

**CALIFICACIÓN DE CONSTRUCCIONES MEDIANTE LA APLICACIÓN DE ALGORITMOS DE
INTELIGENCIA ARTIFICIAL EMPLEANDO FOTOGRAFÍAS DE FACHADAS.**

**CALIFICACIÓN DE CONSTRUCCIONES MEDIANTE LA APLICACIÓN DE ALGORITMOS DE
INTELIGENCIA ARTIFICIAL EMPLEANDO FOTOGRAFÍAS DE FACHADAS.**

HASBLEIDY LORENA ALDANA URREA

**UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS
FACULTAD DE INGENIERÍA
PROYECTO CURRICULAR DE INGENIERÍA CATASTRAL Y GEODESIA
BOGOTÁ D.C
2021**

**CALIFICACIÓN DE CONSTRUCCIONES MEDIANTE LA APLICACIÓN DE ALGORITMOS DE
INTELIGENCIA ARTIFICIAL EMPLEANDO FOTOGRAFÍAS DE FACHADAS.**

HASLEIDY LORENA ALDANA URREA

**Trabajo de grado para optar al título de
Ingeniera Catastral y Geodesta**

Director:

PHD. EDWIN ROBERT PEREZ CARVAJAL

**UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS
FACULTAD DE INGENIERÍA
PROYECTO CURRICULAR DE INGENIERÍA CATASTRAL Y GEODESIA
BOGOTÁ D.C
2021**

*A Dios y al universo por darme la fuerza, la sabiduría y los valores para alcanzar mis metas.
A mi hermana Angélica por ser incondicional en mi vida, por inspirarme a ser mejor cada día, y por
su inagotable cariño.*

*A mi compañero de vida Camilo por su infinito amor, por su apoyo, paciencia y motivación durante
tantos años.*

*A mi madre por motivarme a seguir el camino de la superación, por todo su afecto y empeño para
poder alcanzar mis metas académicas.*

A Lala, Karina, Annie, Conny y Velita por llenar mi corazón de huellitas de amor.

A mis amigas y amigos por compartir la vida universitaria con una segunda familia.

A mis docentes y a todo aquel que contribuyó en mi formación personal y académica.

A todos los que creyeron en mí y en mis sueños.

Hasbleidy Lorena Aldana Urrea.

Nota de aceptación:

Firma del presidente del Jurado

Firma del Jurado

Firma del Jurado

Bogotá D.C., _____

AGRADECIMIENTOS

Al finalizar este trabajo deseo manifestar un agradecimiento muy especial:

A la Universidad Distrital Francisco José de Caldas, por brindarme las herramientas para realizar mi formación a nivel personal y profesional. Es un orgullo pertenecer a la comunidad académica.

A mi director de proyecto, Edwin Robert Pérez Carvajal por aportarme su conocimiento y por guiarme en el desarrollo de este trabajo. Durante la carrera he recibido su acompañamiento en temas de diversos campos como avalúos e inteligencia artificial aplicada al territorio.

A la Unidad Administrativa Especial de Catastro Distrital (UAECD) por facilitarme la información catastral de los predios de la UPZ 65 Arborizadora.

A mis docentes por compartir su conocimiento profesional. A los que con sus aportes enriquecieron mi aprendizaje y me brindaron su apoyo.

A mi familia, amigos, compañeros y colegas por su compañía, consejo y apoyo en el transcurso de mi vida universitaria.

A todas las personas que directa e indirectamente contribuyeron en la elaboración del presente estudio.

RESUMEN

Con este trabajo se desarrolló un proceso metodológico para identificar los mejores algoritmos de Inteligencia Artificial (IA) que permitan predecir el puntaje de las construcciones de los predios ubicados en la UPZ 65 Arborizadora. La base de datos empleada contiene atributos de tipo numérico y una variable asociada a la fotografía de la fachada de cada construcción. Con el objeto de determinar el aporte de esta última, los resultados se compararon con los obtenidos en una base de datos idéntica sin incluir las fotografías.

Inicialmente se solicitó a la Unidad Administrativa Especial de Catastro Distrital (UAECD) información referente a los inmuebles de tipo residencial de la zona de estudio. Se depuró la base de datos dejando los atributos y variables considerados como necesarios para alcanzar un buen aprendizaje en los algoritmos. Posteriormente se realizó la toma de las fotografías de las fachadas en el aplicativo *Mapas Bogotá* con la extensión de *Street View*, y se anexó a la base de datos. Así mismo, se añadió una variable que describiera el acabado de los materiales observados en la fachada.

En el software WEKA se realizó el preprocesamiento de los datos y posteriormente se emplearon varios algoritmos de los grupos *Bayes*, *Lazy*, *Funciones* y *Árboles de Decisión*. Se seleccionaron los que presentaron resultados con más del 90% de instancias correctamente clasificadas, los cuales fueron *BayesNet*, *IBK* y *RandomForest*. Así mismo, se realizó un análisis más detallado en el algoritmo *BayesNet* por ser el clasificador más destacado al obtener los mejores resultados.

Finalmente, se aplicaron los algoritmos seleccionados (*BayesNet*, *IBK*, *RandomForest*) en el conjunto de datos con variables numéricas y se escogió nuevamente el *BayesNet* por presentar la mayor cantidad de instancias correctamente clasificadas.

Así, se llegó a la comparación del algoritmo con el mayor desempeño en cada una de las bases de datos, el cual fue el *BayesNet*. Como resultado del análisis realizado, se determinó que las fotografías y la variable correspondiente a los acabados de las fachadas, aportaron información fundamental para alcanzar el buen aprendizaje de los algoritmos *BayesNet*, *IBK* y *RandomForest*. Los resultados obtenidos dan cuenta de la efectividad de la IA implementada en el proceso catastral de la obtención del puntaje de las construcciones de tipo residencial. Además demostró ser un proceso rápido, fácil y con un costo de operación bajo.

Los resultados presentados en este estudio permitieron establecer que los algoritmos de Inteligencia Artificial (IA) son una herramienta útil, rápida y de bajo costo para realizar el proceso de calificación de las construcciones. Cabe señalar que es necesario que este puntaje se encuentre actualizado ya que sirve como base para realizar el cálculo del avalúo catastral, y la liquidación del Impuesto Predial Unificado (IPU). Así se garantiza que las características físicas actuales del inmueble estén acordes con el valor liquidado por concepto del IPU.

CONTENIDO

INTRODUCCIÓN	12
1 PLANTEAMIENTO DEL PROBLEMA	15
2 OBJETIVOS	17
2.1 OBJETIVO GENERAL	17
2.2 OBJETIVOS ESPECÍFICOS	17
3 MARCO REFERENCIAL	18
4 MARCO CONCEPTUAL	20
4.1 CONCEPTOS BÁSICOS DE CATASTRO	20
4.1.1 <i>Catastro</i>	20
4.1.2 <i>Determinación de los avalúos catastrales</i>	20
4.1.3 <i>Reconocimiento Predial</i>	21
4.1.4 <i>Calificación de las construcciones</i>	21
4.1.5 <i>Tipificación de las construcciones</i>	23
4.1.6 <i>Impuesto predial Unificado (IPU)</i>	24
4.1.7 <i>Chip</i>	24
4.2 INTELIGENCIA ARTIFICIAL	24
4.3 ALGORITMOS DE INTELIGENCIA ARTIFICIAL	25
4.3.1 <i>Teorema de Bayes</i>	25
4.3.2 <i>Árboles de decisión</i>	26
4.3.3 <i>Lazy (K- vecino)</i>	27
4.4 SOFTWARE WEKA	27
4.4.1 <i>Estadístico Kappa</i>	28
4.4.2 <i>Error absoluto medio (Mean Absolute Error)</i>	28
4.4.3 <i>Raíz del error cuadrático medio (Root Mean Absolute Error)</i>	29
4.4.4 <i>Error absoluto relativo (Relative Absolute Error)</i>	29
4.4.5 <i>Raíz del error cuadrático relativo (Root Relative Squared Error)</i>	29
4.4.6 <i>Matriz de confusión</i>	29
5 METODOLOGÍA	30
5.1 ÁREA DE ESTUDIO	30
5.2 RECOPIACIÓN Y DEPURACIÓN DE INFORMACIÓN CATASTRAL DE LOS PREDIOS	31
5.2.1 <i>Nombre y tipo de las variables empleadas</i>	32
5.3 TOMA DE LAS FOTOGRAFÍAS DE LAS FACHADAS DE LOS PREDIOS	33
5.4 PREPARACIÓN DE LOS DATOS	34
5.4.1 <i>Duplicación de las bases de datos</i>	37

5.5	PREPROCESAMIENTO DE INFORMACIÓN EN EL SOFTWARE WEKA	37
5.6	ENTRENAMIENTO DE LOS ALGORITMOS.....	39
5.7	EVALUACIÓN DE LOS ALGORITMOS	40
5.8	ANÁLISIS DE LOS RESULTADOS	41
5.8.1	<i>Resultados obtenidos de la base de datos con fotografías.....</i>	<i>41</i>
6	ANÁLISIS COMPARATIVO DE LOS RESULTADOS OBTENIDOS EN BASES DE DATOS CON IMÁGENES Y SIN IMÁGENES	49
6.1	RESULTADOS OBTENIDOS EN LA BASE DE DATOS SIN FOTOGRAFÍAS.....	49
6.2	COMPARACIÓN DE LOS RESULTADOS	50
7	CONCLUSIONES Y PERSPECTIVAS.....	54
7.1	CONCLUSIONES	54
7.2	PERSPECTIVAS	55
	BIBLIOGRAFÍA.....	57
	ANEXOS.....	60
•	ANEXO 1: RESULTADOS DEL ALGORITMO BAYESNET CON EL MODO DE PRUEBA USE TRAINING SET.....	60
•	ANEXO 2: RESULTADOS EN CADA PREDICCIÓN DEL ALGORITMO BAYESNET CON EL MODO DE PRUEBA SUPPLIED TEST SET.....	63
•	ANEXO 3: RESULTADOS DEL ALGORITMO BAYESNET CON EL MODO DE PRUEBA PERCENTAGE SPLIT	64
•	ANEXO 4: RESULTADOS DEL ALGORITMO BAYESNET CON EL MODO DE PRUEBA CROSS-VALIDATION	65

LISTA DE FIGURAS

Figura 4-1. Software Weka.	27
Figura 5-1. Diagrama metodología proyecto.	30
Figura 5-2. Zona de estudio UPZ 65 Arborizadora.	31
Figura 5-3. Aplicativo Mapas Bogotá.	33
Figura 5-4. Visualizador Street View.	33
Figura 5-5. Base de datos en formato .xlsx.	37
Figura 5-6. Base de datos en formato .csv.	38
Figura 5-7. Histograma de datos.	38
Figura 5-8. Histograma de los datos con filtro <i>MathExpression</i>	39
Figura 5-9. Partición de datos en los modos de prueba.	40
Figura 5-10. Resultados estadísticos del algoritmo BayesNet.	43
Figura 5-11. Errores de predicción modo de prueba Use Training Set.	44
Figura 5-12. Errores de predicción modo de prueba Supplied Test Set.	45
Figura 5-13. Errores de predicción modo de prueba Percentage Split.	46
Figura 5-14. Errores de predicción modo de prueba Cross- Validation.	46
Figura 6-1. Comparación de resultados obtenidos para las bases de datos 1 y 2, modo <i>Use Training Set</i>	51
Figura 6-2. Comparación de resultados obtenidos para las bases de datos 1 y 2, modo <i>Supplied Test Set</i>	51
Figura 6-3. Comparación de resultados obtenidos para las bases de datos 1 y 2, modo <i>Cross- Validation</i>	52
Figura 6-4. Comparación de resultados obtenidos para las bases de datos 1 y 2, modo <i>Percentage Split</i>	52

LISTA DE TABLAS

Tabla 4-1. Tipos de fachadas según el IGAC.....	22
Tabla 4-2. Rango de puntajes de construcción según el IGAC.....	23
Tabla 5-1. Nombre y tipo de variables.....	32
Tabla 5-2. Fachadas empleadas para el estudio.....	34
Tabla 5-3. Datos por estrato.....	35
Tabla 5-4. Datos por variable Acabado de fachada.....	36
Tabla 5-5. Datos por variable Acabado de fachada, base de datos Validación.....	36
Tabla 5-6. Rango de puntaje por tipo.....	36
Tabla 5-7. Variables en las bases de datos.....	37
Tabla 5-8. Algoritmos empleados en el experimento.....	40
Tabla 5-9. Cantidad de datos empleados en cada modo de evaluación.....	41
Tabla 5-10. Resultados obtenidos en la base de datos con fotografías.....	41
Tabla 5-11. Comparación de los resultados obtenidos en los algoritmos.....	42
Tabla 5-12. Errores estadísticos obtenidos en el algoritmo BayesNet.....	43
Tabla 5-13. Resultados matrices de confusión.....	48
Tabla 6-1. Resultados obtenidos en la base de datos sin fotografías.....	49
Tabla 6-2. Comparación de los resultados obtenidos.....	50

INTRODUCCIÓN

En Colombia, el avalúo catastral es un valor que sirve como base para realizar la liquidación del Impuesto Predial Unificado (IPU) [1]. Para cada predio, el avalúo catastral es la suma de los avalúos parciales de terreno y construcción, obtenidos mediante la investigación y análisis estadístico del mercado inmobiliario. El valor del terreno tiene en cuenta variables como la norma del suelo, la actividad económica, el estado de las vías, entre otras. Para la construcción se determina un puntaje el cual se obtiene mediante una visita, en donde el operador observa las características y establece un número que indica las condiciones de la edificación [2]. Para ello es necesario considerar el estado de conservación y los cambios en la construcción, representados en cuatro variables principales: estructura, acabados principales, baño y cocina. Adicionalmente, se requiere información asociada al complemento industrial y sus datos generales [3]. En Bogotá, la Unidad Administrativa Especial de Catastro Distrital (UAECD), es la entidad encargada de determinar el avalúo catastral mediante el Censo Inmobiliario el cual se realiza cada año en todos los predios de la ciudad [4].

Por otra parte, el Impuesto Predial Unificado (IPU) es un gravamen de orden municipal, recaudado a todos los propietarios de los predios existentes dentro de su autoridad territorial [5]. La tarifa del IPU se determina mediante los rangos del avalúo catastral. En la actualidad, el IPU se considera una de las principales fuentes de ingreso de rentas tributarias locales en varios países del mundo, incluyendo Colombia. Sin embargo, en un estudio reciente se encontró que en el mayor número de unidades territoriales, este recaudo no alcanza su máximo potencial. La razón principal de esta problemática en la mayor parte del territorio colombiano, está asociada a la desactualización catastral, afectando la base gravable y el recaudo municipal [6].

Entre tanto, la Inteligencia Artificial (IA) es la combinación de algoritmos que son creados para imitar los comportamientos y capacidades del ser humano. La ingeniería con enfoque en IA propone la creación de sistemas y máquinas que posean un comportamiento inteligente, similar al que sería llevado a cabo por una persona, teniendo en cuenta factores como la capacidad de adaptación a entornos cambiantes y la creatividad. Las áreas de estudio en las que se desarrolla la IA son muy variadas, ya que van desde las Tecnologías de la Información, hasta la Neurociencia, Psicología, la Ciencia Cognitiva, la Física, las Matemáticas, entre otras [7].

