



POSGRADOS

MAESTRÍA EN TRANSFORMACIÓN DIGITAL E INNOVACIÓN

RPC-SO-13-NO.734-2024

OPCIÓN DE TITULACIÓN:

ARTÍCULOS PROFESIONALES DE ALTO NIVEL

TEMA:

APLICACIÓN DE RANDOM FOREST PARA LA
PREDICCIÓN DEL TIEMPO DE APROBACIÓN
EN DE TRÁMITES DE CONSTRUCCIÓN
DESDE EL GAD MUNICIPAL DEL CANTÓN
CUENCA

AUTOR:

ESTEBAN PAÚL GARZÓN CAMPOS

DIRECTOR:

FABIÁN LEONARDO CUESTA ASTUDILLO

CUENCA – ECUADOR
2026

Autor:**Esteban Paúl Garzón Campos**

Ingeniero de Sistemas.

Magister en Contabilidad y Auditoría.

Candidato a Magíster en Transformación Digital e Innovación por la Universidad Politécnica Salesiana – Sede Cuenca.

egarzonc@est.ups.edu.ec

Dirigido por:**Fabián Leonardo Cuesta Astudillo**

Ingeniero Informático.

Magister en Ciencias de la Computación.

Magister en Gerencia Empresarial.

Máster Universitario en Neuromarketing.

Magister e Auditoria Integral.

Máster Universitario en Inteligencia Artificial.

fcuesta@ups.edu.ec

Todos los derechos reservados.

Queda prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra para fines comerciales, sin contar con autorización de los titulares de propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual. Se permite la libre difusión de este texto con fines académicos investigativos por cualquier medio, con la debida notificación a los autores.

DERECHOS RESERVADOS

2026 © Universidad Politécnica Salesiana.

CUENCA – ECUADOR – SUDAMÉRICA

ESTEBAN PAÚL GARZÓN CAMPOS

Aplicación de Random Forest para la predicción del tiempo de aprobación en de trámites de construcción desde el GAD municipal del cantón Cuenca

DEDICATORIA

A mi Esposa y mis Hijos cuyo tiempo para compartir se ha visto sacrificado para llevar adelante este objetivo.

A mi Mamá y Hermanos, a mis Suegros, Cuñados y Sobrinos y toda la familia que comparten conmigo la alegría de alcanzar un nuevo reto.

AGRADECIMIENTO

A la Universidad Politécnica Salesiana por poner a disposición de la comunidad estas oportunidades de aprendizaje y crecimiento con esa visión de progreso para nuestra Ciudad y el País.

A mi Tutor y Director de la Maestría en Transformación Digital e Innovación por el apoyo en la realización de esta investigación y por la excelente gestión para llevar adelante esta maestría.

Al GAD Municipal del cantón Cuenca por brindar las facilidades para la realización de esta investigación.

Tabla de Contenido

Resumen	7
Abstract	8
Introducción	9
Determinación del Problema.....	14
Marco teórico referencial.....	16
3.1 Machine Learning en trámites administrativos.....	16
3.2 Random Forest como modelo de referencia.....	17
3.3 Hiperparámetros en Random Forest para pronósticos cuantitativos.....	18
3.4 Métricas de evaluación.....	19
Materiales y metodología.....	20
4.1 Fundamentación del estado del arte	20
4.2 Recolección y preparación de los datos	21
4.3 Implementación y evaluación del modelo Random Forest.....	22
4.4 Variables de estudio	23
Escenario 1.....	24
Escenario 2.....	26
Escenario 3.....	27
4.5 Pregunta de investigación	28
Resultados y discusión.....	30
5.1 Entrenamiento y resultados con el escenario 1	30
5.2 Entrenamiento y resultados con el escenario 2	34
5.3 Entrenamiento y resultados con el escenario 3	37
5.4 Discusión.....	40
Conclusiones.....	41
Referencias (Bibliografía)	42

Aplicación de Random Forest para la Predicción del Tiempo de Aprobación de Trámites de Construcción del Gad Municipal del Cantón Cuenca

Autor(es):

Esteban Paúl Garzón Campos

Resumen

Este artículo propone utilizar como una fuente secundaria, la información histórica de los tramites de construcción aprobados por el GAD Municipal del cantón Cuenca en la aplicación del modelo de machine learning Random Forest para predecir el tiempo de aprobación de los proyectos.

Se partió con un conjunto de datos inicial de 47.485 registros, luego del tratamiento de los datos, se obtuvo un conjunto de datos final con 9.841 registros que se utilizaron para la aplicación del modelo.

Con el conjunto de datos depurado, se establecieron tres escenarios, el primero con un conjunto de variables independientes cualitativas y cuantitativas, el segundo con variables únicamente cualitativas y el tercero únicamente con variables cuantitativas. Para la predicción se implementó el algoritmo de Random Forest en Python utilizando el modelo RandomForestRegressor y se ajustaron los hiperparámetros clave mediante el método GridSearchCV de Python.

Como parte del análisis se determinaron las variables predictoras de mayor importancia en cada uno de los escenarios y se obtuvieron métricas de evaluación como: Error absoluto medio (MAE), Error cuadrático medio (MSE), Raíz del error cuadrático medio (RMSE), Coeficiente de determinación (R^2).

En este artículo se concluye que la aplicación del modelo de aprendizaje automático Random Forest, en base a variables relacionadas con las características del proyecto, su ubicación y condiciones administrativas, no permite predecir con alta precisión el tiempo de aprobación de los trámites de construcción en el GAD Municipal del cantón Cuenca, considerando los valores obtenidos en las métricas de evaluación utilizadas.

Palabras clave:

Machine Learning, Random Forest, Procesos Administrativos, Estimación de Tiempos.

Abstract

This article proposes using historical information on construction procedures approved by the Municipal GAD of the Cuenca canton as a secondary source in the application of the Random Forest machine learning model to predict project approval times.

It started with an initial dataset of 47,485 records; after data processing, a final dataset of 9,841 records was obtained, which was used for the model application.

With the cleaned dataset, three scenarios were established: the first with a set of qualitative and quantitative independent variables, the second with only qualitative variables, and the third with only quantitative variables. For prediction, the Random Forest algorithm was implemented in Python using the RandomForestRegressor model, and key hyperparameters were tuned using Python's GridSearchCV method.

As part of the analysis, the most important predictor variables were determined for each scenario, and evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) were obtained.

This article concludes that the application of the Random Forest machine learning model, based on variables related to project characteristics, location, and administrative conditions, does not allow for high-accuracy prediction of approval times for construction procedures in the Municipal GAD of the Cuenca canton, considering the values obtained in the evaluation metrics used.

Keywords:

Machine learning, Random Forest, Administrative Processes, Time Estimation.

Introducción

La Ley Orgánica para la Optimización y Eficiencia de Trámites Administrativos, publicada en el Registro Oficial Suplemento No. 353 en los artículos 4, 5, 7 y 10, establece un marco normativo orientado a simplificar los procesos administrativos mediante el uso de tecnologías de información, reducción de requisitos innecesarios y establecimiento de plazos máximos para la resolución de trámites. Esta ley contempla principios como la celeridad, interoperabilidad, simplificación y control posterior, así como, la obligación de las entidades públicas de utilizar medios digitales para la gestión de trámites y la comunicación con los administrados (Asamblea Nacional del Ecuador, 2018).

La Ley Orgánica para la Optimización y Eficiencia de Trámites Administrativos (Asamblea Nacional del Ecuador, 2018), en su artículo 2 determina como ámbito de la misma a las entidades que integran el régimen autónomo descentralizado; además, en el Ecuador los Gobiernos Autónomos Descentralizados (GAD) Municipales tienen competencias establecidas en otras normas y algunas de ellas de manera exclusiva; un ejemplo de ello, es el caso de la gestión de trámites de construcción, conforme lo establece el artículo 264 de la Constitución de la República del Ecuador (Asamblea Constituyente del Ecuador, 2008) y el artículo 55 del Código Orgánico de Organización Territorial, Autonomía y Descentralización (COOTAD), al indicar que es atribución del nivel municipal ejercer control sobre el uso y ocupación del suelo (Asamblea Nacional del Ecuador, 2010).

Una de las funciones de los gobiernos autónomos descentralizados municipales que establece el artículo 54 del Código Orgánico de Organización Territorial, Autonomía y Descentralización (COOTAD) es, establecer el régimen de uso del suelo y urbanístico, para lo cual determinará las condiciones de urbanización, parcelación, lotización, división o cualquier otra forma de fraccionamiento de conformidad con la planificación cantonal, asegurando porcentajes para zonas verdes y áreas comunales.

Es importante anotar que, la aprobación de trámites de construcción se ha convertido en un componente crítico del desarrollo urbano, dado que su agilidad o demora incide directamente en los procesos de inversión, planificación y crecimiento habitacional. No obstante, múltiples factores como la carga administrativa y la heterogeneidad de los tipos de trámite han generado variabilidad en los tiempos de aprobación, lo cual se traduce en incertidumbre para los ciudadanos y un uso ineficiente de los recursos institucionales (Cuenca, 2024).

Esta realidad evidencia la necesidad de herramientas que permitan comprender, modelar y anticipar los tiempos de aprobación, lo que permitiría mejorar la planificación interna, atender de forma prioritaria los casos complejos y generar expectativas realistas entre los ciudadanos. En este sentido, los avances en transformación digital e innovación tecnológica ofrecen una oportunidad estratégica para repensar y optimizar los procesos administrativos, haciendo uso de técnicas emergentes como el aprendizaje automático (machine learning).

Los modelos de aprendizaje automático (machine learning) se presentan como alternativas importantes para generar predicciones robustas en entornos administrativos complejos. En particular, algoritmos como Random Forest, Gradient Boosting, XGBoost y redes neuronales se han aplicado con éxito en predicciones de duración de trámites, productividad, tiempos de ejecución y análisis de eficiencia institucional (Bijalwan et al., 2024).

