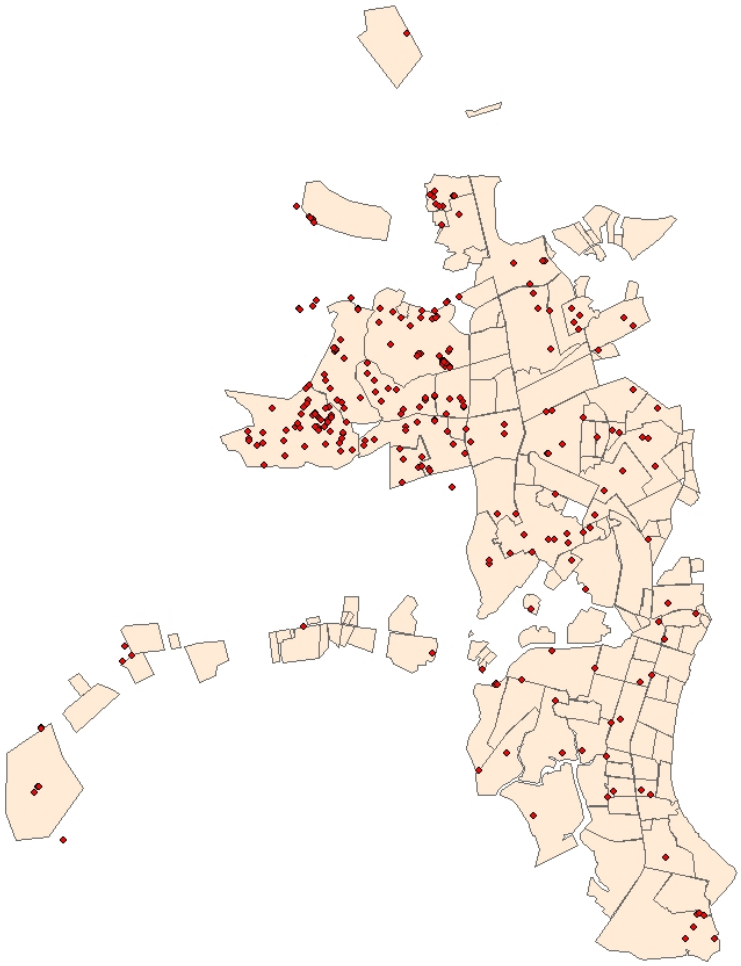


Figura 3.2.
Nota: *Elaborado en ArcGis, representa los registros históricos de eventos de inundaciones a nivel sectorial (2012-2023).*



El gráfico muestra los registros históricos de eventos de inundaciones que fueron facilitados por el CIP-RRD. En él se encontraron 455 eventos de inundaciones dispersos por sectores en Guayaquil desde 2012 hasta 2023,. Con estos registros, se ploteó un mapa de sectores para conocer la distribución de eventos de inundaciones por sector y se calcularon las probabilidades de sufrir este tipo de eventos naturales para cada sector.

3.1.1 Análisis IPC

Tras la implementación del método de importancia de permutación condicional, se determinan cuáles son las variables de mayor relevancia en este estudio: la proporción de viviendas con acceso a servicios de recolección de residuos en los sectores evaluados, la tasa de analfabetismo, la tasa de provisión de agua potable, la tasa de superficie dedicada a la agricultura, la tasa de cobertura de alcantarillado, el número de viviendas impactadas, la incidencia de pobreza, la probabilidad de afectación, el nivel de urbanización y la tasa de desempleo registrado en cada uno de los sectores de Guayaquil considerados en esta investigación.

3.1.2 Costos de implementación y viabilidad

Este proyecto de investigación presenta una alta viabilidad económica, dado que tras la recopilación exhaustiva de datos concernientes a factores geográficos, exposición, vulnerabilidad y condiciones económicas, se procedió a la implementación del modelo de aprendizaje automático Random Forest. Este modelo se utilizó para estimar el área promedio afectada en cada sector, basándose en los predictores más relevantes identificados mediante el método de importancia de permutación condicional.