Es necesario que los gobiernos locales tengan como prioridad la actualización del catastro, especialmente en los municipios con una alta tasa de desactualización y de informalidad predial. Es importante tener la información catastral actualizada, ya que asegura un mayor control sobre los recaudos que se hacen por concepto del IPU. Los múltiples algoritmos que desarrolla la IA se postulan como una alternativa para actualizar el puntaje de las construcciones periódicamente. Esto garantizaría que la utilización de este puntaje sea adecuada para obtener el cálculo del avalúo catastral y a su vez, el IPU. En adición, tener en cuenta los cambios en la construcción y mejoras

realizadas recientemente en los predios, procurará una mayor precisión, además de reducir el tiempo y los costos empleados por un operador en campo para este proceso.

En este sentido, el desarrollo de esta investigación tiene como propósito identificar los mejores algoritmos de IA, que permitan predecir la calificación de las construcciones de los predios ubicados en la UPZ Arborizadora. Para ello se emplearán datos de los predios de la zona de estudio procedentes de la UAECD y se realizará la depuración de la información. Se contará con dos bases de datos idénticas, tanto en variables como en atributos. La principal diferencia radica en que en una de estas se adicionará la variable imagen, que corresponderá a la foto de la fachada de los predios, adquiridas de la plataforma de Mapas Bogotá con la extensión de Street View. Para el procesamiento de la información se empleará el software WEKA, y los algoritmos disponibles para cada una de las bases de datos. La predicción resultante se clasificará en las tipologías definidas por el Instituto Geográfico Agustín Codazzi (IGAC) [8]. Así, los resultados se analizarán y se definirán los modelamientos con los mayores porcentajes de asertividad.

Así, este proyecto servirá para determinar los alcances de la IA implementada en el proceso catastral de determinación de la calificación de las construcciones. Además de aportar a la actualización anual requerida para liquidar el IPU correctamente, reducción de costos y tiempos de ejecución en labores de campo.

Este documento se encuentra fraccionado en 7 capítulos. El capítulo 1 aborda el planteamiento del problema en donde se determina, se orienta y se justifica el desarrollo de esta investigación.

El capítulo 2 presenta el objetivo general y los objetivos específicos definidos para el desarrollo de este estudio.

En el capítulo 3 se adelanta el marco referencial, los antecedentes de esta investigación, y se expone la importancia del estudio y la aplicación de la IA en varios procesos catastrales. Además, se presentan algunos ejemplos de investigaciones en torno a este tema y que son desarrolladas en múltiples países del mundo.

El capítulo 4 aborda el marco conceptual en el que se presentan los conceptos asociados al avalúo catastral, el reconocimiento predial, la calificación de construcciones, el proceso de liquidación del IPU, la IA y algunos de los algoritmos que fueron aplicados en el experimento. Finalmente se exponen ciertas herramientas y utilidades del software empleado en el experimento (Weka).

El capítulo 5 presenta el proceso metodológico, que permitió realizar la predicción del puntaje de las construcciones mediante las fotografías de las fachadas y datos numéricos adicionales. Adicionalmente se muestran los resultados de los algoritmos con la mayor cantidad de instancias correctamente clasificadas, y a su vez los criterios estadísticos del algoritmo con el mejor resultado.

En el capítulo 6 se desarrolla el análisis de los resultados obtenidos al aplicar los algoritmos en una base de datos idéntica a la inicial, sin incluir los atributos correspondientes a las fotografías de las fachadas. A partir de este análisis, se realiza la comparación entre los resultados entregados por el software en cada base de datos, y se determina el aporte de las fotografías en la predicción resultante.

Finalmente, en el capítulo 7 se exponen las conclusiones del trabajo, se realizan las recomendaciones y se proponen futuras investigaciones en las que se tenga en cuenta el uso de la IA alrededor de procesos catastrales.

1 PLANTEAMIENTO DEL PROBLEMA

En Colombia el avalúo catastral se determina por la sumatoria de los avalúos parciales realizados a terrenos y construcciones de un territorio. Para calcular el valor del terreno se tiene en cuenta variables como: la norma de uso del suelo, la actividad económica del sector, los servicios públicos disponibles, las vías y su estado, la topografía, entre otros. En la construcción se tiene en cuenta el tipo de estructura, los acabados, el baño, la cocina, el estado de conservación, la edad de la edificación, el uso, entre otros. Estas variables son valoradas por las condiciones que presente el inmueble al momento de la visita de un operador en campo [9].

En Bogotá, la UAECD es la encargada de establecer el avalúo catastral anualmente y además de notificar a los propietarios de los inmuebles de la respectiva actualización, adjuntando el certificado catastral (documento en el que se encuentra establecido el valor del avalúo, áreas construidas y áreas de terreno) [4].

Ahora bien, el IPU es un gravamen real que se aplica a los predios que están ubicados en una unidad territorial, se genera una vez se realiza la formación de los predios y se calcula sobre la base gravable del avalúo catastral en cada vigencia [10].

Si se realiza una comparación del IPU respecto al Producto Interno Bruto (PIB) entre los países de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y América Latina y el Caribe (ALC), se observa que este tributo tiene un rendimiento muy bajo. En el caso de los países OCDE, el IPU representa en promedio el 1% del PIB, mientras que para ALC ningún país alcanza ese porcentaje. El recaudo de este impuesto tiene su mayor potencial en las áreas urbanas, por lo que en ALC puede ser una fuente importante de ingresos al ser la región más urbanizada del mundo (teniendo más del 80% de su población residente en zonas urbanas). En Colombia, se cuenta con una tasa de urbanización del 85,7%, por lo que el IPU podría alcanzar el 1,4% del PIB [11].

Por otro lado, el desarrollo e investigación de la IA tiene como objetivo crear entornos que simulen el funcionamiento de la mente humana en artefactos con sistemas informáticos. La IA cuenta con múltiples campos de aplicación, los cuales recurren a diferentes técnicas para lograr modelamientos óptimos [12]. Una de estas técnicas de aprendizaje es la denominada Machine Learning o Aprendizaje Automático que consiste en crear programas basados en el entrenamiento a partir de ejemplos de información [13]. Esta técnica proporciona procedimientos que están asociados al tratamiento de datos y el análisis predictivo, de modo que son elementos que pueden ser clave para el futuro del Catastro [14]. La IA dispone de herramientas que pueden facilitar el registro actualizado de procesos catastrales que son necesarios para el desarrollo urbano y rural del país, ya que garantiza que la información sea real y vigente [15].

En razón a lo anterior, este proyecto se basa en la implementación de la IA como una posible propuesta para optimizar costos y tiempos, en un proceso catastral que requiere estar actualizado y que es un insumo para realizar el cálculo del avalúo catastral. En tanto, para determinar la calificación de las construcciones por medio de la IA, se evaluarán varios algoritmos proporcionados por el software WEKA, aplicados a un conjunto de datos pertenecientes a la UPZ 65 Arborizadora, y se compararán los resultados. Esto conlleva a la siguiente pregunta de investigación, ¿Cuáles algoritmos de aprendizaje automático predicen con mayor exactitud el puntaje de la construcción de los predios pertenecientes a la UPZ 65 Arborizadora?

2 OBJETIVOS

2.1 Objetivo general

Evaluar el desempeño de la inteligencia artificial aplicada a algoritmos de clasificación para predecir el puntaje de la construcción mediante fotos de las fachadas de los predios pertenecientes a la UPZ 65 Arborizadora.

2.2 Objetivos específicos

Recopilación y depuración de información asociada a los predios pertenecientes a la UPZ 65 Arborizadora, proveniente de la UAECD.

Obtención de las fotografías de las fachadas de los predios por medio del aplicativo de Mapas Bogotá con extensión de Street View.

Estandarización de las imágenes y datos mediante un preprocesamiento que permita reducir los errores en los algoritmos.

Aplicación de los algoritmos seleccionados mediante un conjunto de datos de entrenamiento y de validación, en el software WEKA.

Interpretación de los resultados obtenidos, mediante herramientas incluidas en el software WEKA.

3 MARCO REFERENCIAL

La IA tiene la capacidad de dar respuestas rápidas a un problema y mediante el aprendizaje puede inferir las respuestas de los patrones desconocidos que son insertados al sistema [15]. Estas cualidades permiten su aplicación en las diferentes áreas del catastro. A continuación, se presenta una visión general de algunos proyectos catastrales que han implementado la IA como una herramienta para el tratamiento de la información, resolución de problemas o mejora de los resultados que son obtenidos por los métodos tradicionales.

En 2004 un grupo de investigadores de la entidad catastral denominada Gerencia del Catastro, creó una Red Neuronal (RNA) para valorar los inmuebles de uso *vivienda colectiva* en Madrid, España. El objetivo fue dar una primera aproximación, tomando variables como: distancia del municipio al centro de la ciudad, tamaño del municipio, vetustez, estado de conservación de la construcción, entre otros. En el estudio se seleccionó el uso *vivienda colectiva*, por ser el predominante en esta región, contando con cerca de dos millones de viviendas, que corresponde a dos tercios del total de los inmuebles. La estructura de la RNA es de tres capas con veinte neuronas y una conexión de propagación hacia adelante [3].

En el año 2016, la Universidad de la Laguna planteó un proyecto piloto que permitió detectar, clasificar y ubicar automáticamente señales horizontales que se encuentran en los carriles de la red viaria de Canarias, España. Entre las señales se encuentra: mantener el sentido de circulación (recto), giro a izquierda, giro a derecha, ceda el paso, ausencia de señal, entre otras. Como insumo se hace uso de una Ortofoto (Representación fotográfica de la superficie terrestre a escala, sin errores, ni deformaciones) que contiene información de la zona urbana de Canarias, con una resolución de 10 cm/pixel. El experimento se desarrolla con una red neuronal convolucional, la cual da muy buenos resultados con un 98,48% de aciertos sobre el conjunto de datos de validación. Este proyecto piloto es tan exitoso que da paso a otro proyecto con mayor complejidad que toma fuerza con el apoyo de la institución de Infraestructura de Datos Espaciales de Canarias (GRAFCAN) [16].

Un grupo de investigadores de la ciudad de Córdoba (España) en 2017, desarrolló una red neuronal de tipo Perceptrón Multicapa con el fin de estimar el precio de los locales comerciales. Se consideraron variables determinantes como superficie, estado de conservación, localización, entre otras. En este modelo se obtienen muy buenos resultados referentes al proceso de formación de los precios y un ajuste a las no linealidades del mercado, que incluso muestran más exactitud que los obtenidos por el método tradicional (Modelo Hedónico (MH)) [17].

En el año 2019 en la ciudad de Morelia Mochiacán (México), se empleó una red neuronal multicapa *Back Propagation* (la cual realiza la corrección de los errores y modifica los pesos de las neuronas), para predecir los precios de los inmuebles de esta ciudad. Los mejores resultados se obtuvieron aplicando la función tangente hiperbólica y sigmoidea. Este proyecto da a conocer la viabilidad para estimar los precios de los inmuebles mediante redes neuronales multicapa ya que los errores absolutos medios son muy bajos [18].

Para Latinoamérica, en la Provincia de Córdoba, Argentina se realiza un proyecto en el que se implementa la técnica de aprendizaje Modelos de Valuación Masiva Automatizada (AVM), la cual realiza la predicción de valores de terreno, mediante la aplicación de algoritmos y modelos matemáticos. Los algoritmos que se aplican son: Random Forest, Redes Neuronales y Support Vector Machine. Se selecciona el modelo con la predicción más exacta en cada localidad o zona rural, teniendo en cuenta los indicadores de desempeño y una instancia de validación cruzada [11].

Otro caso en Latinoamérica, es en la ciudad de Córdoba (Argentina). En la revista Vivienda y Ciudad, se expone un experimento en el cual se realiza la valuación masiva del valor unitario del suelo (VUT), mediante la aplicación del algoritmo Random Forest. La técnica geo-estadística Kriging Ordinario es usada para el tratamiento de los residuos. Los resultados en la predicción del modelo son muy buenos y exponen los métodos de aprendizaje automático como una herramienta útil en procesos de valuación masiva. Este modelo garantiza la reducción del tiempo empleado por los métodos convencionales y la constante actualización del valor del suelo [19].

En Colombia los investigadores del área del catastro tienen como prioridad la optimización de los procesos de actualización. En 2017, en la ciudad de Pereira, se realiza un modelo de IA que permite calcular el avalúo catastral de un predio urbano perteneciente a este territorio. Los algoritmos usados son: procesos gaussianos, máquinas de soporte vectorial, redes neuronales y regresión lineal múltiple. La comparación de los resultados de estos algoritmos se realiza mediante validación cruzada anidada. El mejor modelo es determinado por los resultados obtenidos en el error cuadrático medio (EMC), el error absoluto medio (EAM) y el coeficiente de determinación (r^2) [20].

4 MARCO CONCEPTUAL

En la actualidad, Colombia tiene tan sólo el 5% de su territorio actualizado, lo que implica un déficit de la información sobre la cual se realiza el avalúo catastral y el IPU. Por esa razón la estructura tributaria del país es obsoleta, ya que no se cuenta con información actualizada y confiable para la asignación de las tarifas del IPU. Esto dificulta que los contribuyentes cancelen montos acordes a los bienes inmuebles de su propiedad, y contribuya a un bajo recaudo de este impuesto en muchas zonas del país [21]. La IA se postula como un optimizador de costos y tiempo en varios procesos catastrales, y también puede ampliar el enfoque multipropósito que se espera alcanzar en los próximos años en el territorio colombiano. Esta vinculación de IA y catastro permite la continua interacción entre los datos y los formatos disponibles que pueden ser manejados de forma global. Entre estos, destacan los formatos .shp de los Sistemas de Información Geográfica (SIG), los de diseño asistido por ordenador (CAD) como el .dxf, o los raster, o como las imágenes en formato .jpg de la fachada de los predios [14].

4.1 Conceptos básicos de catastro

4.1.1 Catastro

El catastro es una herramienta de registro administrativo que tiene el objetivo de impulsar políticas para la toma de decisiones referentes a la planificación del territorio. Su función es obtener la plena identificación física, jurídica, fiscal y económica de los bienes inmuebles que pertenecen al estado. El catastro debe contar con los instrumentos para abordar eficientemente la administración de tierras, mercados de propiedad, proteger los derechos de los propietarios, soportar el manejo de la tierra y el desarrollo sostenible de infraestructuras que cubran las necesidades de la población en aumento a largo plazo. La implantación de un impuesto territorial ha sido uno de los principales objetivos para la generación del catastro en muchos países [22].

4.1.2 Determinación de los avalúos catastrales

El avalúo catastral es el valor que se le da a cada predio, se obtiene por la suma del valor del avalúo del terreno y de la construcción. Su objetivo final es servir como base gravable para aplicar las tarifas diferenciales a los predios y liquidar el IPU [23]. Para la correcta determinación del avalúo catastral es necesario tener en cuenta los procedimientos que se llevan a cabo en la formación, actualización y conservación catastral.

De acuerdo con ([24], [25] y [26]), la formación catastral es el proceso mediante el cual se obtiene por primera vez la información de carácter físico, jurídico, fiscal y económico de los predios urbanos y rurales de un municipio. A partir de ahí se elabora el plano del municipio y las cartas catastrales con la identificación de cada predio. Luego se determinan las zonas homogéneas geoeconómicas y se realiza el avalúo teniendo en cuenta los valores unitarios definidos por las autoridades catastrales para edificaciones y terrenos. La oficina de Catastro correspondiente liquida el IPU para los propietarios y elabora las listas para el respectivo cobro.

La actualización es el procedimiento que implica renovar la información obtenida en la formación catastral. Se tienen en cuenta los aspectos físicos y jurídicos del catastro, y otro tipo de transformaciones del predio y su alrededor, que puede afectar el avalúo, como el uso, productividad, obras públicas, condiciones del mercado, entre otras.