Diversos estudios destacan la aplicabilidad del machine learning en entornos administrativos:

- Tax et al. (2017) integraron LSTM con logs de eventos para predecir la duración remanente de procesos administrativos, logrando alta precisión incluso con secuencias incompletas.
- Polato et al. (2018) demostraron que los modelos híbridos (LSTM + clustering) permiten mejorar la predicción de duración de trámites al adaptarse dinámicamente al tipo de proceso.

- Senderovich et al. (2017) utilizaron técnicas de minería de colas para estimar retrasos en procesos administrativos.
- Suriadi et al. (2017) analizaron trazas de eventos con algoritmos supervisados y no supervisados para estimar tiempos de aprobación.
- Verenich et al. (2019) desarrollaron un benchmark para comparar técnicas de predicción de tiempo restante en procesos administrativos.
- Breuker et al. (2016) aplicaron autómatas probabilísticos para predecir la siguiente actividad en trámites administrativos.
- Senderovich et al. (2015) aplicaron teoría de colas para modelar cuellos de botella en instituciones públicas.
- Folino et al. (2014) integraron modelos de clustering y regresión para predecir tiempos de ejecución en procesos complejos.
- Rogge-Solti y Weske (2013) utilizaron redes de Petri estocásticas para estimar tiempo restante en procesos burocráticos.
- Gómez et al. (2020) utilizaron modelos de ML para predecir la duración de casos en el sistema judicial colombiano.
- Sun y Medaglia (2019) mapearon los desafíos y oportunidades del uso de inteligencia artificial en el sector público, incluyendo la predicción de tiempos de atención y eficiencia en la prestación de servicios administrativos en entornos gubernamentales.

Estos estudios evidencian que la predicción del tiempo de aprobación de trámite no solo es técnicamente viable, sino que contribuye directamente a la eficiencia institucional y mejora la experiencia ciudadana. La integración de estas tecnologías en los GAD permitiría priorizar casos complejos, anticipar demoras, distribuir mejor la carga de trabajo y establecer tiempos estimados de respuesta con mayor confianza.

En este contexto, las técnicas de aprendizaje automático han demostrado ser efectivas para el análisis predictivo en procesos administrativos y logísticos. El algoritmo Random Forest, en particular, se destaca por su capacidad para manejar grandes volúmenes de datos heterogéneos, identificar relaciones no lineales entre

variables y ofrecer resultados de alta precisión en tareas de regresión y clasificación. Este modelo de Random Forest construye múltiples árboles de decisión a partir de subconjuntos aleatorios de datos y características, y luego consolida los resultados mediante votación o promedio, reduciendo así el riesgo de sobreajuste y mejorando la robustez general del modelo (Yamashiro y Nonaka, Estimation of processing time using machine learning and real factory data for optimization of parallel machine scheduling problem. *Operations Research Perspectives*, 2021).

Aplicaciones previas de este algoritmo en contextos públicos y urbanos han mostrado resultados alentadores. Por ejemplo, Navarrin et al. (2017) emplearon redes neuronales tipo LSTM para predecir el tiempo restante de procesos de negocio en entornos corporativos, incorporando atributos contextuales del proceso como tipo de tarea, hora de inicio y recursos involucrados. De forma análoga, estudios como el de Bijalwan et al. (2024) han utilizado modelos de Random Forest para predecir la productividad de empleados municipales en India, encontrando que este tipo de técnicas permiten priorizar decisiones de gestión de talento con base en variables históricas de rendimiento.

Además, investigaciones como las de Battisti et al. (2023) han demostrado la viabilidad de aplicar procesos automáticos para trámites relacionados con la construcción, utilizando modelos digitales como Building Information Modeling (BIM) y algoritmos computacionales para realizar validaciones normativas. Aunque estos estudios se han desarrollado en otros países, aportan referentes valiosos para la digitalización del proceso de revisión y aprobación de planos en municipios latinoamericanos.

El presente proyecto utiliza la información histórica de los tramites de construcción aprobados por el GAD Municipal del cantón Cuenca en la aplicación del modelo de machine learning Random Forest para predecir el tiempo de aprobación de los proyectos. Se realizó la validación de los resultados para medir su desempeño y determinar la fiabilidad del modelo y su utilidad en la priorización de casos.

Para el efecto, en este proyecto se planteó como objetivo “Predecir el tiempo de aprobación en trámites de construcción del GAD Municipal del cantón Cuenca mediante la aplicación de Random Forest, con base en las características del proyecto, su ubicación, y condiciones administrativas como una herramienta para la optimización de los procesos.”

Adicionalmente, se establecieron los siguientes objetivos específicos:

1. Fundamentar el estado del arte del uso de machine learning en trámites administrativos, enfocándose en estudios sobre predicción de tiempos y tramitología pública.
2. Generar el conjunto de datos históricos de trámites de construcción del GAD Municipal del cantón Cuenca que sea representativo y enriquecido y facilite el entrenamiento del modelo Random Forest, recopilando y transformando variables técnicas, geográficas y administrativas.
3. Implementar el modelo de Random Forest para predecir los tiempos de aprobación de trámites de construcción utilizando validación cruzada y métricas como MAE, RMSE y R^2 para medir su desempeño.

Determinación del Problema

Desde el punto de vista normativo, la Ley Orgánica para la Optimización y Eficiencia de Trámites Administrativos (Registro Oficial Suplemento No. 353, 2018) establece obligaciones claras para las entidades públicas, incluyendo el uso de tecnologías digitales, la interoperabilidad entre sistemas y el cumplimiento de plazos máximos para resolver trámites. Esta ley representa un mandato legal para innovar en los procesos administrativos, y al mismo tiempo abre una oportunidad para que los gobiernos locales adopten tecnologías de vanguardia como la inteligencia artificial y el aprendizaje automático.

Desde la perspectiva de desarrollo institucional, el Gobierno Autónomo Descentralizado Municipal del cantón Cuenca gestiona una alta carga de trámites relacionados con la planificación urbana, entre los cuales se destacan los permisos de construcción, licencias urbanísticas y revisiones de planos arquitectónicos. Estos trámites, se gestionan con procesos automatizados para su control y seguimiento, sin embargo, de acuerdo a la información que consta en la base de datos de información histórica de los tramites de construcción aprobados por el GAD Municipal del cantón Cuenca, la especificidad de los trámites individuales genera tiempos de respuesta variables. Esta falta de previsibilidad no solo afecta la calidad del servicio brindado a la ciudadanía, sino que también limita la eficiencia interna de la Institución al no contar con métricas claras para gestionar recursos humanos y técnicos.

Por otra parte, desde el punto de vista del impacto económico que los trámites administrativos generan en los ciudadanos, desarrolladores inmobiliarios y constructores, especialmente en lo relacionado con los costos de financiamiento de la inversión debemos considerar que; en proyectos de construcción, la planificación adecuada del flujo de caja y del calendario de ejecución es fundamental para evitar sobrecostos y asegurar la viabilidad financiera. Sin embargo, uno de los factores más inciertos dentro del ciclo de vida de un proyecto urbano es el tiempo requerido

para la obtención de permisos y aprobaciones técnicas por parte del Gobierno Autónomo Descentralizado Municipal del cantón Cuenca.

Marco teórico referencial

En este apartado se definirá el estado del arte en base a estudios relevantes, relacionados con: Machine Learning en trámites administrativos, Random Forest, Hiperparámetros en Random Forest para pronósticos cuantitativos y Métricas de evaluación de modelos predictivos.

3.1 MACHINE LEARNING EN TRÁMITES ADMINISTRATIVOS

En los últimos años, el aprendizaje automático (machine learning) se ha consolidado como una herramienta clave para la modernización. Su capacidad de procesar grandes volúmenes de datos, identificar patrones y generar predicciones confiables ha permitido su aplicación en instituciones públicas, en ámbitos como la justicia, la salud, la planificación urbana y la gestión de trámites gubernamentales (Sun y Medaglia, 2019).

Varios estudios han demostrado el potencial del machine learning para predecir el tiempo de finalización de trámites y reducir la incertidumbre en procesos administrativos. Tax N. et al. (2017) utilizaron redes LSTM (Long Short-Term Memory) para predecir la duración remanente de casos en ejecución, logrando precisión incluso con secuencias incompletas. En otro caso, Polato M. et al. (2018) demostraron que la combinación de modelos LSTM con técnicas de clustering permite adaptarse de forma dinámica al tipo de trámite, mejorando la exactitud en la predicción.

Bijalwan et al. (2024) aplicaron y compararon seis algoritmos de Machine Learning para predecir la productividad de empleados municipales, destacando el buen rendimiento de XGBoost en términos de precisión y eficiencia.

En relación a la calidad de los datos, Suriadi et al. (2017) propusieron un enfoque sistemático para limpiar registros de eventos utilizando minería de procesos,

logrando un incremento en la calidad de los datos y por lo tanto la optimización de las operaciones de los modelos predictivos.

En base a los artículos citados se confirma que la utilización de machine learning para la predicción de tiempos de aprobación de trámites de construcción del GAD del cantón Cuenca puede ser viable.

3.2 RANDOM FOREST COMO MODELO DE REFERENCIA

Entre las diferentes técnicas de machine learning, el algoritmo Random Forest ha mostrado especial eficacia en tareas de clasificación y regresión vinculadas a procesos administrativos y productivos. Este modelo, introducido por Breiman, se basa en múltiples árboles de decisión a partir de subconjuntos aleatorios de datos y variables, combinando sus resultados mediante votación o promedio (Yu et al., 2015).

En entornos industriales, Yamashiro y Nonaka (2021) utilizaron Random Forest para estimar tiempos de procesamiento en fábricas reales, demostrando que este algoritmo puede reducir el makespan de la producción hasta en un 30 %. Estos hallazgos, aunque provenientes de sectores distintos, confirman la versatilidad de Random Forest en la optimización de procesos y tiempos, haciéndolo relevante para trámites administrativos que presentan variabilidad significativa.