3.2 Estimación de hiperparámetros

En concordancia con la literatura científica pertinente, se optó por entrenar exclusivamente el algoritmo Random Forest. Según las conclusiones de Behrang Bidadian, Aaron E. Maxwell y Michael P. Strager, este algoritmo se destaca como la opción preferente para la modelización de eventos naturales tales como inundaciones. Para el entrenamiento del modelo, se seleccionaron

todas las variables asociadas con las consecuencias y antecedentes de inundaciones. Se asignó el 80% de los datos al conjunto de entrenamiento y el 20% restante al conjunto de evaluación.

Con el objetivo de optimizar los hiperparámetros del algoritmo Random Forest, se diseñó una matriz de parámetros. Mediante la técnica de búsqueda exhaustiva (grid search), se exploró cada combinación posible de parámetros para evaluar el desempeño del modelo y seleccionar la configuración óptima que minimiza la función de pérdida. Los parámetros óptimos resultaron ser: $mtry = 7$, que representa el número de variables seleccionadas al azar en la bifurcación de un nodo interno; $trees = 975$, que indica el número de árboles en el ensamblaje; y $min = 11$, que corresponde al número mínimo de observaciones requeridas en un nodo hoja.

Figura 3.3.
Importancia de la permutación condicional

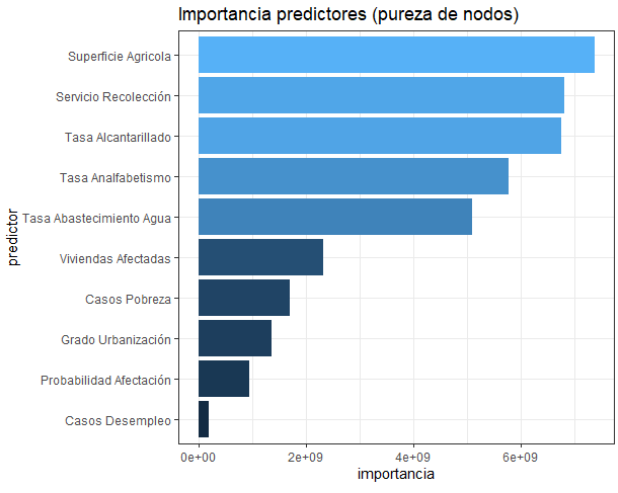


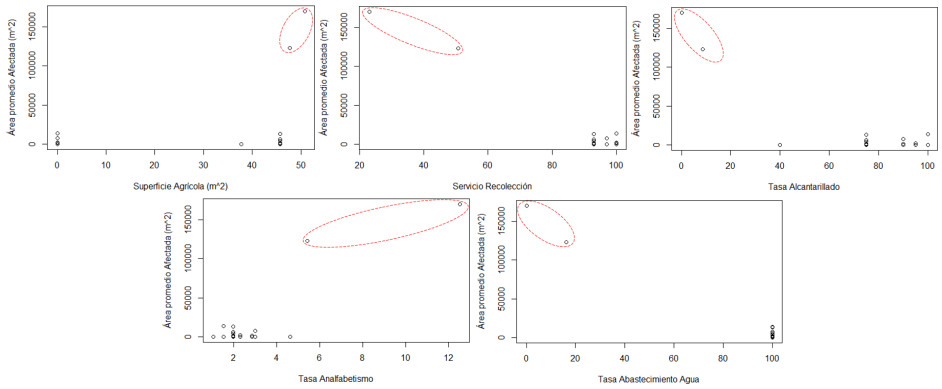
Figura 3.4.
Resultados del modelo de Random Forest

Métrica	Estimación
Tipo	Regresión
Número De Árboles	975
Tamaño De Muestra	24
Número De Variables Independientes	10
Mtry	7
Tamaño Del Nodo Objetivo	11
Modo De Importancia De Variable	ninguno
Regla De División	varianza
Error De Predicción OOB	746603522
R^2	0.5698229

3.3 Efectos de los predictores

Además, se evaluó el efecto que ejerce la presencia de una elevada tasa de superficie agrícola, bajos porcentajes de servicio de recolección de residuos en el sector, niveles bajos o nulos de servicio de red de alcantarillado, tasas de analfabetismo significativas y la falta de servicios adecuados de abastecimiento de agua sobre el área promedio de daños en metros cuadrados en un sector de Guayaquil.