La conservación catastral es el método que garantiza la coherencia entre el Notariado, el Registro y el Catastro. Se realiza mediante el registro de las variaciones que haya sufrido el predio, como el cambio de propietario, de límites, ampliaciones, demoliciones, reajustes anuales de los avalúos y la inscripción de nuevos predios. Toda esta información se consigna en los documentos catastrales pertinentes [26].

4.1.3 Reconocimiento Predial

Proceso mediante el cual se registra en los documentos catastrales requeridos, la información actualizada referente a los aspectos físicos y jurídicos de los predios [27].

4.1.4 Calificación de las construcciones

Según el manual propuesto por la Dirección de Sistemas de Información y Catastro ([8],[28]), la calificación de las construcciones es el procedimiento para reconocer, clasificar y calificar las construcciones de los predios. El operador realiza una visita al predio y registra en la ficha predial el puntaje total obtenido de las características observadas en la construcción. Los elementos a calificar son: estructura, acabados principales, baño y cocina.

En la estructura se identifican los siguientes componentes:

- **Armazón:** se refiere a la estructura que sostiene la edificación. Son los cimientos, columnas, vigas, entrepisos, entre otros.
- **Muros:** son los muros divisorios que no hacen parte de la estructura principal de la construcción.
- **Cubierta:** corresponde al material del techo del exterior de la edificación
- **Estado de conservación:** es la condición física que se observa en la edificación, depende de la calidad de los materiales, la edad de la construcción y las actividades de mantenimiento que se le hayan realizado durante su uso. Los grados de calificación son: Malo, regular, bueno y excelente.

Los acabados principales corresponden a los elementos más visibles en la construcción, que cumplen la función de impermeabilizar, aislar o dar estética. En ocasiones existen edificaciones que no cuentan con acabados. En este ítem se califica la fachada, el cubrimiento de muros, los pisos y la conservación.

- **Fachada:** son los materiales de los que está constituido el frente de la edificación en donde se ubica la nomenclatura de identificación del predio. Esta calificación es un elemento representativo de la calidad de la vivienda al momento de la visita. Se tiene en cuenta el material y tamaño de: la puerta, las ventanas y los cubrimientos de los muros. De acuerdo a las características mencionadas los tipos de fachada se clasifican en:

- **Sencilla:** las puertas son de madera sin ningún tipo de trabajo, o son láminas metálicas muy sencillas. Cuenta con ventanas de tamaño pequeño y tiene pañetes sencillos, con pinturas económicas o ladrillo común a la vista.
- **Regular:** las puertas son de madera elaborada, o en metal con lámina doblada, puede tener rejas de ornamentación. Puede tener ladrillo a la vista o con pañete.
- **Buena:** las puertas y ventanas presentan trabajos manuales trabajados, con terminados como la pintura en aceite. El cubrimiento es pañete en cemento, pintura en vinilo y materiales con cualidades de más calidad.
- **Lujosa:** es una fachada con diseños arquitectónicos especiales, con puertas y ventanas amplias hechas con materiales que tienen trabajos manuales detallados de alto costo y muy buena calidad.

La Tabla 4-1 contiene una muestra gráfica de los tipos de fachada mencionados anteriormente:

Fachada sencilla	Fachada regular
	
Fachada buena	Fachada lujosa
	

Tabla 4-1. Tipos de fachadas según el IGAC.

Fuente: IGAC [8].

- **Cubrimiento de muros:** se califica teniendo en cuenta el material que cubre los muros de la construcción, se tendrá en cuenta el material que predomine.
- **Pisos:** se califican el material de los pisos de la edificación, se tiene en cuenta el material predominante, si se encuentra en partes iguales en el predio, se elige el piso con el material más costoso.

- **Estado de conservación:** se califica el estado de mantenimiento y conservación de los elementos que conforman los acabados principales.

Y por último se tiene baños y cocina, en este caso el puntaje se asigna de acuerdo al tamaño, enchape, cantidad y calidad del mobiliario, y estado de conservación. Se califica únicamente el baño principal, los secundarios no se tienen en cuenta.

4.1.5 Tipificación de las construcciones

Los inmuebles de uso residencial son tipificados de acuerdo a la calificación de la construcción que se tenga como resultado final en la ficha predial. Según el IGAC [8] la tipificación se da mediante la agrupación de los rangos de puntaje como se observa en la Tabla 4-2.

TIPO	RANGO DE PUNTAJE
1	1-10
2	11-28
3	29-46
4	47-64
5	65-82
6	83-100

Tabla 4-2. Rango de puntajes de construcción según el IGAC.
Fuente: IGAC [8].

- ❖ **Tipo 1:** se refiere a las viviendas que no tienen planeación al momento de su construcción, ya que son fabricadas con materiales de desecho en las paredes y techos. Están ubicadas en zonas marginales que no están urbanizadas, por lo que carecen de servicios públicos básicos, generalmente no están en condiciones sanitarias óptimas, hay escasez de vías de comunicación y transporte urbano.
- ❖ **Tipo 2:** en este rango de puntaje se encuentran viviendas que pueden o no ser planeadas para su construcción. Los materiales que se usan son madera, prefabricado, ladrillo o bloque, cubiertas de zinc o teja de barro. Los acabados son muy sencillos y pueden tener algunos servicios públicos domiciliarios. Se encuentran ubicadas en la periferia por lo que pueden tener algunas vías de acceso pero sin pavimentar, así como pocos medios de transporte urbano.
- ❖ **Tipo 3:** las características de estas viviendas están constituidas por fachadas calificadas como regulares, con muros de ladrillo o bloque. Las cubiertas son en placa de concreto o teja en asbesto-cemento. Cuentan con servicios públicos básicos, con vías pavimentadas por lo que se facilita el acceso al medio de transporte urbano.
- ❖ **Tipo 4:** las viviendas pertenecientes a este rango tienen la estructura en material de concreto, los muros son de ladrillo o bloque recovado, pueden ser estucados y con pintura. Además, las cubiertas son de placa de concreto o teja en asbesto cemento. Cuentan con la totalidad de servicios públicos básicos. En adición, las vías son pavimentadas facilitando el acceso a la zona.

- ❖ **Tipo 5:** las edificaciones están totalmente terminadas. Las fachadas son calificadas como buenas, por lo que cuentan con acabados de calidad. La estructura es en concreto y los muros son de ladrillo o bloque. Las cubiertas se encuentran fabricadas en placa de concreto o está compuesta por materiales de buena calidad. Los servicios públicos con los que cuentan son básicos y complementarios. Alrededor hay grandes zonas verdes, sociales y recreativas, además de parqueaderos y vías de acceso en condiciones óptimas.
- ❖ **Tipo 6:** en esta tipología las viviendas tienen características propias de inmuebles lujosos como grandes áreas, acabados con materiales de muy buena calidad. Cuentan con todos los servicios públicos básicos más complementarios. Cuentan con amplias zonas verdes, sociales y recreativas, así mismo se encuentran parqueaderos y las vías de acceso están en buenas condiciones.

4.1.6 Impuesto predial Unificado (IPU)

En Colombia, el IPU es un gravamen que cobran los municipios por la propiedad raíz ubicada en el perímetro de su jurisdicción [5]. En la Ley 44 de 1990 [29] se estableció el IPU como la unificación de los impuestos: predial, parques y arborización, estratificación socioeconómica y la sobretasa al levantamiento catastral.

Las autoridades catastrales como el IGAC, las oficinas de catastro de Bogotá, Cali, Medellín y Antioquia son las encargadas de fijar los avalúos catastrales que sirven como base para realizar la liquidación del IPU. Otra forma de obtener la base del IPU es mediante el autoevalúo que determine el propietario, el cual no puede ser inferior al avalúo catastral del año anterior. Para fijar las tarifas, los Concejos Municipales deben tener en cuenta el estrato socioeconómico, el uso del suelo en el perímetro urbano y la vetustez de la formación o actualización del catastro [21].

Por otra parte, el 10% del recaudo total de este impuesto debe ser destinado a un fondo de vivienda de estrato bajo, así como la obtención de terrenos para proyectos de construcción de vivienda de interés social (VIS). En Bogotá, los inmuebles que son exentos de pagar el impuesto predial son: los salones comunales de propiedad de las Juntas de Acción Comunal, los predios de propiedad de Iglesias o Comunidades religiosas, predios de la Defensa Civil Colombiana, predios pertenecientes a parques naturales o parques públicos que son propiedad del estado y los edificios declarados monumentos nacionales [26].

4.1.7 Chip

Es un código homologado que sirve para identificar individualmente los predios. En Bogotá la entidad encargada de asignar la numeración para cada predio es la UAECD [4].

4.2 Inteligencia Artificial

La IA es la ciencia que se encarga de la creación de máquinas y artefactos que simulen ciertos procesos realizados por humanos. Las máquinas son creadas con un sistema que tenga la capacidad de analizar, comprender y razonar para ejecutar tareas específicas [14] La IA requiere sistemas informáticos ya que las entidades creadas deben interpretar de forma correcta los datos

proporcionados, aprender de dichos datos mediante la aplicación de los diferentes algoritmos, y lograr la solución de problemas mediante el conocimiento adquirido [30].

El aprendizaje automático es el conjunto de actividades que permiten mejorar los sistemas, a partir de la corrección de los errores cometidos y de las herramientas de los algoritmos que facilitan la sustracción del conocimiento de los datos insertados [12].

Los métodos de clasificación abordan los datos que pertenecen a diferentes clases, las cuales se denominan etiquetas discretas. El objetivo de este método es predecir las clases teniendo en cuenta la clasificación previa de los atributos [31].

4.3 Algoritmos de inteligencia artificial

Weka contiene una extensa colección de algoritmos para ser aplicados sobre los datos mediante interfaces implementadas con Java. Se presentan a continuación los algoritmos BayesNet, IBK y RandomForest.

4.3.1 Teorema de Bayes

El Teorema de Bayes se emplea para realizar la búsqueda de la hipótesis más probable dado un conjunto de datos de entrenamiento y un conjunto a priori que determina la probabilidad de cada una de las hipótesis formuladas [32]. Para encontrar las probabilidades de las hipótesis a posteriori, se aplica la siguiente ecuación:

$$h_{MAP} \equiv \underset{h \in H}{\text{arg max}} \frac{p(D|h) * p(h)}{p(D)} \quad (1)$$

Donde:

- D es el conjunto de datos de entrenamiento.
- h es cada una de las hipótesis formuladas dadas en el conjunto H .
- MAP corresponde con máximo a posteriori, es decir, el mayor valor obtenido una vez se realiza el cálculo.

4.3.1.1 Aprendizaje Bayesiano

Este método de clasificación se basa en modelos estadísticos. El teorema de Bayes es la base para el cálculo de las probabilidades requeridas en los algoritmos. Las características más importantes son: la restricción de independencia de los atributos y el funcionamiento global en la base de datos haciendo que los cambios pequeños no sean significativos [33]. La fórmula matemática se presenta en la siguiente ecuación:

$$P(h/D) = \frac{P(h/D) * p(h)}{p(D)} \quad (2)$$

Donde:

- $P(h)$ es la probabilidad a priori de la hipótesis h . En el algoritmo representa la probabilidad presentada por una determinada clasificación del conjunto de datos de entrenamiento.

- $P(D)$ es la probabilidad de observación para el conjunto de entrenamiento D , siendo este el conjunto usado para clasificar.
- $P(D/h)$ es la probabilidad de observación para el conjunto de entrenamiento D en un universo en el que se verifica la hipótesis h .
- $P(h/D)$ es la probabilidad a posteriori de h , una vez se ha observado el conjunto de entrenamiento D .

4.3.1.1.1 Redes Bayesianas

Una red bayesiana representa la distribución de probabilidades que se tiene en un conjunto de variables. Así mismo se establecen las dependencias existentes entre los atributos a través de un grafo dirigido acíclico, con nodos y arcos que representan las variables y dependencias respectivamente. En razón a lo anterior, se asigna una tabla de probabilidades condicionadas a cada uno de los nodos del grafo [34].

4.3.1.1.2 Algoritmo BayesNet

Algoritmo empleado mayormente para clasificación. Establece la representación de variables y sus dependencias condicionales por medio de un grafo acíclico dirigido. Su funcionamiento está basado en la Teoría de Bayes [33].

4.3.2 Árboles de decisión

Clasifican las instancias mediante la organización de un árbol convencional: de la raíz hasta las hojas. Cada nodo interior del árbol contiene una prueba definida a un atributo, las hojas del árbol son las respuestas de clasificación de las instancias, y las ramas son las descendientes de un nodo interior y corresponden a un valor posible del atributo que se evaluó en ese nodo. Finalmente, se tiene un árbol con un conjunto de atributos que son evaluados y separados en el proceso [33].

4.3.2.1 Algoritmo RandomForest

Realiza particiones sucesivas en los atributos, buscando que la variable y el valor umbral maximicen la homogeneidad de las particiones finales. La homogeneidad de las particiones se puede medir con el índice de Gini:

$$G = \sum_{k=1}^K f_k * (1 - f_k) \quad (3)$$

Donde:

- k es cada clase presente en el total de particiones.
- K es el total de clases presentes en el total de particiones
- f_k es la proporción de los casos en la partición que pertenece a la clase k .

Si se trata de un árbol complejo, se debe calcular el índice de Gini como la suma de todas sus particiones. Las particiones se detienen cuando se alcanza una homogeneidad. Luego de este proceso inicia la poda en la que se realiza la validación cruzada para evitar que el árbol se sobreajuste a los datos de entrenamiento. A continuación, a cada partición realizada a las variables

le corresponde la clase que presente más frecuencia. Los nuevos casos se clasifican en función del lugar en donde se sitúan en el espacio de variables [35].

4.3.3 Lazy (K- vecino)

Los clasificadores basan su funcionamiento en reglas de vecindad en donde buscan un conjunto de prototipos denominado K , siendo este el patrón a clasificar. La búsqueda del vecino más cercano se realiza en el conjunto de referencia denominado como R . Así mismo se emplea una distancia *Euclídea* δ para medir la proximidad [36].

4.3.3.1 Algoritmo IBK

Basa su aprendizaje en las experiencias antiguas para resolver nuevos problemas. La metodología que se aplica en este algoritmo es el método del vecino más cercano. Por las características que presenta se emplean mayormente como algoritmos clasificadores [37].

4.4 Software Weka

Weka es un software desarrollado por la Universidad de Waikato en Nueva Zelanda. Permite realizar transformaciones, clasificación, regresión, clustering, asociación y visualización sobre los datos insertados. Las herramientas contenidas están orientadas a la extensibilidad, por lo que añadir funcionalidades resulta ser un procedimiento sencillo para el usuario. Weka ofrece una amplia gama de algoritmos asociados a máquinas de conocimiento con implementación en Java. El formato empleado por el software es ARFF (“Attribute Relation File Format”) [38].



Figura 4-1. Software Weka.
Fuente: Manual Weka [39].

Según el manual Weka ([39], [40]) el software ofrece las siguientes opciones para realizar la evaluación de los algoritmos:

1. **Use training set:** el entrenamiento se ejecuta con todos los datos disponibles y luego se aplica de nuevo sobre los mismos.
2. **Supplied test set:** esta opción permite seleccionar un archivo externo para evaluar el algoritmo empleado con los datos iniciales.
3. **Percentage split:** el usuario define el porcentaje con el que se realizará la partición de los datos que se emplearán en el entrenamiento. El restante será usado por el software para evaluar el algoritmo por medio de la validación.
4. **Cross-validation:** el software realiza una validación cruzada del número de particiones que el usuario defina (folds). La validación cruzada se logra dividiendo los datos en n partes. Por cada parte, se construye el clasificador con las $n-1$ partes restantes y se evalúa con esta. Es decir, por cada una de las n particiones la validación cruzada es estratificada cuando cada una de las partes conserva las propiedades de la muestra inicial o dicho de otra manera, un porcentaje de elementos de cada clase.