Tyralis (2017) indica que el pronóstico de series temporales mediante algoritmos de machine learning ha ganado popularidad recientemente. Random Forest es un algoritmo de aprendizaje automático implementado en el pronóstico de series temporales; sin embargo, la mayoría de sus propiedades de pronóstico han permanecido inexploradas.

Artículos recientes como el de Merza et al. (2024) indican que Random Forest es uno de los modelos de machine learning más populares, utilizado tanto para problemas de clasificación como de regresión. Como modelo de conjunto,

demuestra una alta precisión predictiva y baja varianza, además de ser fácil de aprender y optimizar.

Random Forest es un algoritmo flexible con una amplia gama de aplicaciones y funciona bien con un gran número de conjuntos de datos. Además, es inmune a las suposiciones estadísticas, así como a la carga de preprocesamiento y puede manejar un conjunto de datos grande con alta dimensionalidad y valores faltantes (Zhu, 2020).

Basados en estas fuentes bibliográficas se considera oportuno utilizar Random Forest en la predicción del tiempo de aprobación de trámites de construcción en el GAD Municipal del cantón Cuenca.

3.3 HIPERPARÁMETROS EN RANDOM FOREST PARA PRONÓSTICOS CUANTITATIVOS

La mayoría de los modeladores dedican mucho tiempo a ejecutar la sintonización de hiperparámetros en su modelo y puede que no haya ninguna mejora en el rendimiento del modelo (Trithipkaiwanpon y Taetragool, 2021). El desempeño del algoritmo Random Forest depende de la correcta configuración de sus hiperparámetros, que determinan la capacidad predictiva y la eficiencia del modelo. Entre los más relevantes se encuentran:

- `n_estimators` (número de árboles): controlar este parámetro permite balancear entre precisión y tiempo de cómputo.
- `max_depth` (profundidad máxima): regula la complejidad de los árboles; valores elevados pueden causar sobreajuste.
- `max_features` (número de variables consideradas en cada división): influye en la diversidad de los árboles y en la reducción de la varianza.
- `min_samples_split` y `min_samples_leaf`: definen el tamaño mínimo de muestras en nodos y hojas, afectando la generalización del modelo.

Probst et al. (2019) realizaron un análisis de sensibilidad de estos parámetros, concluyendo que una correcta optimización no solo mejora la precisión, sino también la estabilidad del modelo en diferentes dominios. En el contexto de trámites administrativos, la optimización de hiperparámetros es esencial para lograr un equilibrio entre precisión y costos computacionales.

3.4 MÉTRICAS DE EVALUACIÓN

La validación de modelos predictivos en entornos cuantitativos requiere el uso de métricas que midan con precisión la exactitud y robustez de las predicciones. Entre las más utilizadas destacan:

Error Absoluto Medio (MAE): mide la desviación promedio entre valores predichos y reales, fácil de interpretar.

Raíz del Error Cuadrático Medio (RMSE): otorga mayor penalización a los errores grandes, útil en escenarios donde retrasos significativos tienen alto impacto.

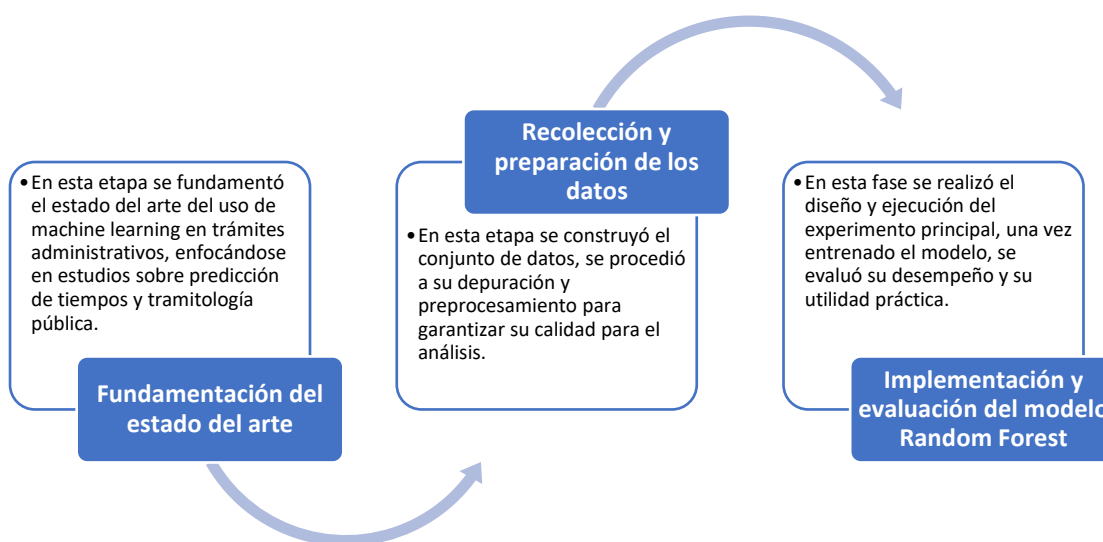
Coefficiente de Determinación (R^2): evalúa la proporción de variabilidad explicada por el modelo, indicando su capacidad explicativa.

En estudios aplicados, Bijalwan et al. (2024) utilizaron estas métricas para evaluar la productividad de empleados municipales, mientras que Yamashiro y Nonaka (2021) validaron su modelo en procesos fabriles, ambos confirmando la efectividad de Random Forest en predicciones cuantitativas. Estas métricas permiten comparar de manera objetiva el rendimiento de diferentes algoritmos, garantizando decisiones basadas en evidencia.

Materiales y metodología

El presente Trabajo de Titulación adopta un enfoque cuantitativo, experimental, considerando que se buscó evaluar el comportamiento del modelo predictivo Random Forest, en función de un conjunto de variables extraídas de trámites reales de construcción. La experimentación se realizó aplicando y ajustando dicho modelo sobre datos históricos, midiendo su capacidad para predecir el tiempo de aprobación de nuevos trámites. A continuación, se detalla la metodología utilizada por fases:

Figura 1. Metodología



4.1 FUNDAMENTACIÓN DEL ESTADO DEL ARTE

En esta etapa se fundamentó el estado del arte del uso de machine learning en trámites administrativos, enfocándose en estudios sobre predicción de tiempos y tramitología pública. Incluyó las siguientes actividades:

- Revisar literatura científica y técnica, realizar búsquedas en bases académicas (Scopus, IEEE Xplore, SpringerLink, Google Scholar) y filtrar artículos entre 2010 y 2025 con foco en estudios aplicados a servicios públicos, gobierno electrónico, BPM y trámites burocráticos.

- Se clasificaron estudios relevantes, agruparon los estudios encontrados en categorías como: Machine Learning en trámites administrativos, Random Forest, Hiperparámetros en Random Forest para pronósticos cuantitativos y Métricas de evaluación de modelos predictivos.
- Se analizaron enfoques metodológicos y técnicas aplicadas, se evaluaron los métodos de recolección de datos, preparación de conjuntos de entrenamiento, y técnicas de validación empleadas, haciendo énfasis en estudios que aplican modelos como Random Forest con variables estructuradas.
- Se justificó el uso de Random Forest, comparando el rendimiento del algoritmo Random Forest frente a otros modelos en escenarios similares y se identificaron razones técnicas por las cuales es adecuado para este caso.
- Se redactó del marco del estado del arte, se integraron los hallazgos en un documento organizado por secciones: Machine Learning en trámites administrativos, Random Forest como modelo de referencia, Hiperparámetros en Random Forest para pronósticos cuantitativos y Métricas de evaluación.

4.2 RECOLECCIÓN Y PREPARACIÓN DE LOS DATOS

En esta etapa se construyó el conjunto de datos, se procedió a su depuración y preprocesamiento para garantizar su calidad para el análisis. Considerando que la recolección de datos es un paso fundamental para tener éxito en la obtención de resultados (Hernandez Mendoza y Duana Avila, 2020), se utilizó una fuente secundaria de datos que disminuye el uso de recursos y tiempo, aunque esto signifique tener dificultad con la calidad de los datos (Borda Pérez et al., 2013), en este paso se incluyeron las siguientes actividades:

- Gestionar con las unidades técnicas del GAD Municipal del cantón Cuenca para acceder a las bases de datos históricas de trámites de construcción

desde el año 2007 a la presente fecha (sistemas como SIGTRAM y otros repositorios institucionales).

- Se extrajeron registros con la siguiente información: número de trámite, tipo de trámite, año de ingreso, mes de ingreso, día de ingreso, parroquia, área de terreno, tipo de intervención, área de intervención, tipo de edificación, número de edificaciones, número de departamentos, número de habitaciones, número de servicios higiénicos, número de pisos, personería, fecha de ingreso y fecha de aprobación.
- Se consolidó la información en un único archivo estructurado en formato Excel con un total de 47.485 registros.

Tomando en cuenta que el objetivo principal de la preparación de los datos consiste en organizarlos de manera que puedan ser procesados por los programas de construcción de modelos que hayan sido elegidos y, al mismo tiempo, asegurar que los datos se hallan de tal forma que se pueda obtener el mejor modelo posible del conjunto de datos (i Solé, 2019), se aplicaron los siguientes procesos:

- Normalización de los datos para las variables categóricas: tipo de trámite, parroquia, tipo de intervención, tipo de edificación y personería.
- Fusión y Categorización del atributo objetivo: tiempo de aprobación (días de despacho), definido como la diferencia entre fecha de ingreso y fecha de aprobación.
- Tratamiento de falta de datos, para lo cual, se definió no tratar los casos que presentan un atributo no observado, se eliminaron los registros incompletos o inconsistentes obteniendo un conjunto de datos final con 9.841 registros.