Figura 3.5.
Efecto de los predictores significativos



Los sectores con mayores áreas de superficie agrícola, un menor porcentaje de servicios de recolección, un menor porcentaje de alcantarillado, una mayor tasa de analfabetismo y una menor tasa de abastecimiento de agua tienden a experimentar una mayor área promedio de daños en eventos de inundaciones.

A partir de este análisis, ahora se tiene una comprensión profunda del impacto potencial en el área promedio dañada de un sector si no se satisfacen los criterios e indicadores mínimos en términos de exposición y vulnerabilidad. Estos indicadores pueden ser inferidos a partir de las variables significativas identificadas por el algoritmo.

3.4 Evaluación del algoritmo

Para la evaluación del rendimiento del algoritmo, se seleccionaron aleatoriamente seis sectores representativos: Febres Cordero, Huancavilca Sur, Estero Salado, Cristo del Consuelo, Ceibos y Mapasingue.

Tabla 3.1.
Predicciones del área promedio de daño por sectores, generadas mediante el algoritmo Random Forest.

Sector	Predicciones (m^2)	Valor Real (m^2)
Febres Cordero	46.04	0.00
Huancavilca Sur	10530.18	13533.40
Estero Salado	4105.40	1215.98
Cristo del Consuelo	5557.13	1683.93
Ceibos	51.62	0.00
Mapasingue	51.70	0.00

Como se puede apreciar en la tabla presentada, los pronósticos generados por el modelo muestran una notable concordancia con los datos reales, reflejando de manera precisa la extensión del daño en metros cuadrados, especialmente en aquellos sectores donde no se reportaron daños.

Durante la estación invernal en Guayaquil, se identifican tres sectores particularmente susceptibles a inundaciones: Chongón, Martha de Roldós y Monte Sinaí. Los pronósticos generados por el algoritmo entrenado para estos tres sectores críticos demostraron una aproximación significativa a los valores empíricamente registrados. Es destacable el caso del sector Martha de Roldós, que habitualmente experimenta inundaciones de magnitud considerable, resultando en la interrupción de las actividades económicas locales, un impacto negativo en la economía del sector y un incremento en la afluencia de personas a los centros de salud adyacentes. El pronóstico para este sector fue 1.35 veces el valor real observado, lo que sugiere una alta precisión predictiva, dada la magnitud de los daños que este sector incurre tras eventos de inundación.

Tabla 3.2.
Predicciones del área promedio de daños en sectores de interés

Sector	Predicciones (m^2)	Valor Real (m^2)
Chongón	125364.70	170192.17
Martha de Roldós	17613.51	13084.79
Monte Sinaí	103211.15	122855.03

El algoritmo es capaz de explicar aproximadamente el 57% de la variabilidad en el área promedio dañada, medida en metros cuadrados por sector, lo cual indica un rendimiento considerablemente aceptable para la modelización de los daños ocasionados por eventos naturales como las inundaciones.

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

En el segmento correspondiente a las conclusiones y recomendaciones, se lleva a cabo un análisis exhaustivo de los resultados obtenidos mediante la aplicación de modelos de aprendizaje automático para la predicción de ratios de daños ocasionados por inundaciones. Se abordan en detalle las métricas fundamentales, enfatizando la precisión alcanzada y su correlación con los objetivos predefinidos al inicio del estudio. Se destacan las fortalezas inherentes al proyecto, haciendo hincapié en la eficacia del modelo Random Forest como el más apropiado para este tipo de análisis, así como en la minuciosa recopilación de datos realizada. Aunque se identifican áreas susceptibles de mejora, se subraya la contribución significativa del estudio en la identificación de los sectores más vulnerables de la ciudad de Guayaquil ante posibles daños derivados de inundaciones.