Los resultados estadísticos obtenidos de aplicar los algoritmos de inteligencia artificial disponibles en Weka son:

4.4.1 Estadístico Kappa

El coeficiente Kappa de Cohen es una medida para estimar el grado de acuerdo o concordancia. Se evalúan varios elementos cualitativos correspondientes a variables categóricas. Toma valores entre 0 y 1, siendo 0 un acuerdo nulo o nula concordancia y 1 si existe total coincidencia entre los observadores [41].

Para realizar el cálculo se realiza la siguiente operación matemática:

$$K = P(A) - \frac{P(E)}{1} - P(E) \quad (4)$$

Donde:

- $P(A)$ es la proporción de veces que coinciden los observadores.
- $P(E)$ es la proporción de veces que se espera que los observadores coincidan por azar

4.4.2 Error absoluto medio (Mean Absolute Error)

Es la diferencia media entre los valores observados y los estimados. Tiene unidades propias por lo que permite compararse únicamente con modelos con errores medidos en las mismas unidades. En este cálculo se suma la totalidad de valores absolutos pronosticados y se divide en el número de datos [42].

$$MAE = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n} \quad (5)$$

4.4.3 Raíz del error cuadrático medio (Root Mean Absolute Error)

Mide la diferencia media entre los valores observados y los estimados. La principal característica de esta medida estadística es que castiga severamente los errores grandes. El cálculo consiste en realizar la sumatoria de la diferencia de los valores reales y pronosticados elevados al cuadrado. Se compara con modelos que empleen las mismas unidades [43].

$$RMAE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}} \quad (6)$$

4.4.4 Error absoluto relativo (Relative Absolute Error)

Esta medida estadística permite determinar la diferencia entre el valor observado y su valor medio. Consiste en dividir las diferencias por la variación de los valores observados. La escala de medición es de 0 a 1. Finalmente se multiplica por 100 para obtener un resultado porcentual. La comparación se permite con modelos de diferentes unidades [42].

$$PEAM = \frac{\sum_{t=1}^n \frac{100|Y_t - \hat{Y}_t|}{Y_t}}{n} \quad (7)$$

4.4.5 Raíz del error cuadrático relativo (Root Relative Squared Error)

Muestra la diferencia entre el valor observado con el valor medio. Se calcula dividiendo las diferencias por la variación de los valores observados en una escala de 0 a 1. Finalmente se multiplica por 100 para obtener un resultado porcentual. Esta medida estadística se puede comparar con modelos con resultados en diferentes unidades [42].

$$RRSE = \sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2}{\sum_{i=1}^N (\bar{\theta}_i - \theta_i)^2}} \quad (8)$$

4.4.6 Matriz de confusión

Es una herramienta que permite visualizar el desempeño de un algoritmo. Representa los tipos de aciertos y errores que presenta el modelo en el proceso de aprendizaje y evaluación. En las columnas $j, j=1\dots N$, se observa el número de predicciones que hay en cada clase. Por otra parte cada fila $i, i=1\dots N$ está asociada a las instancias de la clase real ([44], [45]).

5 METODOLOGÍA

El esquema presentado en la Figura 5-1 muestra de manera sencilla y clara el procedimiento a seguir desde la definición del área de estudio y el tratamiento de los datos, hasta la obtención del resultado final: el algoritmo con la mayor cantidad de instancias correctamente clasificadas. En este apartado se presentan los primeros 8 pasos de esta propuesta metodológica.

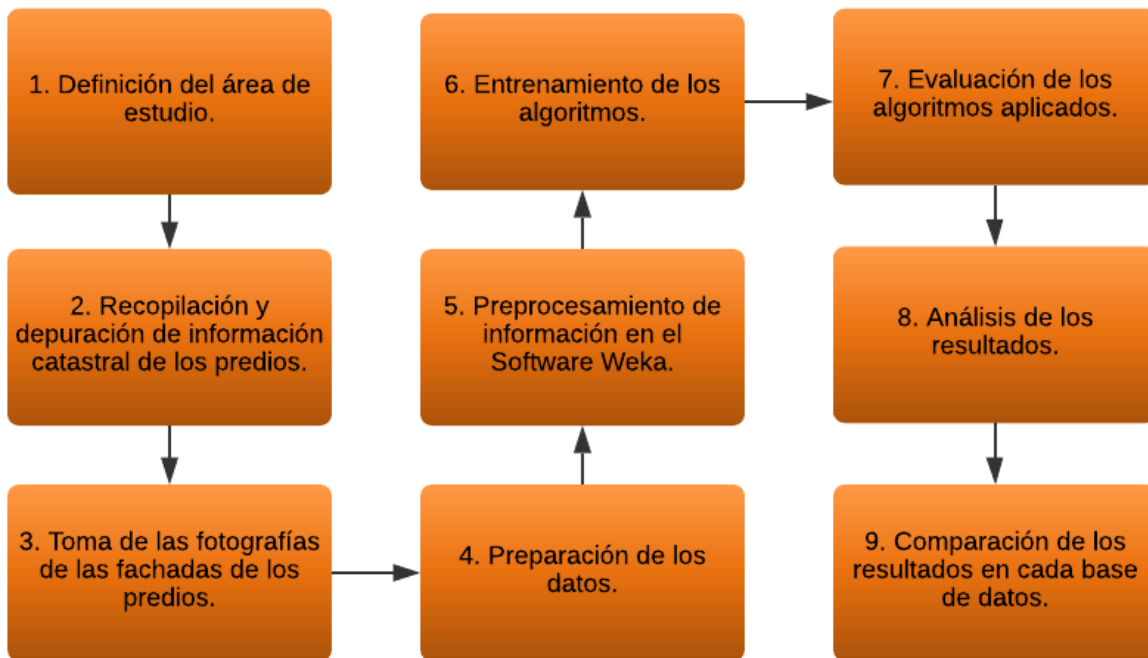


Figura 5-1. Diagrama metodología proyecto.
Fuente: Elaboración propia.

5.1 Área de estudio

Como se muestra en la Figura 5-2, el área de estudio comprende la UPZ 65 Arborizadora, la cual pertenece a la localidad de Ciudad Bolívar, localizada en el suroccidente de la ciudad de Bogotá. Cuenta con 14 barrios, limita al norte con las UPZ Apogeo, Timiza y Carvajal, al oriente con las UPZ Carvajal y Venecia, al sur con las UPZ Tunjuelito y San Francisco, y al occidente con las UPZ Ismael Perdomo y Jerusalem.

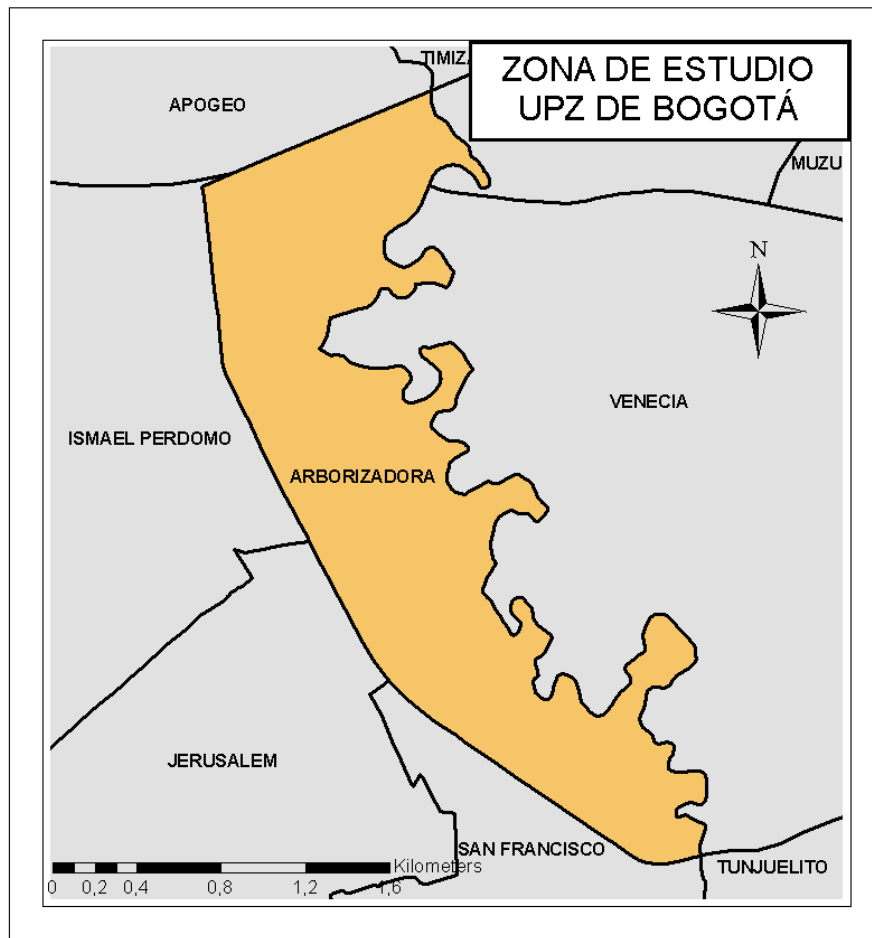


Figura 5-2. Zona de estudio UPZ 65 Arborizadora.
Fuente: Elaboración propia en Software Arcgis.

5.2 Recopilación y depuración de información catastral de los predios

Una vez seleccionada el área de estudio, se procede con la recopilación de información procedente de la UAECD. La base de datos disponible en formato Excel, contiene características asociadas a los predios y a su entorno (área de uso, sector, lote, manzana, matrícula inmobiliaria, chip, valor m² en los años 2014 y 2015, área, número de pisos, entre otros).

En razón al alcance del proyecto, es necesario descartar las construcciones de tipo comercial, industrial y residencial que estén bajo el régimen de propiedad horizontal, además eliminar datos que corresponden a lotes, atributos no válidos o nulos. Para ello, se adelanta una depuración a la base de datos que incluye entre otras cosas, la selección de los atributos que cuentan con un ángulo de visibilidad que permite captar la totalidad de la fachada en el visor de Mapas Bogotá con Street View. Lo anterior, permite garantizar la toma de las fotografías de las edificaciones y construir la base de datos para ejecutar el entrenamiento y validación de los algoritmos en Weka. Dicho esto, la base de datos final solamente contiene información relacionada a los predios de tipo residencial, sin régimen de propiedad horizontal. De allí se seleccionan 372 datos de un total de 28.000.

Como resultado de la depuración de la información, se seleccionan las variables: imagen, área de terreno, área construida en el año 2015, edad, número de pisos, estrato y la variable predictoria: puntaje. Finalmente, se añade una variable denominada: Acabado de fachada, la cual complementa la información contenida en la base de datos del UAECD y permite valorar cualitativamente el estado de la fachada de cada predio.

La descripción de cada variable se presenta a continuación:

- **Imagen:** corresponde a la fotografía de la fachada del predio, la cual se obtiene del visor de mapas Bogotá, con la extensión de Street View.
- **Área de terreno:** es cantidad de terreno con que cuenta el predio, en m².
- **Área construida:** corresponde a la cantidad de construcción habitacional con que cuenta el predio investigado, en m².
- **Edad:** define la edad de construcción del inmueble.
- **Número de pisos:** cantidad de pisos que tiene la construcción.
- **Estrato:** corresponde a la clasificación de las viviendas y/o los predios debido a la diferencia socioeconómica jerarquizada [23].
- **Puntaje:** es la medida cuantitativa según las características físicas de la construcción. En este estudio, corresponde al resultado final obtenido de realizar el proceso llevado a cabo por los algoritmos seleccionados. Este procedimiento implica reconocer, clasificar y calificar los tipos de edificaciones existentes en la base de datos.
- **Acabado de fachada:** para la definición de esta variable se empleó cada una de las fotografías de las fachadas obtenidas y se asignó un atributo que podía ser sencillo, regular, bueno o lujoso, de acuerdo a la observación realizada.

5.2.1 Nombre y tipo de las variables empleadas

En la base de datos las variables tienen asignados los tipos y encabezados que se muestran en la Tabla 5-1.

Nombre variable	Tipo de Variable	Encabezado en base de datos
Imagen	Imagen	IMAGEN
Acabado de fachada	Cadena de caracteres	ACAB_FACHADA
Área de terreno	Numérica	AREA TERRENO
Área construida al año 2015	Numérica	AREA CONSTRUIDA 2015
Edad	Numérica	EDAD
Número de pisos	Numérica	NUM_PIS
Estrato	Numérica	ESTRATO
Puntaje	Cadena de caracteres	PUNTAJE

Tabla 5-1. Nombre y tipo de variables.
Fuente: Elaboración propia.

5.3 Toma de las fotografías de las fachadas de los predios.

La fotografía de cada una de las fachadas para los 372 predios seleccionados se obtiene en el aplicativo de Mapas Bogotá y el visualizador Street View, el cual permite realizar la búsqueda de los predios por medio del identificador “CHIP”.

La Figura 5-3 y la Figura 5-4 ilustran la interfaz gráfica del aplicativo Mapas Bogotá y la extensión Street View.

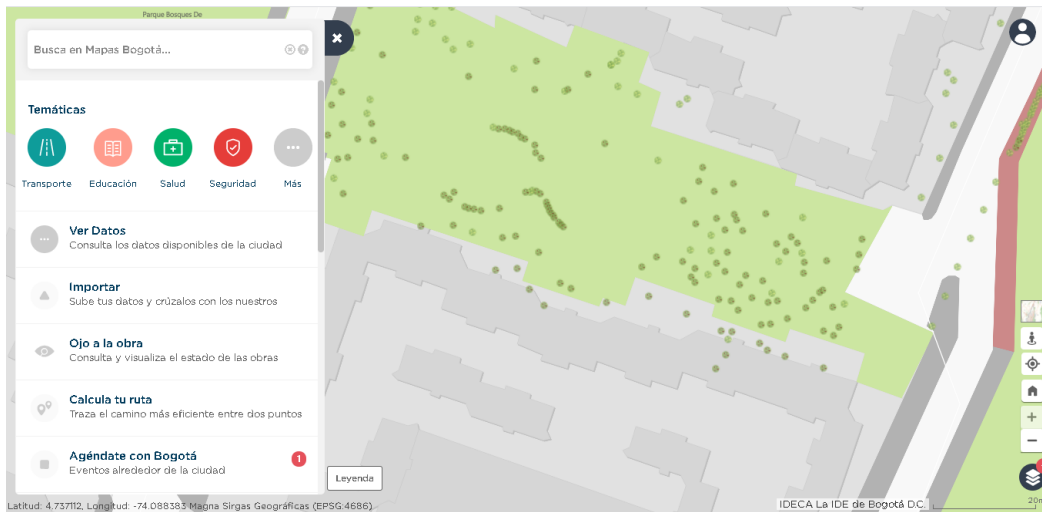


Figura 5-3. Aplicativo Mapas Bogotá.
Fuente: Página web Mapas Bogotá [46].



Figura 5-4. Visualizador Street View.
Fuente: Página web Mapas Bogotá [46].