4.3 IMPLEMENTACIÓN Y EVALUACIÓN DEL MODELO RANDOM FOREST

En esta fase se realizó el diseño y ejecución del experimento principal, una vez entrenado el modelo, se evaluó su desempeño y su utilidad práctica:

- Se dividió el conjunto de datos en subconjuntos de entrenamiento (80%) y prueba (20%) mediante muestreo aleatorio estratificado para mantener la proporción de clases o valores.
- Se implementó el modelo de Random Forest.
- Se ajustaron los hiperparámetros clave mediante el método GridSearchCV de python (número de árboles y profundidad máxima).
- Se entrenó el modelo con los datos procesados, utilizando el tiempo de aprobación (días de despacho) como variable objetivo.
- Se registró el modelo final y su configuración para su análisis posterior y posible reutilización.
- Se aplicó el modelo sobre el conjunto de prueba para obtener predicciones del tiempo de aprobación, para lo cual, se calcularon las métricas de evaluación:
 - Error absoluto medio (MAE)
 - Error cuadrático medio (MSE)
 - Raíz del error cuadrático medio (RMSE)
 - Coeficiente de determinación (R^2)
- Se analizó la importancia de las variables generadas con Random Forest para identificar cuáles factores influyen más en la duración del trámite.
- Se realizó la interpretación de resultados y se validó la utilidad práctica del modelo como herramienta de apoyo a la mejora de los procesos de atención ciudadana y a la planificación de proyectos.

4.4 VARIABLES DE ESTUDIO

Para alcanzar el objetivo general del presente trabajo de titulación, se establecieron tres escenarios con las siguientes variables de estudio:

Escenario 1

Tabla 1. Variables de estudio (Escenario 1)

Nombre de la variable	Tipo	Definición	Dimensión (valor)	Nivel de medición
Tipo de trámite	Independiente	Tipo de trámite ingresado	Texto (ANTEPROYECTOS, PLANOS)	Cualitativo
Año de ingreso	Independiente	Año en el que se ingresa el trámite	numérico	Cuantitativo
Mes de ingreso	Independiente	Mes en el que se ingresa el trámite	numérico	Cuantitativo
Día de ingreso	Independiente	Día en el que se ingresa el trámite	numérico	Cuantitativo
Parroquia	Independiente	Parroquia en la que se encuentra ubicado el predio	Texto (BAÑOS, BELLAVISTA, CAÑARIBAMBA, CHAUCHA, CHECA, CHIQUINTAD, CUMBE, EL BATAN, EL SAGRARIO, EL VALLE, EL VECINO, GIL RAMIREZ DAVALOS, HERMANO MIGUEL, HUAYNA CAPAC, LLACAO, MACHANGARA, MOLLETURO, MONAY, NULTI, OCTAVIO CORDERO PALACIOS, PACCHA, QUINGEO, RICAURTE, SAN BLAS, SAN JOAQUIN, SAN SEBASTIAN, SANTA ANA, SAYAUSI, SIDCAY, SININCAY, SUCRE, TARQUI, TOTORACOCHA, TURI, VICTORIA DEL PORTETE, YANUNCAY)	Cualitativo
Área total del terreno	Independiente	Área total del predio en el que se emplazará el proyecto	Numérico	Cuantitativo
Tipo de intervención	Independiente	Tipo de intervención a	Texto (NUEVA EDIFICACIÓN,	Cualitativo

Nombre de la variable	Tipo	Definición	Dimensión (valor)	Nivel de medición
		realizar en el predio	AMPLIACIÓN, REFORMA, RESTAURACIÓN, RECONSTRUCCIÓN, ACTUALIZACIÓN, ESTADO ACTUAL, READECUACIÓN / REHABILITACIÓN)	
Área de intervención	Independiente	Área a ser intervenida por el proyecto	Numérico	Cuantitativo
Tipo de edificación	Independiente	Tipo de edificación del proyecto	Texto (AISLADA, CONTINUA CON RETIRO FRONTAL, CONTINUA SIN RETIRO FRONTAL, PAREADA CON RETIRO FRONTAL, PAREADA SIN RETIRO FRONTAL, AISLADA CON RETIRO FRONTAL, CONTINUA CON PORTAL, NO DEFINIDO, PATRIMONIAL)	Cualitativo
Número de edificaciones	Independiente	Número de edificaciones del proyecto	Numérico	Cuantitativo
Número de departamentos	Independiente	Número de departamentos	Numérico	Cuantitativo
Número de habitaciones	Independiente	Número de habitaciones	Numérico	Cuantitativo
Número de servicios higiénicos	Independiente	Número de servicios higiénicos	Numérico	Cuantitativo
Número de pisos	Independiente	Número de pisos	Numérico	Cuantitativo
Personería	Independiente	Tipo de personería del dueño del predio	Texto (Natural, Jurídica)	Cualitativo
Días despacho	Dependiente	Número de días transcurridos para el despacho del trámite	Numérico	Cuantitativo

Nota: Elaboración propia del autor.

Escenario 2

Tabla 2. Variables de estudio (Escenario 2)

Nombre de la variable	Tipo	Definición	Dimensión (valor)	Nivel de medición
Tipo de trámite	Independiente	Tipo de trámite ingresado	Texto (ANTEPROYECTOS, PLANOS)	Cualitativo
Año de ingreso	Independiente	Año en el que se ingresa el trámite	numérico	Cuantitativo
Mes de ingreso	Independiente	Mes en el que se ingresa el trámite	numérico	Cuantitativo
Día de ingreso	Independiente	Día en el que se ingresa el trámite	numérico	Cuantitativo
Parroquia	Independiente	Parroquia en la que se encuentra ubicado el predio	Texto (BAÑOS, BELLAVISTA, CAÑARIBAMBA, CHAUCHA, CHECA, CHIQUINTAD, CUMBE, EL BATAN, EL SAGRARIO, EL VALLE, EL VECINO, GIL RAMIREZ DAVALOS, HERMANO MIGUEL, HUAYNA CAPAC, LLACAO, MACHANGARA, MOLLETURO, MONAY, NULTI, OCTAVIO CORDERO PALACIOS, PACCHA, QUINGEO, RICAURTE, SAN BLAS, SAN JOAQUIN, SAN SEBASTIAN, SANTA ANA, SAYAUSI, SIDCAY, SININCAY, SUCRE, TARQUI, TOTORACOCHA, TURI, VICTORIA DEL PORTETE, YANUNCAY)	Cualitativo
Tipo de intervención	Independiente	Tipo de intervención a realizar en el predio	Texto (NUEVA EDIFICACIÓN, AMPLIACIÓN, REFORMA, RESTAURACIÓN, RECONSTRUCCIÓN,	Cualitativo

Nombre de la variable	Tipo	Definición	Dimensión (valor)	Nivel de medición
			ACTUALIZACIÓN, ESTADO ACTUAL, READECUACIÓN / REHABILITACIÓN)	
Tipo de edificación	Independiente	Tipo de edificación del proyecto	Texto (AISLADA, CONTINUA CON RETIRO FRONTAL, CONTINUA SIN RETIRO FRONTAL, PAREADA CON RETIRO FRONTAL, PAREADA SIN RETIRO FRONTAL, AISLADA CON RETIRO FRONTAL, CONTINUA CON PORTAL, NO DEFINIDO, PATRIMONIAL)	Cualitativo
Personería	Independiente	Tipo de personería del dueño del predio	Texto (Natural, Jurídica)	Cualitativo
Días despacho	Dependiente	Número de días transcurridos para el despacho del trámite	Numérico	Cuantitativo

Nota: Elaboración propia del autor.

Escenario 3

Tabla 3. Variables de estudio (Escenario 3)

Nombre de la variable	Tipo	Definición	Dimensión (valor)	Nivel de medición
Tiempo Respuesta	Dependiente (Yi)	Tiempo transcurrido desde el inicio al final de los trámites de avalúo.	minutos	Cuantitativo
Área total del terreno	Independiente	Área total del predio en el que se emplazará el proyecto	Numérico	Cuantitativo
Área de intervención	Independiente	Área a ser intervenida por el proyecto	Numérico	Cuantitativo

Nombre de la variable	Tipo	Definición	Dimensión (valor)	Nivel de medición
Número de edificaciones	Independiente	Número de edificaciones del proyecto	Numérico	Cuantitativo
Número de departamentos	Independiente	Número de departamentos	Numérico	Cuantitativo
Número de habitaciones	Independiente	Número de habitaciones	Numérico	Cuantitativo
Número de servicios higiénicos	Independiente	Número de servicios higiénicos	Numérico	Cuantitativo
Número de pisos	Independiente	Número de pisos	Numérico	Cuantitativo
Días despacho	Dependiente	Número de días transcurridos para el despacho del trámite	Numérico	Cuantitativo

Nota: Elaboración propia del autor.

4.5 PREGUNTA DE INVESTIGACIÓN

Se definió la siguiente pregunta para la investigación

¿La aplicación del modelo de aprendizaje automático Random Forest, utilizando variables relacionadas con las características del proyecto, su ubicación y condiciones administrativas, permite predecir con alta precisión el tiempo de aprobación de los trámites de construcción en el GAD Municipal del cantón Cuenca?

En este sentido,

Hipótesis nula (H_0):

La implementación del modelo de aprendizaje automático Random Forest, basado en variables asociadas a las características del proyecto, su localización geográfica y aspectos administrativos, no permite predecir con alta precisión el tiempo de aprobación de trámites de construcción en el GAD Municipal del cantón Cuenca, contribuyendo significativamente a la optimización de los procesos administrativos.

Hipótesis alternativa (H_1):

La implementación del modelo de aprendizaje automático Random Forest, basado en variables asociadas a las características del proyecto, su localización geográfica y aspectos administrativos, permite predecir con alta precisión el tiempo de aprobación de trámites de construcción en el GAD Municipal del cantón Cuenca, por lo que no contribuye de manera efectiva a la optimización de los procesos administrativos.