En última instancia, se proporcionan recomendaciones pormenorizadas para investigaciones futuras, abordando aspectos tales como la expansión del conjunto de datos, la optimización de los modelos, la evaluación continua y las posibles aplicaciones prácticas. Estas sugerencias tienen como objetivo potenciar aún más el impacto del presente trabajo en el ámbito de estudio, proyectando su influencia hacia futuras investigaciones y aplicaciones prácticas.

4.1 Conclusiones

El enfoque principal de la investigación ha sido la implementación de un modelo de aprendizaje automático destinado a la predicción de tasas de daño causadas por inundaciones. Mediante el análisis de las métricas presentadas en los resultados, se puede evaluar en qué medida se han alcanzado los objetivos planteados al comienzo del proyecto.

El objetivo general de este estudio consistía en desarrollar un modelo estadístico mediante técnicas de aprendizaje automático para cuantificar en m^2 el área afectada por inundaciones, con el objetivo de mejorar las estrategias de mitigación. La obtención de un error cuadrático medio de 0.57, combinado con la creación de un mapa de calor en ArcMap 10.7, proporciona un enfoque positivo en cuanto a las predicciones y su concordancia con las etiquetas reales. Este resultado es crucial para la precisa identificación de áreas vulnerables a daños por inundaciones, cumpliendo así de manera significativa con nuestro objetivo general propuesto.

- La preparación y compilación de un conjunto de datos que incorpora aspectos socioeconómicos, geográficos y temporales con el fin de optimizar la base de datos destinada a la modelización predictiva se llevó a cabo mediante la integración de registros de inspecciones por daños ocasionados por inundaciones proporcionados por SeguraEP, registros catastrales de la totalidad de la ciudad de Guayaquil, y datos de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) de 2017, que contenían información sociodemográfica específica de las unidades de investigación y los sectores de la ciudad.

- Al emplear el conjunto de datos previamente construido, se logró entrenar un modelo

Random Forest con una precisión final del 57%. Aunque es cierto que aún puede mejorarse, el modelo cumple con la capacidad de realizar estimaciones en metros cuadrados del área total afectada por sector debido a inundaciones.

- Mediante el empleo del software ArcMap, se logró una diferenciación visual de las áreas más susceptibles a sufrir daños por inundaciones. Entre estas áreas se destacan Monte Sinaí, Chongón y Florida, las cuales constituyen las tres zonas más vulnerables de la ciudad. Este hallazgo subraya la necesidad de implementar medidas preventivas específicas en dichas áreas con el fin de mitigar los daños ocasionados por eventos de inundación.

4.1.1 Limitaciones

La falta de precisión del modelo se atribuye principalmente a la ausencia de información sociodemográfica actualizada disponible por la unidad geográfica empleada en este estudio. Este problema surge debido al cambio en la categorización de sectores en las nuevas versiones del ENEMDU a partir de 2018, donde la unidad primaria de muestreo ahora se define como una combinación lineal utilizando conglomerados en lugar de sectores. Además, el método de muestreo utilizado para obtener datos en el ENEMDU se basa en muestras aleatorias de sectores censales, lo que significa que no se dispone de información para todos los sectores. Esta limitación resulta en una reducción significativa en la cantidad de registros disponibles para el entrenamiento del modelo, impactando directamente en la precisión de las predicciones.

4.1.2 Implicaciones del trabajo

La identificación de zonas vulnerables a daños por inundaciones posibilita dirigir la atención hacia áreas que demandan asistencia urgente, lo cual optimiza tanto el tiempo como

los recursos disponibles para prevenir posibles daños futuros. Además, esta acción contribuye significativamente a mejorar la calidad de vida de los habitantes de dichos sectores, permitiendo una intervención más eficaz y enfocada en la mitigación de riesgos y la protección de la comunidad.

4.1.3 Trabajos futuros

Se identifican oportunidades claras para mejorar tanto la precisión como la eficiencia del modelo. En futuras investigaciones, sería valioso dirigir esfuerzos hacia la optimización y el afinamiento del modelo, así como considerar un cambio en la unidad de investigación, no limitándose únicamente a sectores de la ciudad de Guayaquil, sino ampliando el alcance a más sectores a nivel general de la provincia. Esta expansión podría aprovechar la mayor cantidad de registros disponibles. En resumen, el trabajo realizado representa un avance significativo en la aplicación de modelos de aprendizaje automático para mejorar la calidad de vida de los habitantes de Guayaquil. Aunque persisten desafíos por abordar, los objetivos inicialmente planteados en el proyecto se han enfrentado de manera efectiva, estableciendo así las bases para futuras investigaciones y aplicaciones prácticas en este campo.