Cada fotografía fue almacenada en una carpeta con el nombre del CHIP correspondiente, y en la base de datos con el mismo identificador adicionando la extensión .png. Lo anterior facilita la búsqueda que realiza el software Weka para la aplicar los filtros de preprocesamiento en las imágenes.

La Tabla 5-2 presenta una muestra de cada una de las tipificaciones de las construcciones trabajadas en el experimento.

<p>TIPO 1</p>	<p>TIPO 2</p>
	
<p>TIPO 3</p>	<p>TIPO 4</p>
	

Tabla 5-2. Fachadas empleadas para el estudio.
Fuente: Elaboración propia.

5.4 Preparación de los datos

El software Weka permite aplicar los modos de evaluación de algoritmos *Use Training set*, *Cross-Validation*, *Percentage Split* y *Supplied Test Set*. En particular, para este último es necesario particionar los datos en dos grupos: uno de entrenamiento y uno de validación. De ahí que de los 372 datos disponibles, se dispone el 80% (332) para entrenamiento y el 20% restante (40) para validación. Además, para garantizar la evaluación de las cuatro tipologías, los 40 datos de validación se dividen en 4 grupos, de 10 datos por cada tipo de construcción.

De igual manera, con el fin de realizar la comparación de los resultados de la validación con la misma cantidad de datos en cada método, para los demás modos de evaluación se emplean los mismos 332 datos seleccionados anteriormente.

Una vez generadas las bases de datos, se identifican las siguientes características de los atributos:

- **Imagen:** de acuerdo al total de predios comprendidos en el conjunto de datos, el registro fotográfico resultante es de 372 imágenes.
- **Área de terreno:** para el análisis desarrollado en este estudio se consideran áreas que van desde los 17 m² hasta los 283,42 m².

La base de datos asociada a la validación con el modo *Supplied Test Set*, presenta áreas de terreno desde los 34,5 m² hasta los 129,2 m².

- **Área construida:** el área construida de los predios empleados en la aplicación de los algoritmos va desde los 18 m² hasta los 494,4 m².

En cuanto a la base de datos para realizar la validación con el método *Supplied test set* el área construida va desde los 34,5 m² hasta los 129,2 m².

- **Edad:** la edad de los inmuebles está en un rango de 1 a 40 años.

El conjunto de datos asociado a la validación con el modo *Supplied Test Set*, presenta edades desde los 2 hasta los 41 años.

- **Número de pisos:** el número de pisos de las construcciones del área de estudio va desde 1 hasta 5 niveles.

- **Estrato:** en la Tabla 5-3 se presenta la distribución de los atributos según el estrato:

ESTRATO	Base de datos completa (332 datos)	Base de datos para validación (40 datos)	DESCRIPCIÓN
	NÚMERO DE DATOS	NÚMERO DE DATOS	
2	214	23	Como se observa, la base de datos completa presenta 214 datos en estrato 2 y 118 para estrato 3. De igual manera en el caso de la base de datos para validación, el número de datos para estrato 2 es 23 y para estrato 3 es de 17.
3	118	17	

Tabla 5-3. Datos por estrato.
Fuente: Elaboración propia.

- **Acabado de fachada:** la base de datos que contiene los 332 datos presenta la distribución que se muestra en la Tabla 5-4.

ACABADO DE FACHADA	NÚMERO DE DATOS	DESCRIPCIÓN
Sencillo	21	El acabado de fachada <i>sencillo</i> está asociado a 21 datos. El acabado de fachadas <i>regular</i> presenta 90 datos. En el caso del acabado de tipo <i>bueno</i> se observa que contiene la mayor parte de los datos con un total de 124. Y por último el acabado de fachada <i>lujoso</i> contiene 97 datos.
Regular	90	
Bueno	124	
Lujoso	97	

Tabla 5-4. Datos por variable Acabado de fachada.
Fuente: Elaboración propia.

Por otra parte, el conjunto de los 40 datos para realizar la validación por el modo *Supplied Test Set* se distribuye como se muestra en la Tabla 5-5:

ACABADO DE FACHADA	NÚMERO DE DATOS	DESCRIPCIÓN
Sencillo	10	Se observa que los acabados de fachada de tipo <i>regular</i> , <i>sencillo</i> , <i>bueno</i> y <i>lujoso</i> presentan una distribución de 10 datos por cada tipo de acabado.
Regular	10	
Bueno	10	
Lujoso	10	

Tabla 5-5. Datos por variable Acabado de fachada, base de datos Validación.
Fuente: Elaboración propia.

- **Puntaje:** los datos asociados a los resultados de predicción que corresponden a la variable puntaje, se modifican de acuerdo a la calificación de las construcciones de tipo residencial mediante la metodología definida por el IGAC [8] mostrada en la Tabla 5-6.

TIPO	RANGO DE PUNTAJE
1	1-10
2	11-28
3	29-46
4	47-64
5	65-82
6	83-100

Tabla 5-6. Rango de puntaje por tipo.
Fuente: IGAC [8].

De acuerdo con las características de los inmuebles pertenecientes a la zona de estudio, los rangos de calificación corresponden al tipo 1, tipo 2, Tipo 3 y tipo 4.

5.4.1 Duplicación de las bases de datos

Esta etapa consiste en crear una nueva base de datos que contiene la duplicación de todos los atributos y las variables: *área de terreno, área construida 2015, edad, número de pisos, estrato y puntaje*. Se garantiza entonces que una de las bases de datos contiene las variables *imagen y acabado de fachada*, y la otra cuenta con datos de tipo numérico. Lo anterior con el objeto de aplicar los mismos algoritmos en las bases de datos y realizar la comparación de los resultados obtenidos en cada caso.

Considerando lo anterior, en la Tabla 5-7 se muestra el conjunto de variables con el que cuenta cada base de datos:

VARIABLE	BASE DE DATOS 1	BASE DE DATOS 2
IMAGEN	Si	No
ACAB_FACHADA	Si	No
AREA DE TERRENO	Si	Si
AREA CONSTRUCCION	Si	Si
EDAD	Si	Si
NUM_PIS	Si	Si
ESTRATO	Si	Si
PUNTAJE	Si	Si

Tabla 5-7. Variables en las bases de datos.

Fuente: Elaboración propia.

5.5 Preprocesamiento de información en el software Weka

Para el cargue y lectura de las bases de datos en el software Weka se transforma el formato de los archivos de Excel (.xlsx) a delimitado por comas (.csv), como se muestra en la Figura 5-5 y Figura 5-6.

	A	B	C	D	E	F	G	H
1	IMAGEN	ACAB_FACHADA	AREA TERRENO	AREA CONSTRUIDA 2015	EDAD	NUM_PIS	ESTRATO	PUNTAJE
2	AAA0016RUWF.png	sencillo	55	113	20	2	2	Tipo 2
3	AAA0016RUYX.png	sencillo	104	157,8	19	2	2	Tipo 2
4	AAA0016RUZM.png	sencillo	49,6	158,1	21	3	2	Tipo 2
5	AAA0016RWBR.png	regular	55,5	113,76	20	2	2	Tipo 3
6	AAA0016RWCX.png	sencillo	229,24	151,73	34	3	2	Tipo 2
7	AAA0016RWEA.png	sencillo	54,3	111	21	2	2	Tipo 2
8	AAA0016RWFT.png	regular	53,8	223,01	29	4	2	Tipo 3
9	AAA0016RWHY.png	regular	51	199,8	34	4	2	Tipo 3
10	AAA0016RWKL.png	sencillo	45,7	116,2	34	3	2	Tipo 2
11	AAA0016RWLW.png	sencillo	48	48	29	1	2	Tipo 2
12	AAA0016RWMS.png	sencillo	137,9	282,27	4	2	2	Tipo 2
13	AAA0016RZSY.png	regular	283,42	494,4	15	2	2	Tipo 3
14	AAA0016RZTD.png	sencilla	72	18	14	1	2	Tipo 1
15	AAA0016RZWW.png	pobre	93	45,8	28	1	2	Tipo 1
16	AAA0016RZXS.png	sencillo	115,1	95,51	14	2	2	Tipo 2

Figura 5-5. Base de datos en formato .xlsx.

Fuente: Elaboración propia.

Imagen	ACAB_FACHADA	AREA_TERRENO	AREA_CONSTRUIDA	2015	EDAD	NUM_PIS	ESTRATO	PUNTAJE
AAA0016RUWF.png	regular	55	113	20	2	2	Tipo 2	
AAA0016RUYX.png	regular	104	157.8	19	2	2	Tipo 2	
AAA0016RUZM.png	regular	49.6	158.1	21	3	2	Tipo 2	
AAA0016RWBR.png	bueno	55.5	113.76	20	2	2	Tipo 3	
AAA0016RWCX.png	regular	229.24	151.73	34	3	2	Tipo 2	
AAA0016RWEA.png	regular	54.3	111	21	2	2	Tipo 2	
AAA0016RWFT.png	bueno	53.8	223.01	29	4	2	Tipo 3	
AAA0016RWHY.png	bueno	51	199.8	34	4	2	Tipo 3	
AAA0016RWKL.png	regular	45.7	116.2	34	3	2	Tipo 2	
AAA0016RWLW.png	regular	48	48	29	1	2	Tipo 2	
AAA0016RWM5.png	regular	137.9	282.27	4	2	2	Tipo 2	
AAA0016RZSY.png	bueno	283.42	494.4	15	2	2	Tipo 3	
AAA0016RZTD.png	sencillo	72	18	14	1	2	Tipo 1	

Figura 5-6. Base de datos en formato .csv.

Fuente: Elaboración propia.

Una vez cargados los archivos en el software y de acuerdo con el manual Weka [39], se aplican los siguientes filtros de preprocesamiento:

1. **NominalToString:** transforma los valores nominales de un atributo en una cadena de caracteres.
2. **SimpleColorHistogramFilter:** entrega los histogramas de las imágenes.

De los 63 histogramas resultantes de la aplicación del filtro, se observa que en algunos casos los resultados son sesgados hacia la derecha o izquierda, ocasionando un bajo rendimiento en la clasificación llevada a cabo por los algoritmos. Como alternativa para estos casos, se utiliza el filtro *MathExpression* que realiza una operación matemática logarítmica logrando así un ajuste con tendencia normal en las frecuencias absolutas con respecto a los valores de los tonos asociados a las imágenes [47]. A continuación, se presenta un ejemplo:

Histograma:

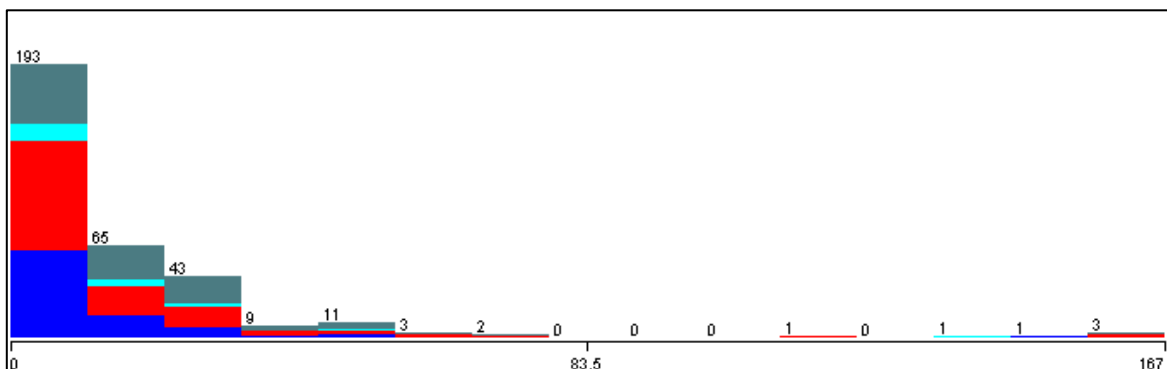


Figura 5-7. Histograma de datos.

Fuente: Elaboración propia.

Como se observa en la Figura 5-7, la información presenta una frecuencia absoluta sesgada hacia los valores de tono de la izquierda. Además se observan las tipificaciones que están asociadas a cada valor.

Aplicación del filtro *MathExpression*:

En la se presenta la aplicación del filtro *MathExpression*. Como se observa en la Figura 5-8, el procedimiento realiza la redistribución de la frecuencia absoluta con respecto a las clases, dando como resultado una distribución normal en los valores de los tonos encontrados en las imágenes.

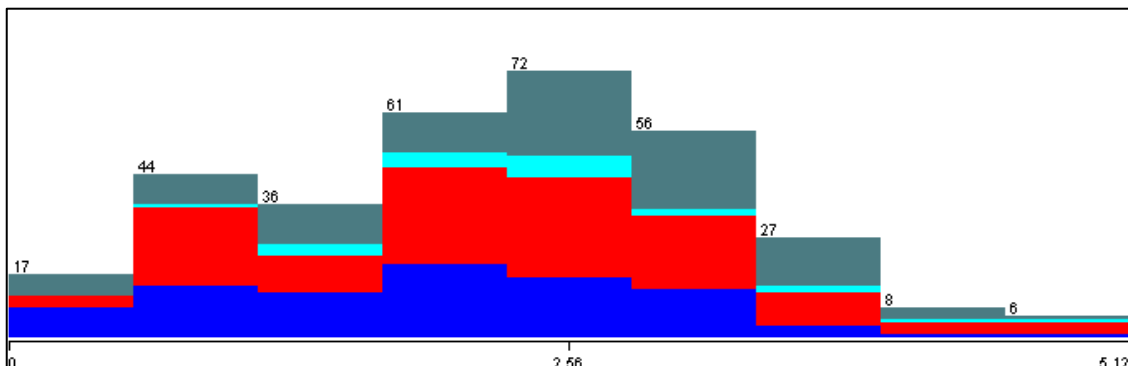


Figura 5-8. Histograma de los datos con filtro *MathExpression*.
Fuente: Elaboración propia.

3. **StringToNominal**: convierte un atributo de tipo cadena a un tipo nominal.
4. **NumericToNominal**: convierte un atributo de tipo numérico a tipo nominal.

En la base de datos de validación se aplican los mismos filtros descritos anteriormente y el archivo resultante se guarda con la extensión solicitada por Weka para realizar la evaluación por el modo *Supplied Test Set* (.arff).

Finalmente, a la base de datos que no contiene imágenes, se aplica únicamente el filtro *NumericToNominal*, considerando que únicamente se cuenta con datos numéricos.

5.6 Entrenamiento de los algoritmos

Para realizar el entrenamiento se cuenta con una base de datos híbrida con información alfanumérica y de imágenes. Así, se tienen datos adicionales a las fotografías que describen las características de los predios de la zona de estudio.

De la revisión bibliográfica y de acuerdo a la naturaleza de las bases de datos, para el entrenamiento y validación se aplican los algoritmos presentados en la Tabla 5-8.

	ALGORITMO	DESCRIPCIÓN
BAYES	BayesNet	Aprendizaje mediante redes bayesianas.
	NaiveBayes	Clasificador de probabilidades.
	NaiveBayesUpdateable	Basado en el clasificador NaiveBayes. Aprende una instancia a la vez.
FUNCIONES	SMO	Realiza la optimización secuencial para la clasificación de vectores.
LAZY	IBK	Clasificador k- vecino más cercano.
	KSTAR	Aplica la función de distancia generalizado al aprendizaje realizado con el vecino más cercano
	LWL	Se basa en el aprendizaje localmente pesado.

ÁRBOLES DE DECISIÓN	ALGORITMO	DESCRIPCIÓN
	DecisionStump	Árboles de decisión de un nivel.
	J48	Algoritmo basado en el árbol de decisión C4.5.
	LMT	Árboles de modelo logístico.
	NBTree	Las hojas del árbol presentan clasificadores NaiveBayes en las hojas.
	RandomForest	Realiza árboles aleatorios.
	RandomTree	En cada nodo existe un conjunto aleatorio de características.
	REPTree	Realiza la poda en el árbol para la reducción de errores.