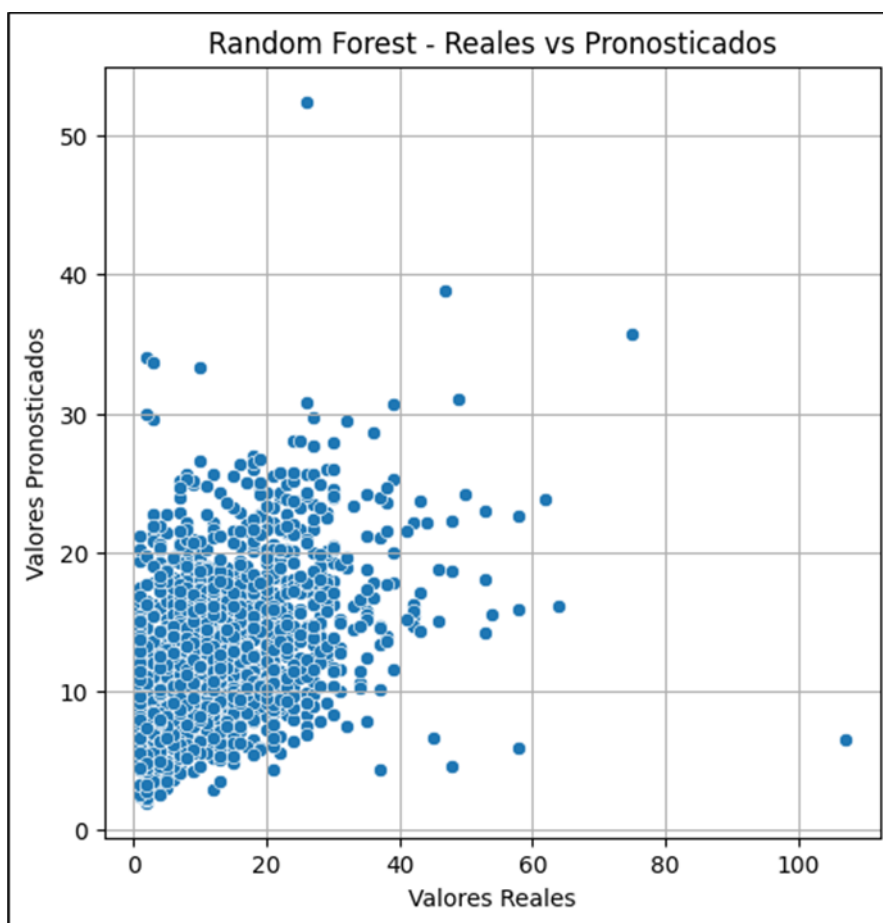
Resultados y discusión

A continuación, se detallan los resultados del entrenamiento y aplicación del algoritmo de Machine Learning: Random Forest para cada una de las bases de datos de los escenarios 1,2,3.

5.1 ENTRENAMIENTO Y RESULTADOS CON EL ESCENARIO 1

En el escenario 1 se utilizaron todas las variables del conjunto de datos (Tipo de trámite, Año de ingreso, Mes de ingreso, Día de ingreso, Parroquia, Área total del terreno, Tipo de intervención, Área de intervención, Tipo de edificación, Número de edificaciones, Número de departamentos, Número de habitaciones, Número de servicios higiénicos, Número de pisos, Personería) para el pronóstico del tiempo de aprobación de trámites (días de despacho), obteniendo los resultados que se describen a continuación.

Figura 2. Relación de tiempo real vs. Pronosticado (Escenario 1)



Nota: Elaboración propia del autor.

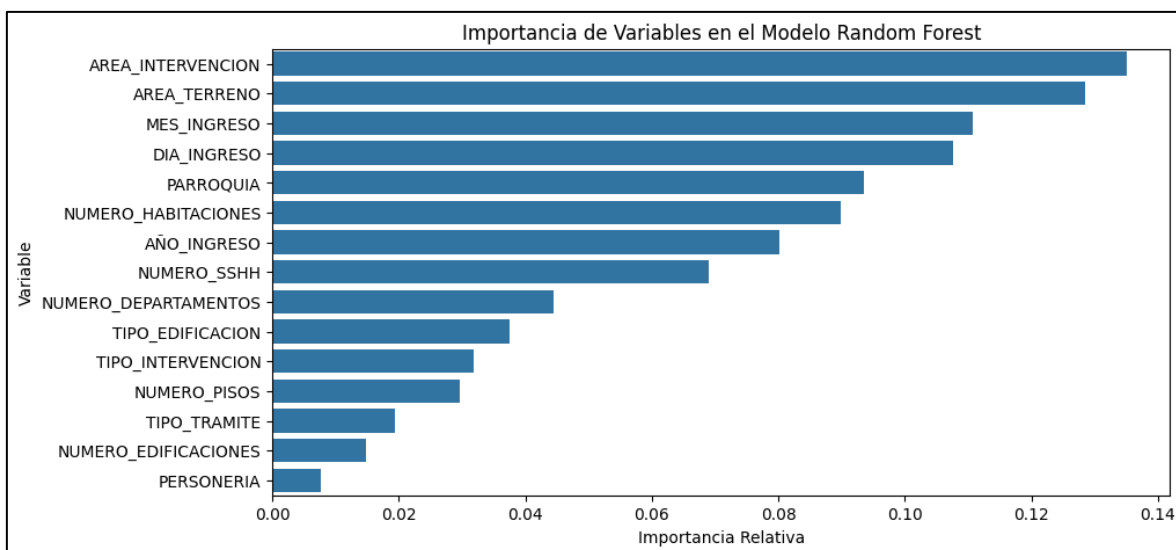
Como se puede evidenciar en la gráfica, la relación entre los valores pronosticados y los valores reales no tienen una aproximación lineal simétrica que es lo que se esperaría de un pronóstico confiable. Los valores de las métricas de evaluación obtenidas para este escenario son las siguientes:

Tabla 4. Métricas de evaluación (Escenario 1)

Métrica	Valor
Error absoluto medio (MAE)	6,363352078
Error cuadrático medio (MSE)	79,13717499
Raíz del error cuadrático medio (RMSE)	8,895907766
Coefficiente de determinación (R^2)	0,665431345

Nota: Elaboración propia del autor.

Figura 3. Importancia de cada variable en el pronóstico del tiempo de aprobación de trámites de construcción en el GAD Cuenca (Escenario 1)



Nota: Elaboración propia del autor.

Por otra parte, en este escenario el Área de Intervención (AREA_INTERVENCION) es la variable independiente con mayor importancia en el pronóstico del tiempo de aprobación de los trámites (días de despacho). En la siguiente tabla se presentan los resultados de la importancia de cada variable en el pronóstico del tiempo de trámites de construcción en el GAD de Cuenca para el escenario 1:

Tabla 5. Importancia de las variables (Escenario 1)

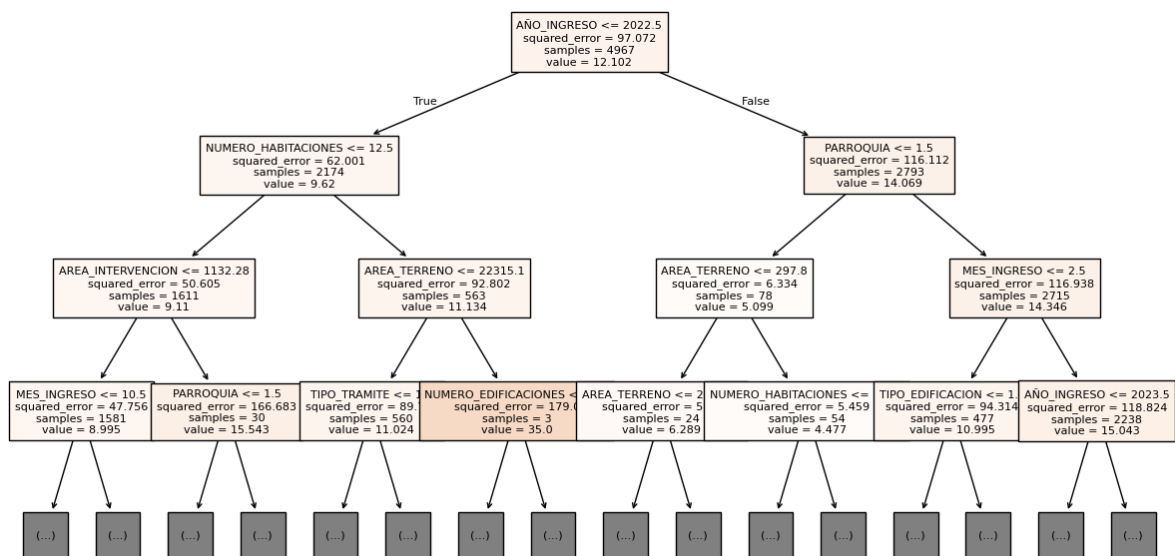
Variable	Importancia
AREA_INTERVENCION	0,135171
AREA_TERRENO	0,128578
MES_INGRESO	0,110695
DIA_INGRESO	0,107650
PARROQUIA	0,093575
NUMERO_HABITACIONES	0,089897
AÑO_INGRESO	0,080162
NUMERO_SSHH	0,069048
NUMERO_DEPARTAMENTOS	0,044456
TIPO_EDIFICACION	0,037505
TIPO_INTERVENCION	0,031897

Variable	Importancia
NUMERO_PISOS	0,029655
TIPO_TRAMITE	0,019347
NUMERO_EDIFICACIONES	0,014770
PERSONERIA	0,007595

Nota: Elaboración propia del autor.

Para la predicción se implementó el algoritmo de Random Forest en Python utilizando el modelo RandomForestRegressor. El diagrama de muestra el árbol de decisión del algoritmo de Random Forest en el pronóstico de tiempo de aprobación de los trámites de construcción en el GAD Cuenca (días de despacho) para el escenario 1.