4.2 Recomendaciones

La experiencia obtenida en este trabajo y en investigaciones previas o relacionadas con la evaluación del área afectada por inundaciones en la ciudad de Guayaquil destaca la dificultad inherente a este tipo de estudios, ya que demandan la colaboración de entidades involucradas en la obtención y mantenimiento de una extensa cantidad de datos. A pesar de estos desafíos, se sostiene que la rentabilidad potencial de tales estudios es significativa, dado que directamente

contribuyen a la reducción del riesgo. La consideración de realizar una comparación con otros modelos es una perspectiva valiosa para fortalecer la robustez y generalización de los resultados obtenidos en la investigación. Aunque el Random Forest ha demostrado ser efectivo, la evaluación de su desempeño frente a otros modelos podría ofrecer insights adicionales sobre la idoneidad de diferentes enfoques para el contexto específico de la ciudad de Guayaquil. Se sugiere llevar a cabo un análisis comparativo con otros algoritmos de aprendizaje automático relevantes para estudios de evaluación de daños por inundaciones. El análisis de la muestra ha facultado la elaboración y predicción de escenarios de daño con un grado de completitud razonable. No obstante, la realización de encuestas en los sectores no incorporados en la muestra de viviendas obtenidas del ENEMDU podría significativamente mejorar los resultados obtenidos. Esto se debe a que algunos sectores no han podido ser analizados con una precisión elevada debido a la carencia de variables sociodemográficas en la muestra actual.

4.2.1 Otras recomendaciones

Este trabajo se materializó gracias a la colaboración fundamental de la Centro Internacional del Pacífico para la Reducción del Riesgo de Desastres (CIP-RRD) y la Municipalidad de Guayaquil, quienes facilitaron el acceso a los datos necesarios. Los resultados obtenidos en este estudio estarán disponibles públicamente, lo que permitirá el desarrollo de planes de emergencia ante inundaciones. Subrayamos la importancia de involucrar a instituciones relacionadas con la temática en este tipo de estudios. Es esencial respaldar la participación activa de las empresas en investigaciones de riesgo, haciéndoles comprender que los mayores beneficiarios son ellos mismos. Al tener un conocimiento más preciso del riesgo y su dimensión, las empresas pueden tomar decisiones informadas para mitigar dicho riesgo. Este

enfoque contribuirá de manera significativa a pequeños avances en la mejora del bienestar y la reducción de uno de los diversos riesgos a los que Guayaquil está expuesto. Estas recomendaciones persiguen el propósito de fortalecer el impacto y la utilidad del proyecto integrador, además de establecer cimientos para investigaciones subsiguientes en el mismo ámbito.

BIBLIOGRAFÍA

Aragón-Durand, F. (2014). *Inundaciones en zonas urbanas de cuencas en América Latina*. Soluciones Prácticas, Lima,Perú, 1 edition.

Caparrini, F. S. (2017). *Introducción al Aprendizaje Automático*. https://www.cs.us.es/ fsancho/?e=75.

Merz, B., Kreibich, H., and Lall, U. (2013). Multi-variate flood damage assessment: a tree-based data-mining approach. *Natural Hazards and Earth System Sciences*, 13(1):53–64.

Mosavi, A., Ozturk, P., and Chau, K.-w. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11).

Nuñez, J. R. and Bastidas, J. G. (2021). Vulnerabilidad a inundaciones en la ciudad de guayaquil. *https://storymaps.arcgis.com/stories/f3f9478a16b343fbb50c7695fabfd9e4*, 1:1.

Zambrano, M. (2018). Métodos estadísticos de machine learning aplicados en el estudio de la accesibilidad web: una revisión de la literatura. *Minerva*, 3:150–157.