Tabla 5-8. Algoritmos empleados en el experimento.
Fuente: revista de Ciencias Informáticas [33].

Una vez obtenidos los resultados de la predicción a través de cada uno de los algoritmos, se seleccionan los que presenten la mayor cantidad de instancias correctamente clasificadas.

5.7 Evaluación de los algoritmos

Los algoritmos son evaluados de acuerdo a las alternativas que el software Weka ofrece. Esta etapa se constituye como un paso clave en el análisis del desempeño que tuvo el algoritmo en el aprendizaje.

En la Figura 5-9 se presentan las alternativas que ofrece Weka para realizar la partición de datos empleados en el entrenamiento y en la validación.

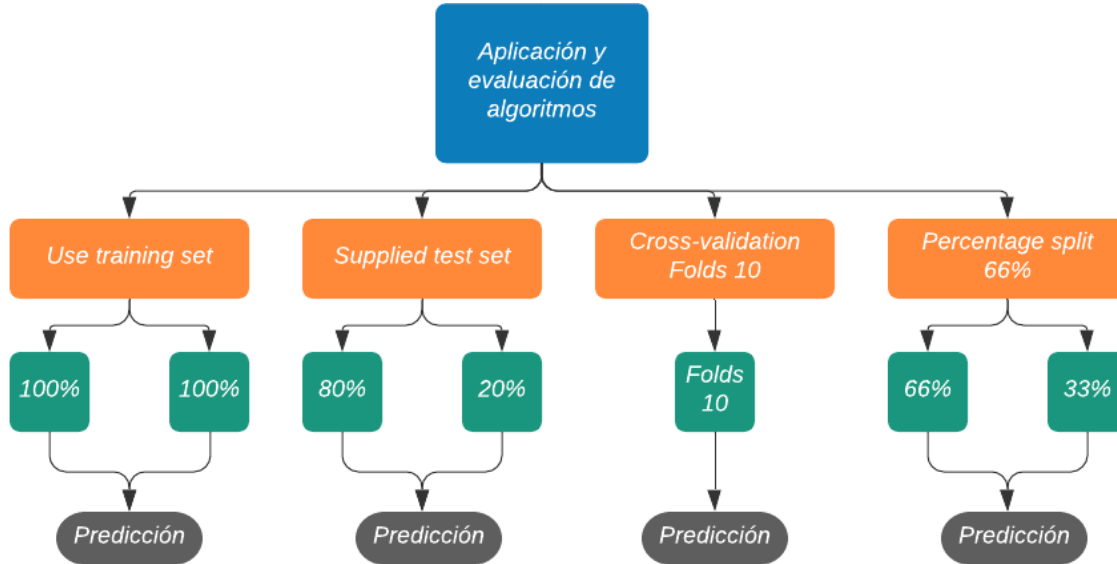


Figura 5-9. Partición de datos en los modos de prueba.
Fuente: Elaboración propia.

La cantidad de datos que se emplearon para realizar el entrenamiento y validación para las dos bases de datos (con imágenes y sin imágenes), se muestran en la Tabla 5-9. Ya que el objetivo es comparar el funcionamiento de los algoritmos, se emplea el mismo número de datos en los dos casos.

Algoritmo	Método de evaluación	Entrenamiento		Validación	
		Porcentaje (%)	Cantidad datos	Porcentaje (%)	Cantidad datos
	Use training set	100%	332	100%	332
	Supplied test set	80%	332	20%	40
	Percentage split	66%	219	33%	113
	Cross- Validation	Folds			
		10			

Tabla 5-9. Cantidad de datos empleados en cada modo de evaluación.
Fuente: Elaboración propia.

5.8 Análisis de los resultados

En este apartado se seleccionan los algoritmos con la mayor cantidad de instancias correctamente clasificadas, y a su vez el que presente el mayor resultado de clasificación. De este último se documentan los errores estadísticos generales y puntuales, así como las matrices de confusión.

5.8.1 Resultados obtenidos de la base de datos con fotografías

En la Tabla 5-10 se presentan los resultados obtenidos en la base de datos con imágenes.

RESULTADOS DE VALIDACIÓN DE LA BASE DE DATOS CON FOTOGRAFÍAS					
ALGORITMO	MODO DE PRUEBA	INSTANCIAS CORRECTAMENTE CLASIFICADAS		INSTANCIAS INCORRECTAMENTE CLASIFICADAS	
		INSTANCIAS	CANTIDAD DE DATOS	INSTANCIAS	CANTIDAD DE DATOS
BayesNet	Use training set	100,00%	332	0,00%	0
	Supplied test set	90,00%	36	10,00%	4
	Cross-validation (Folds 10)	98,19%	326	1,80%	6
	Percentage split 66%	97,34%	110	2,65%	3
Random Forest	Use training set	100,00%	332	0,00%	0
	Supplied test set	90,00%	36	10,00%	4
	Cross-validation (Folds 10)	94,87%	310	6,62%	22
	Percentage split 66%	92,03%	104	7,96%	9
IBK K=4	Use training set	97,89%	325	2,10%	7
	Supplied test set	85,00%	34	15,00%	6
	Cross-validation (Folds 10)	96,98%	322	3,01%	10
	Percentage split 66%	95,57%	108	4,42%	5

Tabla 5-10. Resultados obtenidos en la base de datos con fotografías.
Fuente: Elaboración propia.

Así mismo, los gráficos presentados en la Tabla 5-11 permiten observar una comparación del desempeño obtenido en el aprendizaje de los algoritmos empleando cada modo de evaluación (*Use Training Set*, *Supplied Test Set*, *Cross-Validation* y *Percentage Split*).

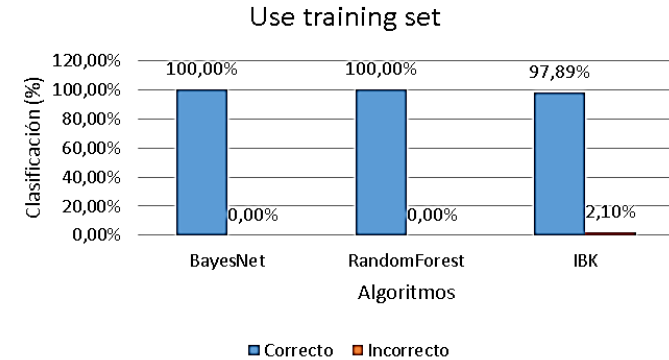
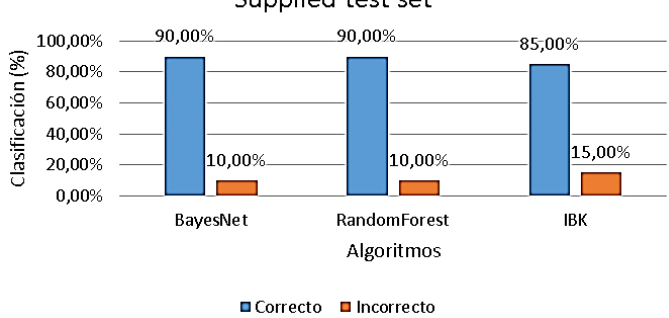
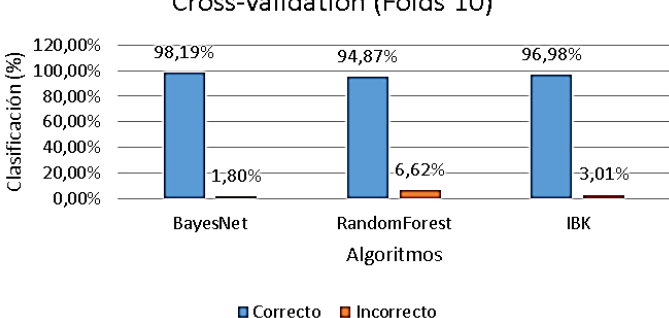
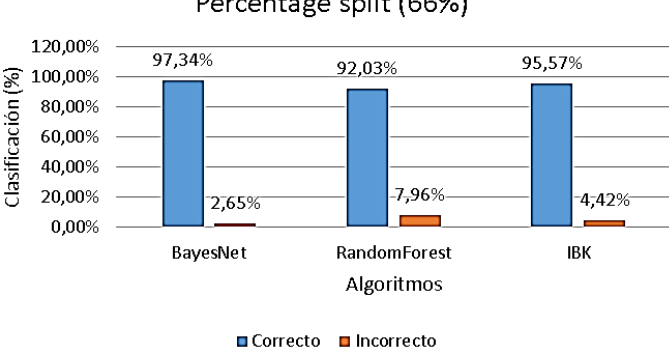
Comparación de los resultados obtenidos en los algoritmos	Análisis												
<p style="text-align: center;">Use training set</p>  <table border="1" data-bbox="235 378 901 735"> <caption>Use training set - Classification Results</caption> <thead> <tr> <th>Algoritmo</th> <th>Correcto (%)</th> <th>Incorrecto (%)</th> </tr> </thead> <tbody> <tr> <td>BayesNet</td> <td>100,00%</td> <td>0,00%</td> </tr> <tr> <td>RandomForest</td> <td>100,00%</td> <td>0,00%</td> </tr> <tr> <td>IBK</td> <td>97,89%</td> <td>2,10%</td> </tr> </tbody> </table>	Algoritmo	Correcto (%)	Incorrecto (%)	BayesNet	100,00%	0,00%	RandomForest	100,00%	0,00%	IBK	97,89%	2,10%	<p>Para el método de evaluación <i>Use Training Set</i> los algoritmos BayesNet y RandomForest obtuvieron un 100% de instancias clasificadas correctamente. En cuanto al algoritmo IBK el resultado es 97,89%, por lo que la diferencia es baja.</p>
Algoritmo	Correcto (%)	Incorrecto (%)											
BayesNet	100,00%	0,00%											
RandomForest	100,00%	0,00%											
IBK	97,89%	2,10%											
<p style="text-align: center;">Supplied test set</p>  <table border="1" data-bbox="235 777 901 1092"> <caption>Supplied test set - Classification Results</caption> <thead> <tr> <th>Algoritmo</th> <th>Correcto (%)</th> <th>Incorrecto (%)</th> </tr> </thead> <tbody> <tr> <td>BayesNet</td> <td>90,00%</td> <td>10,00%</td> </tr> <tr> <td>RandomForest</td> <td>90,00%</td> <td>10,00%</td> </tr> <tr> <td>IBK</td> <td>85,00%</td> <td>15,00%</td> </tr> </tbody> </table>	Algoritmo	Correcto (%)	Incorrecto (%)	BayesNet	90,00%	10,00%	RandomForest	90,00%	10,00%	IBK	85,00%	15,00%	<p>En el modo <i>Supplied Test Set</i> el mejor desempeño lo tienen los algoritmos BayesNet y RandomForest con un 90%, para el caso de IBK aunque el resultado es bueno, se reduce a un 15% el porcentaje de instancias correctamente clasificadas, alcanzando un 85%.</p>
Algoritmo	Correcto (%)	Incorrecto (%)											
BayesNet	90,00%	10,00%											
RandomForest	90,00%	10,00%											
IBK	85,00%	15,00%											
<p style="text-align: center;">Cross-validation (Folds 10)</p>  <table border="1" data-bbox="235 1123 901 1438"> <caption>Cross-validation (Folds 10) - Classification Results</caption> <thead> <tr> <th>Algoritmo</th> <th>Correcto (%)</th> <th>Incorrecto (%)</th> </tr> </thead> <tbody> <tr> <td>BayesNet</td> <td>98,19%</td> <td>1,80%</td> </tr> <tr> <td>RandomForest</td> <td>94,87%</td> <td>6,62%</td> </tr> <tr> <td>IBK</td> <td>96,98%</td> <td>3,01%</td> </tr> </tbody> </table>	Algoritmo	Correcto (%)	Incorrecto (%)	BayesNet	98,19%	1,80%	RandomForest	94,87%	6,62%	IBK	96,98%	3,01%	<p>En el modo de evaluación <i>Cross-Validation</i> se destaca el algoritmo BayesNet por obtener un 98,19% de instancias clasificadas correctamente. El rendimiento más bajo se da en el algoritmo RandomForest con un porcentaje de 94,87%.</p>
Algoritmo	Correcto (%)	Incorrecto (%)											
BayesNet	98,19%	1,80%											
RandomForest	94,87%	6,62%											
IBK	96,98%	3,01%											
<p style="text-align: center;">Percentage split (66%)</p>  <table border="1" data-bbox="235 1480 901 1827"> <caption>Percentage split (66%) - Classification Results</caption> <thead> <tr> <th>Algoritmo</th> <th>Correcto (%)</th> <th>Incorrecto (%)</th> </tr> </thead> <tbody> <tr> <td>BayesNet</td> <td>97,34%</td> <td>2,65%</td> </tr> <tr> <td>RandomForest</td> <td>92,03%</td> <td>7,96%</td> </tr> <tr> <td>IBK</td> <td>95,57%</td> <td>4,42%</td> </tr> </tbody> </table>	Algoritmo	Correcto (%)	Incorrecto (%)	BayesNet	97,34%	2,65%	RandomForest	92,03%	7,96%	IBK	95,57%	4,42%	<p>Por último, el algoritmo BayesNet obtuvo el más alto porcentaje de instancias clasificadas correctamente, alcanzando el 97,34% en el método de evaluación <i>Percentage Split</i>.</p>
Algoritmo	Correcto (%)	Incorrecto (%)											
BayesNet	97,34%	2,65%											
RandomForest	92,03%	7,96%											
IBK	95,57%	4,42%											

Tabla 5-11. Comparación de los resultados obtenidos en los algoritmos.

Fuente: Elaboración propia.

Como se observa en los resultados presentados anteriormente, el algoritmo *BayesNet* logra los porcentajes más altos de instancias correctamente clasificadas en todos los métodos de evaluación. Se tiene el 100% de aciertos en el modo *Use Training Set*, el 90% para el modo *Supplied Test Set*. En el caso de *Cross- Validation* el resultado es 98,19% y en el método *Percentage Split* el total es 97,34%.

El análisis de los siguientes apartados se focaliza en los resultados encontrados en el algoritmo *BayesNet* y parámetros como el número de instancias correcta e incorrectamente clasificadas, los errores estadísticos, el estadístico Kappa y los errores de predicción en cada clasificación.

5.8.1.1 Análisis de los errores del algoritmo BayesNet

En la Tabla 5-12 y en la Figura 5-10. Resultados estadísticos del algoritmo BayesNet. se presentan los resultados estadísticos (*estadístico Kappa, error medio absoluto, error cuadrático medio, error absoluto relativo y error de raíz cuadrada relativo*), proporcionados en cada modo evaluativo del algoritmo BayesNet.

ALGORITMO BAYESNET	USE TRAINING SET	SUPPLIED TEST SET	CROSS-VALIDATION	PERCENTAGE SPLIT
ESTADISTICO KAPPA	1,0000	0,8667	0,9741	0,9626
ERROR MEDIO ABSOLUTO	0,0024	0,0647	0,0248	0,0286
ERROR CUADRATICO MEDIO	0,0173	0,2056	0,0881	0,0898
ERROR ABSOLUTO RELATIVO	0,0069	0,1724	0,0710	0,9734
ERROR DE RAÍZ CUADRADA RELATIVO	0,0414	0,4595	0,2109	0,0265

Tabla 5-12. Errores estadísticos obtenidos en el algoritmo BayesNet.
Fuente: Elaboración propia.