Figura 4. Árbol de decisión Random Forest (Escenario 1)



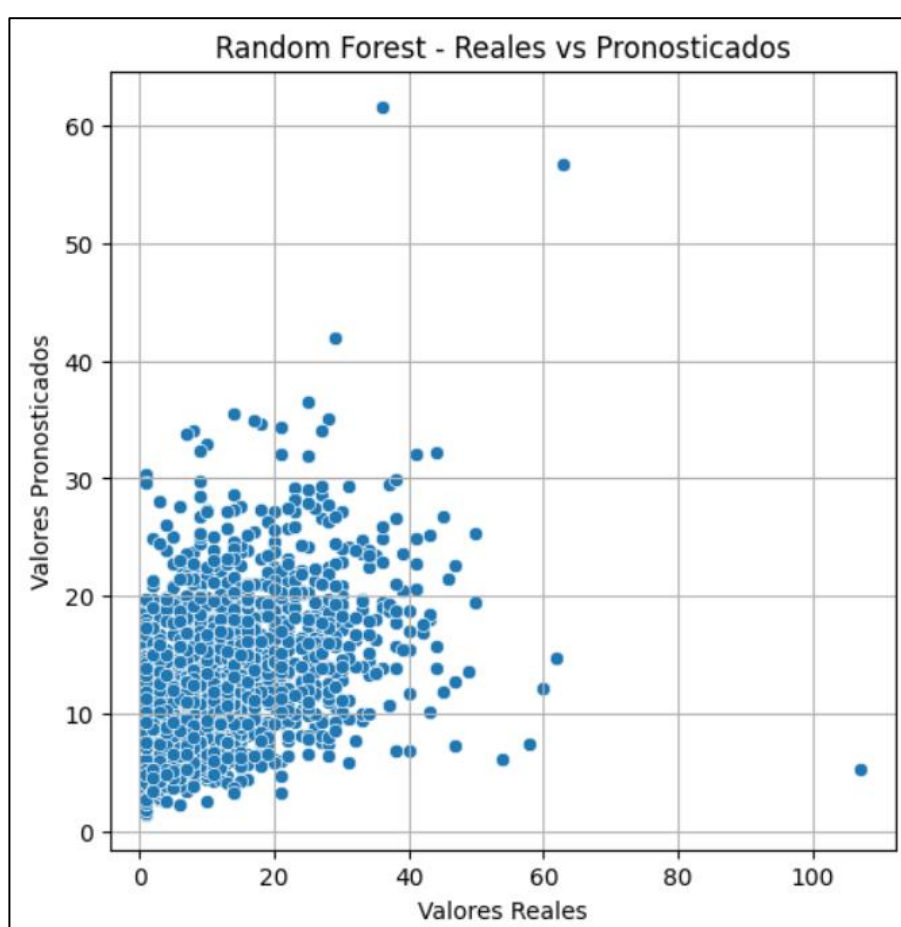
Nota: Elaboración propia del autor.

Para la aplicación del algoritmo se utilizaron los hiperparámetros número de árboles ($n_{estimators}$) = 100 y profundidad máxima (max_depth) = 3.

5.2 ENTRENAMIENTO Y RESULTADOS CON EL ESCENARIO 2

En el escenario 2 se utilizaron las variables cualitativas del conjunto de datos (Tipo de trámite, Año de ingreso, Mes de ingreso, Día de ingreso, Parroquia, Tipo de intervención, Tipo de edificación, Personería) para el pronóstico del tiempo de aprobación de trámites (días de despacho), obteniendo los resultados que se describen a continuación.

Figura 5. Relación de tiempo real vs. Pronosticado (Escenario 2)



Nota: Elaboración propia del autor.

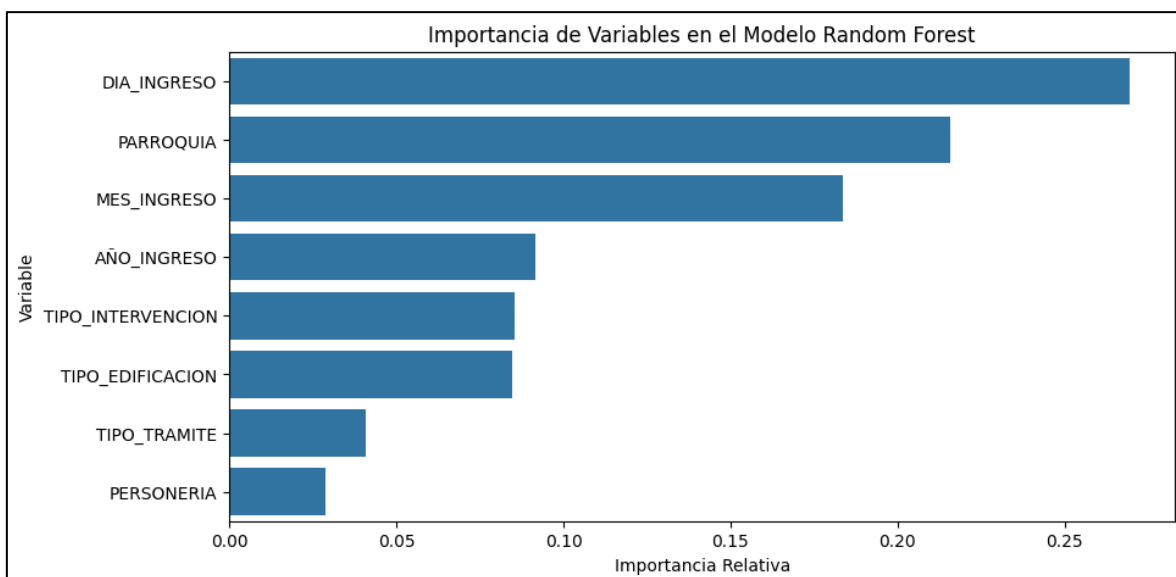
Como se puede evidenciar en la gráfica, la relación entre los valores pronosticados y los valores reales no tienen una aproximación lineal simétrica que es lo que se esperaría de un pronóstico confiable. Los valores de las métricas de evaluación obtenidas para este escenario son las siguientes:

Tabla 6. Métricas de evaluación (Escenario 2)

Métrica	Valor
Error absoluto medio (MAE)	6,3949049
Error cuadrático medio (MSE)	80,7847729
Raíz del error cuadrático medio (RMSE)	8,9880350
Coefficiente de determinación (R ²)	0,6357042

Nota: Elaboración propia del autor.

Figura 6. Importancia de cada variable en el pronóstico del tiempo de aprobación de trámites de construcción en el GAD Cuenca (Escenario 2)



Nota: Elaboración propia del autor.

Por otra parte, en este escenario el Día de Ingreso (DIA_INGRESO) es la variable independiente con mayor importancia en el pronóstico del tiempo de aprobación de los trámites (días de despacho). En la siguiente tabla se presentan los resultados de la importancia de cada variable en el pronóstico del tiempo de trámites de construcción en el GAD de Cuenca para el escenario 2:

Tabla 7. Importancia de las variables (Escenario 2)

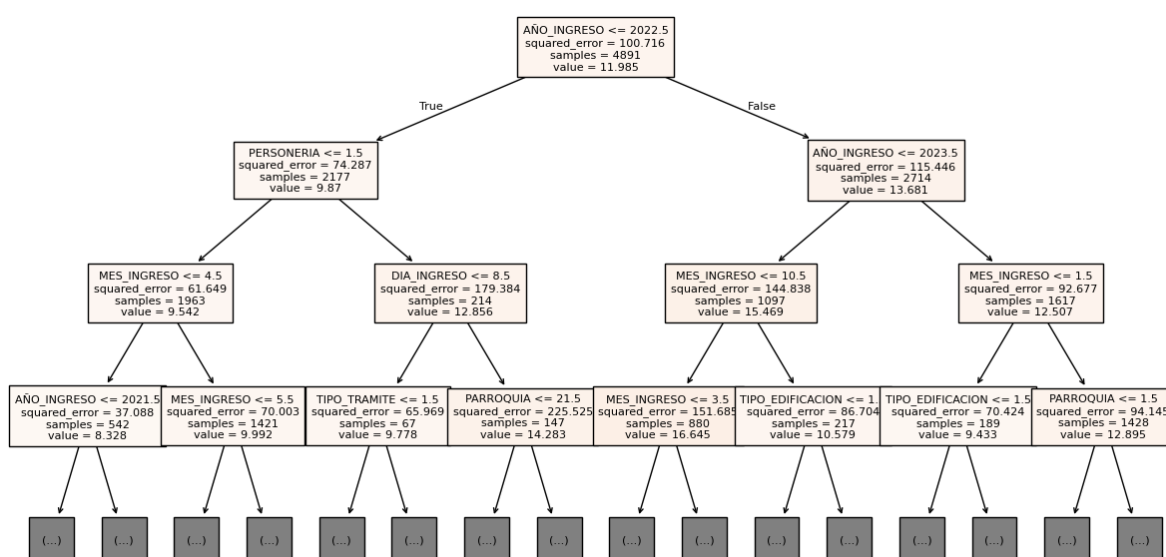
Variable	Importancia
DIA_INGRESO	0,269498
PARROQUIA	0,215626

Variable	Importancia
MES_INGRESO	0,183486
AÑO_INGRESO	0,091462
TIPO_INTERVENCION	0,085519
TIPO_EDIFICACION	0,084824
TIPO_TRAMITE	0,040711
PERSONERIA	0,028874

Nota: Elaboración propia del autor.

Para la predicción se implementó el algoritmo de Random Forest en Python utilizando el modelo RandomForestRegressor. El diagrama de muestra el árbol de decisión del algoritmo de Random Forest en el pronóstico de tiempo de aprobación de los trámites de construcción en el GAD Cuenca (días de despacho) para el escenario 2.

Figura 7. Árbol de decisión Random Forest (Escenario 2)



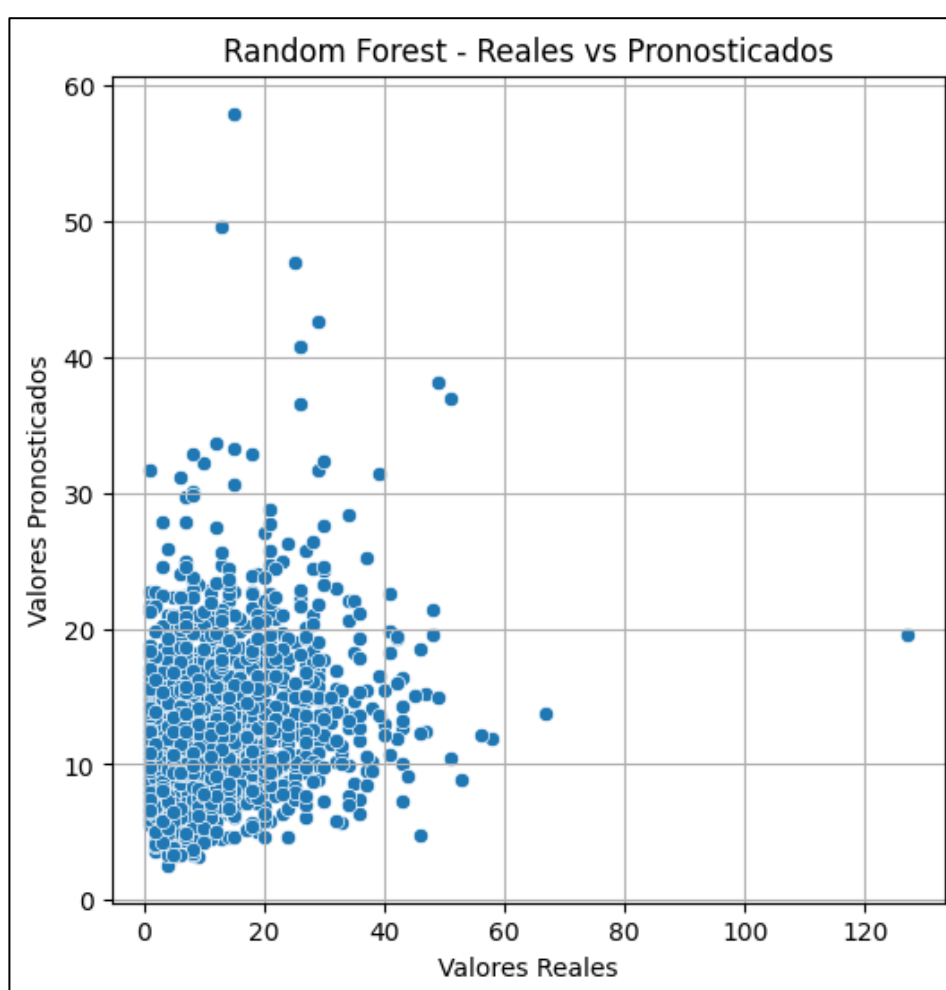
Nota: Elaboración propia del autor.

Para la aplicación del algoritmo se utilizaron los hiperparámetros número de árboles ($n_{estimators}$) = 100 y profundidad máxima (max_depth) = 3.

5.3 ENTRENAMIENTO Y RESULTADOS CON EL ESCENARIO 3

En el escenario 3 se utilizaron las variables cuantitativas del conjunto de datos (Área total del terreno, Área de intervención, Número de edificaciones, Número de departamentos, Número de habitaciones, Número de servicios higiénicos, Número de pisos) para el pronóstico del tiempo de aprobación de trámites (días de despacho), obteniendo los resultados que se describen a continuación.

Figura 8. Relación de tiempo real vs. Pronosticado (Escenario 3)



Nota: Elaboración propia del autor.

Como se puede evidenciar en la gráfica, la relación entre los valores pronosticados y los valores reales no tienen una aproximación lineal simétrica

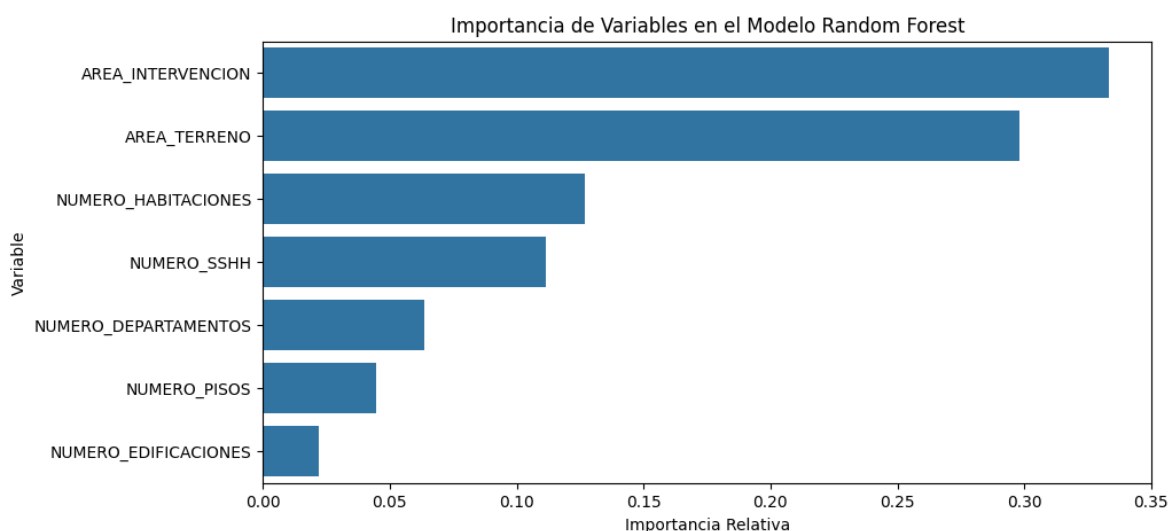
que es lo que se esperaría de un pronóstico confiable. Los valores de las métricas de evaluación obtenidas para este escenario son las siguientes:

Tabla 8. Métricas de evaluación (Escenario 3)

Métrica	Valor
Error absoluto medio (MAE)	7,3035441
Error cuadrático medio (MSE)	98,2227453
Raíz del error cuadrático medio (RMSE)	9,9107389
Coefficiente de determinación (R ²)	0,4676322

Nota: Elaboración propia del autor.

Figura 9. Importancia de cada variable en el pronóstico del tiempo de aprobación de trámites de construcción en el GAD Cuenca (Escenario 3)



Nota: Elaboración propia del autor.

Por otra parte, en este escenario el Área de Intervención (AREA_INTERVENCION) es la variable independiente con mayor importancia en el pronóstico del tiempo de aprobación de los trámites (días de despacho). En la siguiente tabla se presentan los resultados de la importancia de cada variable en el pronóstico del tiempo de trámites de construcción en el GAD de Cuenca para el escenario 3:

Tabla 9. Importancia de las variables (Escenario 3)

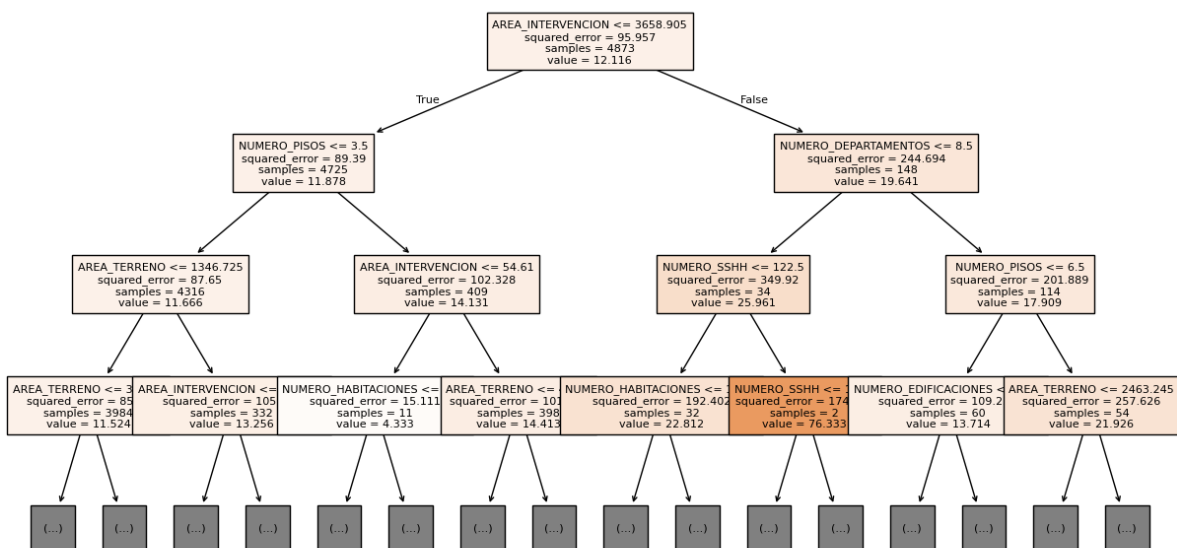
Variable	Importancia
AREA_INTERVENCION	0,333514

Variable	Importancia
AREA_TERRENO	0,298315
NUMERO_HABITACIONES	0,126680
NUMERO_SSHH	0,111467
NUMERO_DEPARTAMENTOS	0,063379
NUMERO_PISOS	0,044777
NUMERO_EDIFICACIONES	0,021869

Nota: Elaboración propia del autor.

Para la predicción se implementó el algoritmo de Random Forest en Python utilizando el modelo RandomForestRegressor. El diagrama de muestra el árbol de decisión del algoritmo de Random Forest en el pronóstico de tiempo de aprobación de los trámites de construcción en el GAD Cuenca (días de despacho) para el escenario 3.