Los resultados obtenidos anteriormente se presentan de forma gráfica en la Figura 5-10.

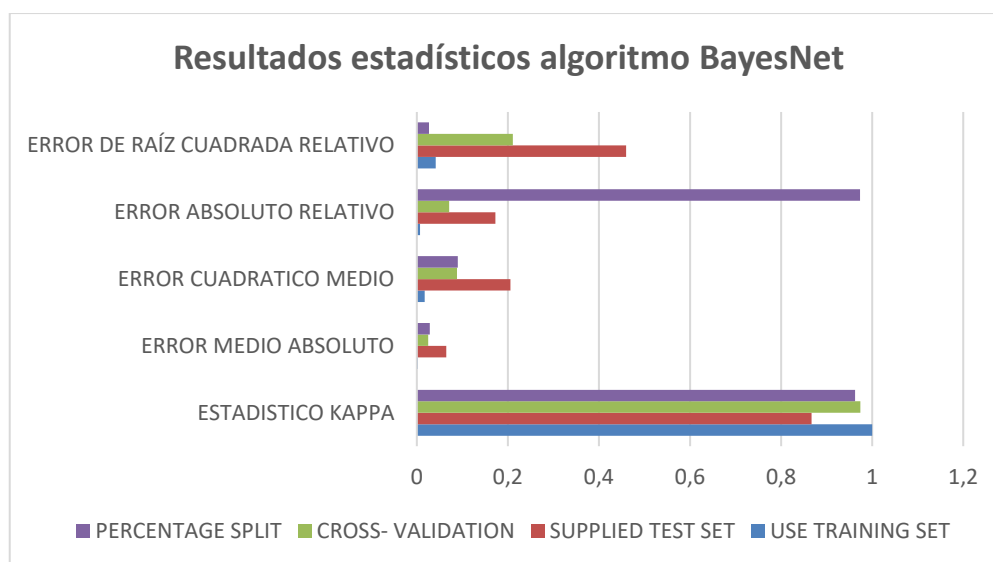


Figura 5-10. Resultados estadísticos del algoritmo BayesNet.
Fuente: Elaboración propia.

A propósito de los resultados, se encuentra que el modo *Use Training Set* obtiene los errores más bajos con 0,0024 en el error medio absoluto, 0,0173 en el error cuadrático medio, 0,69% en el error absoluto relativo, 4,14% en el error de raíz cuadrada relativo y un estadístico Kappa de 1.

En contraste, el método *Supplied Test Set* arroja errores de 0,0647 en el error medio absoluto, 0,2048 en el error cuadrático medio, 17,24% en el error absoluto relativo, 45,95% en el error de raíz cuadrada relativo y un estadístico Kappa de 0,86. Según los resultados encontrados, aunque sean los mayores errores observados, son errores bajos. Además, de acuerdo con el estadístico Kappa se tiene una alta correlación entre la predicción obtenida y los atributos de la base de datos.

Pese a que el error absoluto relativo del modo *Percentage Split* es alto (97,34%), los demás errores de este método son bajos, teniendo 0,02 en el error medio absoluto, 0,08 en el error cuadrático medio y 2,65% en el error de raíz cuadrada relativo. Así mismo la correlación en los datos es alta ya que el resultado del estadístico Kappa es de 0,96.

Los errores obtenidos por el método *Cross- Validation* son bajos en el error absoluto relativo con 7,10%, en el error cuadrático medio con 0,08 y en el error medio absoluto con 0,02. Igualmente, el estadístico Kappa indica una alta correlación con los datos.

En general los errores estadísticos son bajos, por lo que dan cuenta del buen desempeño del algoritmo BayesNet.

5.8.1.1.1 Errores de predicción en cada modo de prueba

A continuación se presentan los errores resultantes en cada una de las predicciones realizadas por el algoritmo BayesNet. Estos resultados indican la pureza con que son clasificadas las instancias, las cuales toman valores en un orden de 0 a 1, siendo 0 el menor y 1 el mayor grado de pureza [44].

- **Modo de prueba Use Training Set**

En la Figura 5-11 se presentan los resultados para cada predicción realizada, mediante el modo evaluativo *Use Training Set*.

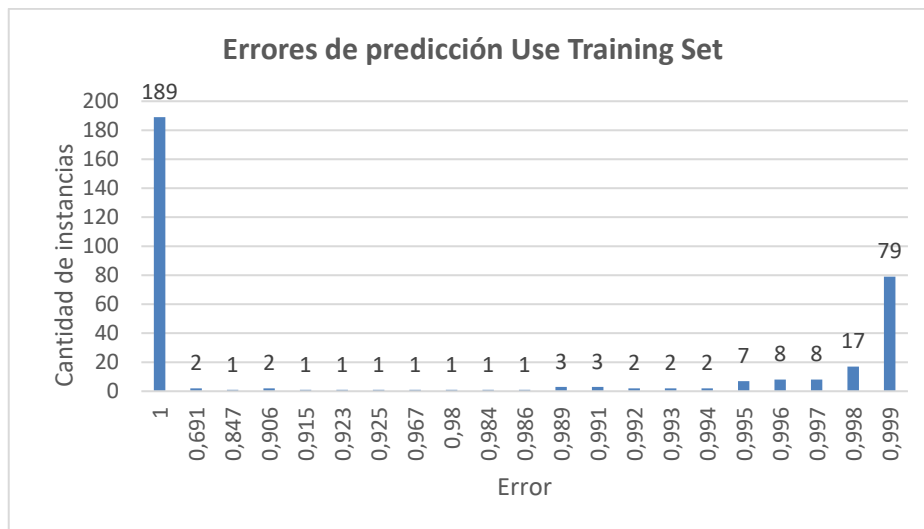


Figura 5-11. Errores de predicción modo de prueba Use Training Set.
Fuente: Elaboración propia.

Como se puede observar, la mayoría de las instancias se clasificaron correctamente (189) con un error de predicción igual a 1. Los errores de predicción más bajos fueron obtenidos tan sólo en 2 instancias con un total de 0,691. Del restante, el mayor número de instancias son clasificadas con errores muy bajos que son cercanos a 1 (0,998 y 0,999).

En el Anexo 1 se presentan los resultados obtenidos en la clasificación realizada en cada una de las instancias mediante el modo de evaluación *Use Training Set*. Es necesario resaltar que la totalidad de los campos son clasificados de forma correcta.

- **Modo de prueba Supplied Test Set**

Los resultados obtenidos para cada una de las predicciones realizadas, mediante el modo evaluativo *Supplied Test Set* se presentan en la Figura 5-12.

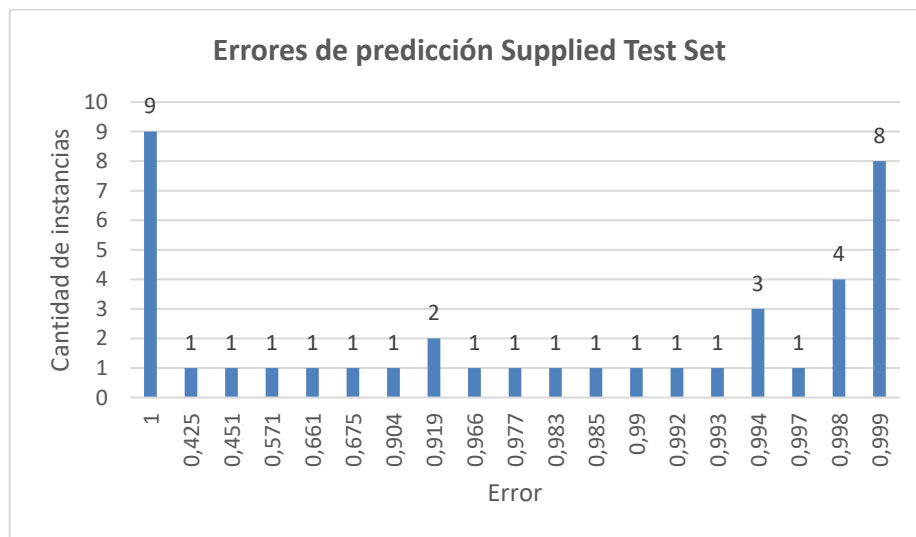


Figura 5-12. Errores de predicción modo de prueba Supplied Test Set.
Fuente: Elaboración propia.

Los errores de predicción en su mayoría están clasificados como 1 o muy cercano a este valor (0,999). Lo anterior significa que casi todas las instancias están clasificados correctamente.

En el Anexo 2 se presentan los resultados obtenidos en la clasificación realizada en cada una de las instancias mediante el modo evaluativo *Supplied Test Set*. Es importante destacar que los campos resaltados en rojo corresponden a instancias incorrectamente clasificadas.

- **Modo de prueba Percentage Split**

En la Figura 5-13 se presentan los resultados entregados por el software, en cada una de las predicciones realizadas mediante el modo evaluativo *Percentage Split*.

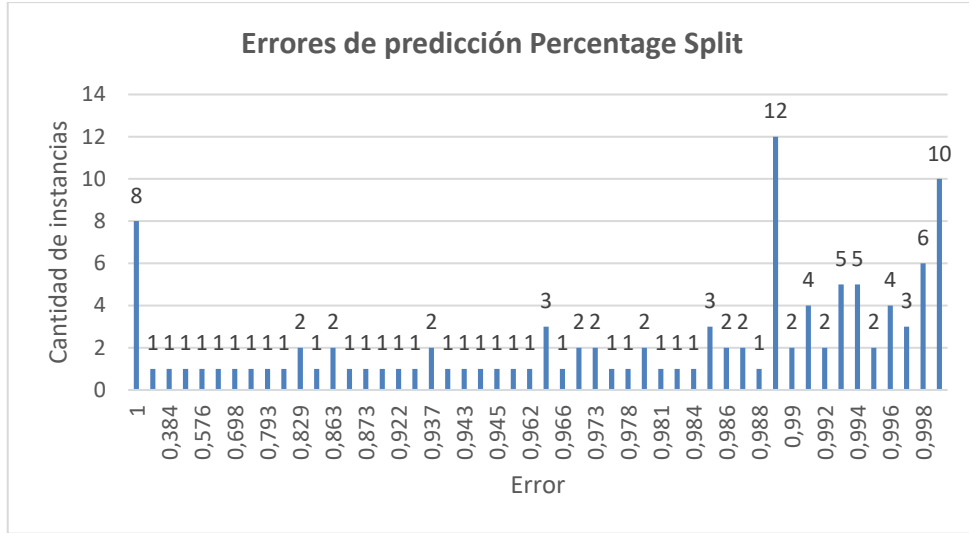


Figura 5-13. Errores de predicción modo de prueba Percentage Split.
Fuente: Elaboración propia.

En el modo de evaluación *Percentage Split* se obtienen buenos resultados, los cuales oscilan entre 0,9 y 1 en la mayoría de las instancias clasificadas.

En el Anexo 3 se presentan los resultados obtenidos en la clasificación realizada en cada una de las instancias mediante este modo evaluativo (*Percentage Split*). Es preciso señalar que los campos que se encuentran resaltados en rojo son instancias incorrectamente clasificadas.

- **Modo de prueba Cross- Validation**

Los resultados obtenidos para cada una de las predicciones realizadas, mediante el modo evaluativo *Cross- Validation* se presentan en la Figura 5-14.

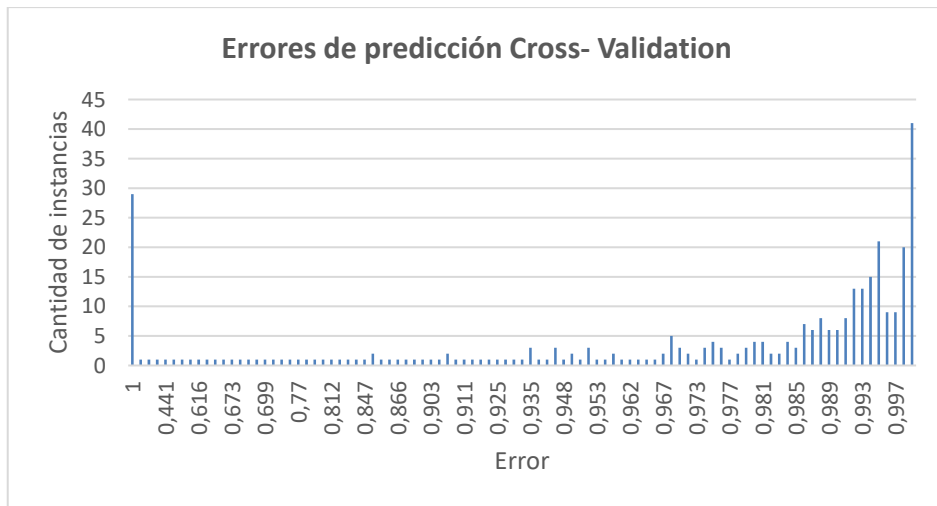


Figura 5-14. Errores de predicción modo de prueba Cross- Validation.
Fuente: Elaboración propia.

En las particiones realizadas en el modo de prueba *Cross- Validation*, se puede observar que los resultados de clasificación en los conjuntos resultantes están entre 0,9 y 1 para la mayor parte de las instancias. Según lo anterior, la mayoría de las instancias son correctamente clasificadas.

En el Anexo 4 se presentan los resultados de la clasificación realizada en cada una de las instancias evaluadas con el modo de prueba *Cross- Validation*. Cabe señalar que los campos resaltados en rojo son instancias incorrectamente clasificadas.

Finalmente, de acuerdo con los resultados la mayor parte de las clasificaciones realizadas en cada modo de prueba (*Use Training Set, Supplied Test Set, Percentage Split, Cross- Validation*), muestran errores de predicción en valores iguales o muy cercanos a 1, por lo que en general se alcanza un alto grado de pureza.

5.8.1.1.2 Resultados de las matrices de confusión

Las matrices de confusión facilitan la interpretación de los errores en la clasificación final del algoritmo. En efecto, las instancias que son clasificadas de forma correcta se encuentran en la diagonal de la matriz, los valores restantes son incorrectos. Es preciso señalar que los encabezados de las columnas son las clases, y las filas son el resultado de la clasificación [48]. De modo que la manera adecuada de interpretar los resultados es verificando la correspondencia de los encabezados de las columnas con las filas. Si no existe tal relación, se determina la cantidad y las clases en que fueron clasificadas las instancias [44].

En la Tabla 5-13 se explicarán las matrices de confusión resultantes en cada modo de evaluación del algoritmo BayesNet.