Figura 10. Árbol de decisión Random Forest (Escenario 3)



Nota: Elaboración propia del autor.

Para la aplicación del algoritmo se utilizaron los hiperparámetros número de árboles ($n_estimators$) = 100 y profundidad máxima (max_depth) = 3.

5.4 DISCUSIÓN

Una vez analizados los resultados de los tres escenarios se evidencia que para los escenarios 1 y 2 el coeficiente de determinación presenta un ajuste moderado del modelo y para el escenario 3 el ajuste del modelo es bajo. Por otra parte, tanto el Error absoluto medio (MAE) como la Raíz del Error Cuadrático Medio (RMSE) indican que el escenario con promedio de error menor en las estimaciones es el escenario 1 que incluye todas las variables del conjunto de datos.

Tabla 10. Métricas de evaluación para los tres escenarios

Métrica	Escenario 1	Escenario 2	Escenario 3
Error absoluto medio (MAE)	6,36335208	6,3949049	7,3035441
Error cuadrático medio (MSE)	79,137175	80,784773	98,2227453
Raíz del error cuadrático medio (RMSE)	8,89590777	8,988035	9,9107389
Coefficiente de determinación (R ²)	0,66543135	0,6357042	0,4676322

Nota: Elaboración propia del autor.

Analizando la importancia de las variables utilizadas para pronosticar el tiempo de aprobación de los trámites de construcción en el GAD Cuenca, los resultados indican que la variable con más importancia es el Área de Intervención (AREA_INTERVENCION). Cuando se utilizan solamente variables cualitativas, la variable Día de Ingreso (DIA_INGRESO) es la variable independiente con mayor importancia.

Conclusiones

En base al análisis de los resultados de los tres escenarios propuestos, se puede concluir que la combinación de variables cualitativas y cuantitativas generan mejores resultados en la predicción del tiempo de aprobación de los trámites de construcción en el GAD Municipal del cantón Cuenca utilizando el modelo de aprendizaje automático Random Forest.

La aplicación del modelo de aprendizaje automático Random Forest, en base a variables relacionadas con las características del proyecto, su ubicación y condiciones administrativas, no permite predecir con alta precisión el tiempo de aprobación de los trámites de construcción en el GAD Municipal del cantón Cuenca, considerando los valores obtenidos en las métricas de evaluación utilizadas.

Referencias (Bibliografía)

Asamblea Constituyente del Ecuador. (20 de Octubre de 2008). CONSTITUCIÓN DE LA REPÚBLICA DEL ECUADOR. *Registro Oficial 449*. Montecristi, Manabí, Ecuador.

Asamblea Nacional del Ecuador. (19 de Octubre de 2010). CODIGO ORGANICO DE ORGANIZACION TERRITORIAL, COOTAD. *Registro Oficial Suplemento 303*. Quito, Pichincha, Ecuador.

Asamblea Nacional del Ecuador. (23 de Octubre de 2018). LEY ORGÁNICA PARA LA OPTIMIZACIÓN Y EFICIENCIA DE TRÁMITES ADMINISTRATIVOS. *Registro Oficial Suplemento No. 353*. Quito, Pichincha, Ecuador.

Battisti, K., Dörn, M., Eggeling, E., Eichler, C., Loës, J., Scherret, J., . . . Ullrich, T. (2023). An Automatic Process for the Application of Building Permits. *Buildings*, 13(78), 17. <https://doi.org/10.3390>

Bijalwan, P., Gupta, A., Mendiratta, A., Johri, A., y Asif, M. (2024). Predicting the Productivity of Municipality Workers: A Comparison of Six Machine Learning Algorithms. *Economies*, 12(16), 19. <https://doi.org/10.3390>

Borda Pérez, M., Tuesca Molina, R., y Navarro Lechuga, E. (2013). *Métodos cuantitativos: Herramientas para la investigación en la salud*. Colombia: Universidad del Norte.

Breuker, D., Matzner, M., Delfmann, P., y Becker, J. (2016). Comprehensible predictive models for business processes. *MIS Quarterly*, 40(4), 1009-1034. <https://doi.org/10.25300/MISQ/2016/40.4.02>

Calles-García, J., y González-Pérez, P. (2011). *La Biblia del Footprinting*.

Cuenca, C. C. (10 de 09 de 2024). ACTA DE LA SESIÓN EXTRAORDINARIA DEL CONCEJO MUNICIPAL DEL CANTÓN CUENCA DEL MARTES 10 DE SEPTIEMBRE DEL 2024.

ACTA DE LA SESIÓN EXTRAORDINARIA DEL CONCEJO MUNICIPAL DEL CANTÓN
CUENCA DEL MARTES 10 DE SEPTIEMBRE DEL 2024. Cuenca, Azuay, Ecuador.

- Folino, G., Greco, G., y Pontieri, L. (2014). Discovering interesting patterns from event logs: A clustering and multi-perspective process mining approach. *Data & Knowledge Engineering*, 87, 1-24. <https://doi.org/10.1016/j.datak.2013.07.001>
- Gómez, A., Camargo, J., y Ayala, D. (2020). Predicción del tiempo de resolución de casos judiciales mediante aprendizaje automático. *Revista Facultad de Ingeniería Universidad de Antioquia*, 96, 10-21. <https://doi.org/10.17533/udea.redin.20200587>
- Hernandez Mendoza, S., y Duana Avila, D. (2020). Técnicas e instrumentos de recolección de datos. *Boletín Científico de las Ciencias Económico Administrativas del ICEA*, 9(17), 51 - 53. <https://doi.org/10.29057/icea.v9i17.6019>
- i Solé, R. S. (2019). Preparación de datos. *uoc.edu*, 56.
- Merza, H. M., Sbeity, I., Dbouk, M., Al Abidin Ibrahim, Z., y Kadhim, A. S. (2024). A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Springer Science and Business Media*, 1145 SCI(308639), 386 - 403. https://doi.org/10.1007/978-3-031-53717-2_37
- Navarrin, N., Beatrice, V., Mirko, P., y Alessandro, S. (2017). LSTM networks for data-aware remaining time prediction of business process instances. IEEE Symposium Series on Computational Intelligence. *University of Padova, Department of Mathematics*, 8. <https://doi.org/10.1109>
- Polato, M., Sperduti, A., Burattin, A., y de Leoni, M. (2018). Time and activity sequence prediction of business process instances. *Data Mining and Knowledge Discovery (Springer)*, 32(6), 1716 - 1763. <https://doi.org/10.1007/s10618-018-0579-6>

- Probst, P., Wright, M. N., y Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3), 1942-4787. <https://doi.org/doi.org/10.1002/widm.1301>
- Rogge-Solti, A., y Weske, M. (2013). Prediction of remaining service execution time using stochastic Petri nets with arbitrary firing delays. *Lecture Notes in Computer Science (Springer)*, 8094, 389 - 404. https://doi.org/10.1007/978-3-642-45005-1_27
- Senderovich, A., Gergatsoulis, M., y Mandelbaum, A. (2017). Queue mining for delay prediction in multi-stage processes. *Elsevier*, 64, 98–119. <https://doi.org/10.1016/j.is.2016.08.003>
- Senderovich, A., Weidlich, M., Gal, A., y Mandelbaum, A. (2015). Queue mining—Predicting delays in service processes. *Lecture Notes in Business Information Processing (Springer)*, 215, 42 - 57. https://doi.org/10.1007/978-3-319-07881-6_4
- Sun, T., y Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly (Elsevier)*, 36(2), 250 - 261. <https://doi.org/10.1016/j.giq.2018.09.008>
- Suriadi, S., Andrews, R., ter Hofstede, A., y Wynn, M. (2017). Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. *Information Systems (Elsevier)*, 64, 132 - 150. <https://doi.org/10.1016/j.is.2016.07.011>
- Tax, N., Verenich, I., La Rosa, M., y Dumas, M. (2017). LSTM networks for data-aware remaining time prediction of business process instances. *Information Systems (Elsevier)*, 64, 1 - 16. <https://doi.org/10.1016/j.is.2016.07.005>
- Trithipkaiwanpon, T., y Taetragool, U. (2021). Sensitivity analysis of random forest hyperparameters. *Institute of Electrical and Electronics Engineers*

Inc.(9454885), 1163 - 1167. <https://doi.org/10.1109/ECTI-CON51831.2021.9454885>

Tyralis, H. (2017). Variable selection in time series forecasting using random forests. *MDPI AG*, 10(114), 25. <https://doi.org/10.3390/a10040114>

Verenich, I., Dumas, M., Rosa, M. L., Maggi, F. M., y Teinmaa, I. (2019). Survey and cross-benchmark comparison of remaining time prediction methods in business process monitoring. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(4), 34. <https://doi.org/10.1145/3331449>

www.elhacker.net. (s.f.). *www.elhacker.net*.
https://www.elhacker.net/trucos_google.html

Yamashiro, H., y Nonaka, H. (2021). Estimation of processing time using machine learning and real factory data for optimization of parallel machine scheduling problem. *Operations Research Perspectives (Elsevier)*, 8, 100 - 196.
<https://doi.org/10.1016/j.orp.2021.100196>

Yamashiro, H., y Nonaka, H. (2021). Estimation of processing time using machine learning and real factory data for optimization of parallel machine scheduling problem. *Operations Research Perspectives. Econstor*, 8, 9.
<https://doi.org/10.1016>

Yu, S., Chen, Y., Huang, Q., Kang, Y., He, R., y Gu, S. (2015). Using the random forest method for classification and regression in hydrology. *CRC Press/Balkema*(113959), 213 - 218. <https://doi.org/10.1201/b18180-35>

Zhu, T. (2020). Analysis on the applicability of the random forest. *Institute of Physics Publishing*, 1607(12123), 6. <https://doi.org/10.1088/1742-6596/1607/1/012123>