Matriz de confusión para cada modo de evaluación	Análisis
<p style="text-align: center;">Use training set</p> <pre> === Confusion Matrix === a b c d <-- classified as 90 0 0 0 a = Tipo 2 0 124 0 0 b = Tipo 3 0 0 21 0 c = Tipo 1 0 0 0 97 d = Tipo 4 </pre>	<p>Los resultados obtenidos muestran que la diagonal de predicciones está rodeada con valores de cero. Según lo anterior, es posible establecer que las instancias están clasificadas correctamente en su totalidad.</p>
<p style="text-align: center;">Supplied test set</p> <pre> === Confusion Matrix === a b c d <-- classified as 10 0 0 0 a = Tipo 2 0 10 0 0 b = Tipo 3 1 1 6 2 c = Tipo 1 0 0 0 10 d = Tipo 4 </pre>	<p>Como se observa en las filas 1, 2 y 4, el total de las instancias de Tipo 2, 3 y 4 son clasificadas correctamente. Por otro lado, las instancias incorrectas se encuentran asociadas al Tipo 1. Se verifica en la fila 3 que de las 10 instancias, se registra 1 dato clasificado como Tipo 2, 1 dato como Tipo 3 y 2 datos como Tipo 4.</p>

Matriz de confusión para cada modo de evaluación	Análisis
<p style="text-align: center;">Cross- Validation</p> <pre> === Confusion Matrix === a b c d <-- classified as 87 1 0 2 a = Tipo 2 0 124 0 0 b = Tipo 3 0 0 21 0 c = Tipo 1 1 0 2 94 d = Tipo 4 </pre>	<p>De acuerdo con los resultados de las filas 2 y 3 presentadas en la matriz de confusión, las instancias asociadas a los Tipos 3 y 1 (124 y 21), son clasificadas correctamente.</p> <p>Así mismo, las instancias incorrectas se encuentran asociadas a los Tipo 2 y Tipo 4, ubicados en las filas 1 y 4 respectivamente. En el Tipo 2 se tiene 1 calificación como Tipo 3, y en el Tipo 4 se tienen 2 calificaciones como Tipo 1.</p>
<p style="text-align: center;">Percentage split</p> <pre> === Confusion Matrix === a b c d <-- classified as 32 0 0 1 a = Tipo 2 0 39 0 0 b = Tipo 3 1 0 8 1 c = Tipo 1 0 0 0 31 d = Tipo 4 </pre>	<p>Según los resultados obtenidos en las filas 2 y 4, las instancias asociadas a los Tipos 3 y 4 (39 y 31) son clasificadas de forma correcta.</p> <p>Además, se verifica en las filas 1 y 3 que las instancias incorrectas se encuentran asociadas a los Tipo 2 y Tipo 1. En el caso del Tipo 2 se tiene una instancia de Tipo 4. Para el Tipo 1 existe 1 instancia clasificada como Tipo 2, y 1 instancia como Tipo 4.</p>

Tabla 5-13. Resultados matrices de confusión.
Fuente: Elaboración propia.

Los resultados obtenidos en las matrices de confusión dan cuenta del aprendizaje adquirido mediante la aplicación del algoritmo BayesNet. La máxima cantidad de errores de clasificación obtenida en uno de los modos de evaluación fue 6, la cual corresponde a la modalidad *Cross-Validation*. Si se compara con el total de instancias (332), este resultado es muy bajo.

6 ANÁLISIS COMPARATIVO DE LOS RESULTADOS OBTENIDOS EN BASES DE DATOS CON IMÁGENES Y SIN IMÁGENES

En este capítulo se presenta el paso número 9 de la propuesta metodológica, el cual corresponde a la comparación de los resultados obtenidos para cada base de datos (con imágenes y sin imágenes). De modo que se aplican los algoritmos que demostraron tener un mejor desempeño en el ejercicio que incluyó imágenes. Dicho esto, se selecciona: *BayesNet*, *IBK* y *RandomForest*₁. Los resultados se presentan en el siguiente apartado:

6.1 Resultados obtenidos en la base de datos sin fotografías

En la Tabla 6-1 se presentan los resultados obtenidos en la base de datos sin imágenes.

RESULTADOS DE VALIDACIÓN DE LA BASE DE DATOS SIN FOTOGRAFÍAS					
ALGORITMO	MODO DE PRUEBA	INSTANCIAS CORRECTAMENTE CLASIFICADAS		INSTANCIAS INCORRECTAMENTE CLASIFICADAS	
		INSTANCIAS	CANTIDAD DE DATOS	INSTANCIAS	CANTIDAD DE DATOS
BayesNet	Use training set	92,46%	307	7,53%	25
	Supplied test set	65,00%	26	35,00%	14
	Cross-validation (Folds 10)	56,02%	186	43,97%	146
	Percentage split 66%	51,32%	58	48,67%	55
Random Forest	Use training set	100,00%	332	0,00%	0
	Supplied test set	50,00%	20	50,00%	20
	Cross-validation (Folds 10)	51,50%	171	48,49%	161
	Percentage split 66%	46,01%	52	53,98%	61
IBK K=4	Use training set	67,46%	224	32,53%	108
	Supplied test set	60,00%	24	40,00%	16
	Cross-validation (Folds 10)	60,24%	200	39,75%	132
	Percentage split 66%	64,60%	73	35,39%	40

Tabla 6-1. Resultados obtenidos en la base de datos sin fotografías.

Fuente: Elaboración propia.

Como se observa, el algoritmo *BayesNet* alcanza el porcentaje más alto de instancias correctamente clasificadas en todos los modos de prueba. Conviene señalar que el mejor desempeño es obtenido mediante el método evaluativo *Use Training Set*, alcanzando un 92,46% de aciertos en la clasificación. De igual manera, para el método de evaluación *Supplied Test Set* se obtuvo un 65% de instancias correctamente clasificadas. Y por último, en el caso de *Cross-Validation* y *Percentage Split* se observa que el rendimiento en las instancias correctamente clasificadas del algoritmo bajó ampliamente a un 56,02% y 51,32% respectivamente.

¹ Ver Tabla 5-10.

En el siguiente apartado se realiza la comparación de los resultados que fueron obtenidos en las bases de datos con imágenes y sin imágenes.

6.2 Comparación de los resultados

Con el objetivo de evitar confusiones en el análisis de la comparación de los resultados obtenidos en las bases de datos, en este apartado se asignan los nombres descritos a continuación:

- **“Base de datos 1”**: corresponde al conjunto de datos que incluye las fotografías de las fachadas.
- **“Base de datos 2”**: hace referencia a la base de datos que no contiene las imágenes.

Así, la comparación se realiza únicamente con el algoritmo BayesNet por alcanzar el mayor número de instancias correctamente clasificadas en todos los modos de prueba (*Use Training Set, Supplied Test Set, Cross- Validation y Percentage Split*). En la Tabla 6-2 se presenta la comparación de los resultados obtenidos por este algoritmo en las bases de datos 1 y 2.

ALGORITMO	MODO DE PRUEBA	BASE DE DATOS 1= CON IMÁGENES				BASE DE DATOS 2= SIN IMÁGENES			
		INSTANCIAS CORRECTAMENTE CLASIFICADAS		INSTANCIAS INCORRECTAMENTE CLASIFICADAS		INSTANCIAS CORRECTAMENTE CLASIFICADAS		INSTANCIAS INCORRECTAMENTE CLASIFICADAS	
		INSTANCIAS	CANTIDAD DE DATOS	INSTANCIAS	CANTIDAD DE DATOS	INSTANCIAS	CANTIDAD DE DATOS	INSTANCIAS	CANTIDAD DE DATOS
BayesNet	Use Training Set	100,00%	332	0,00%	0	92,46%	307	7,53%	25
	Supplied Test Set	90,00%	36	10,00%	4	65,00%	26	35,00%	14
	Cross-Validation (Folds 10)	98,19%	326	1,80%	6	56,02%	186	43,97%	146
	Percentage Split 66%	97,34%	110	2,65%	3	51,32%	58	48,67%	55

Tabla 6-2. Comparación de los resultados obtenidos.

Fuente: Elaboración propia.

A continuación se presenta de forma gráfica los resultados obtenidos en cada modo de prueba, por el algoritmo BayesNet para las bases de datos 1 y 2.

- **Modo de prueba Use Training Set**

En la Figura 6-1 se presentan los resultados obtenidos en cada una de las bases de datos, mediante el modo evaluativo *Use Training Set*.

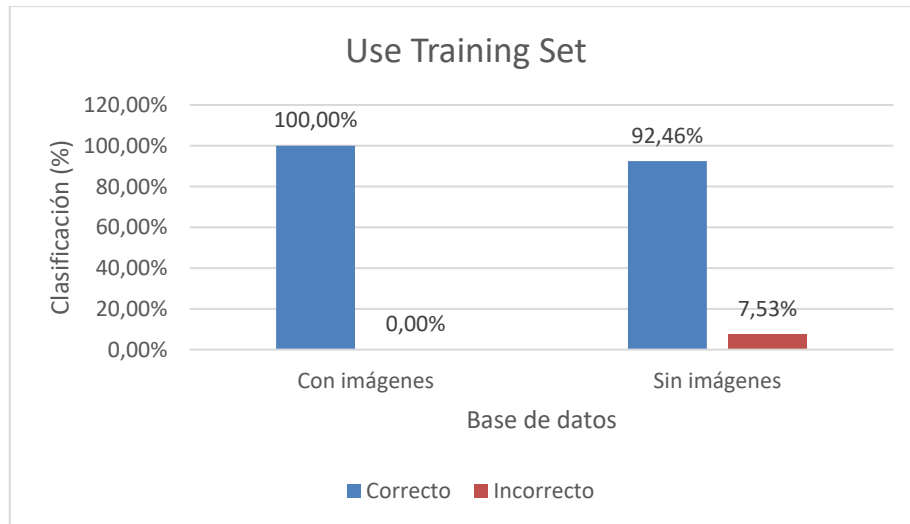


Figura 6-1. Comparación de resultados obtenidos para las bases de datos 1 y 2, modo *Use Training Set*.
Fuente: Elaboración propia.

Como se observa, la base de datos 1 alcanza un porcentaje de 100% en las instancias correctamente clasificadas. En el caso de la base de datos 2 el desempeño se reduce a un 92,46%.

- **Modo de prueba Supplied Test Set**

En la Figura 6-2 se muestran los resultados para el modo evaluativo *Supplied Test Set*.

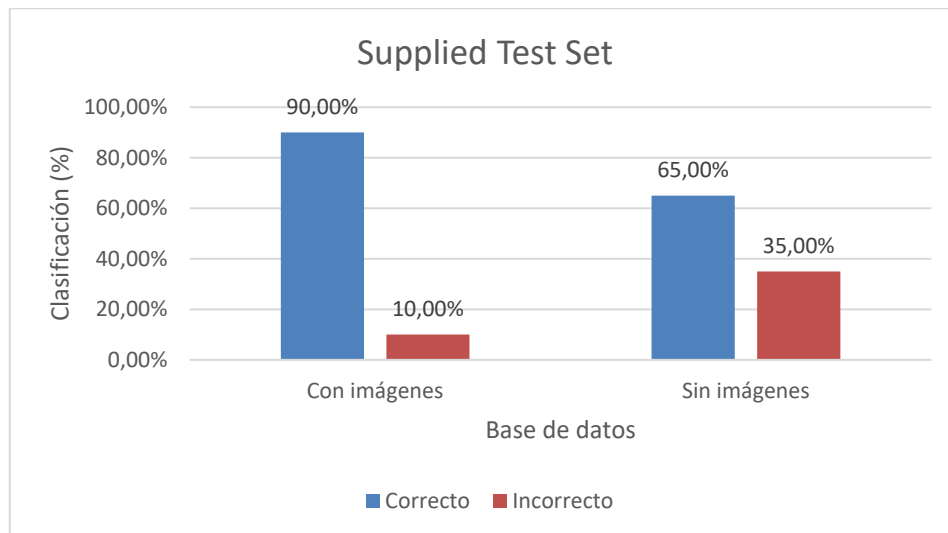


Figura 6-2. Comparación de resultados obtenidos para las bases de datos 1 y 2, modo *Supplied Test Set*.
Fuente: Elaboración propia.

En el modo de evaluación *Supplied Test Set*, el mejor desempeño lo tiene la base de datos 1, presentando un 90% de aciertos. Por otro lado, la base de datos que no contiene las imágenes clasificó de forma correcta sólo el 65% de las instancias.

- **Modo de prueba Cross- Validation (Folds 10)**

En la Figura 6-3 se presentan los resultados para el modo evaluativo *Cross- Validation*.

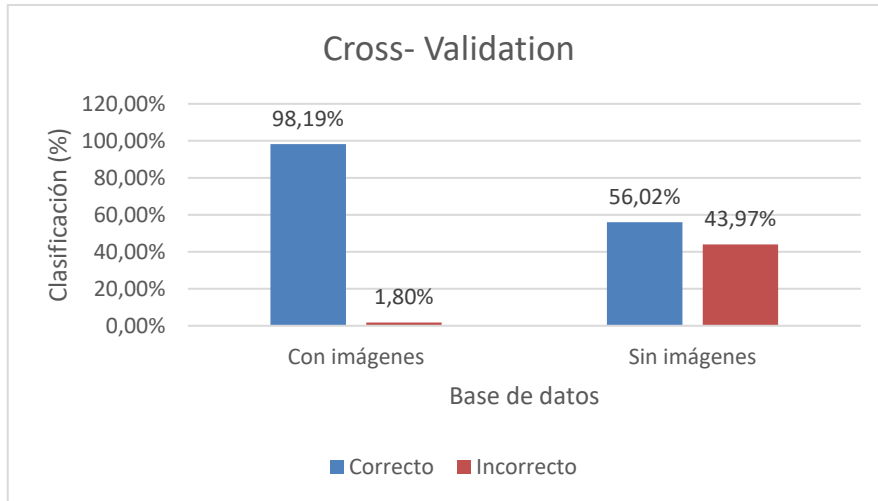


Figura 6-3. Comparación de resultados obtenidos para las bases de datos 1 y 2, modo *Cross- Validation*.
Fuente: Elaboración propia.

De acuerdo con los resultados obtenidos en el modo de prueba *Cross- Validation*, las bases de datos 1 y 2 presentan un porcentaje de instancias correctamente clasificadas de 98,19% y de 56,02% respectivamente.

Modo de prueba Percentage Split 66%

En la Figura 6-4 se presentan los resultados alcanzados para las bases de datos 1 y 2, mediante el modo de prueba *Percentage Split*.

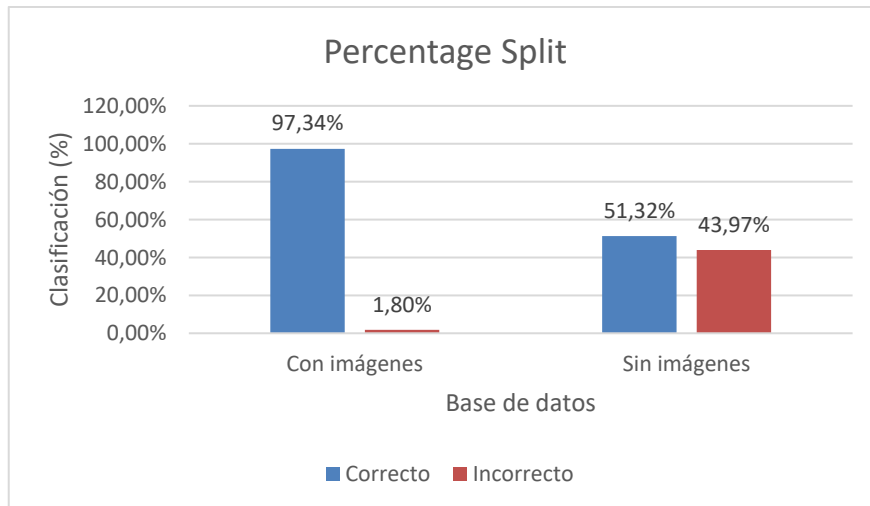


Figura 6-4. Comparación de resultados obtenidos para las bases de datos 1 y 2, modo *Percentage Split*.
Fuente: Elaboración propia.

Como se observa, la base de datos 1 arroja un 97,34% de aciertos en la clasificación. Así mismo, la base de datos 2 entrega un 51,32% de instancias correctamente clasificadas.

Según los resultados obtenidos en los porcentajes de instancias correctamente clasificadas, la discrepancia más alta se da en el modo de prueba *Percentage Split* presentando un 97,34% y un 51,32% en las bases de datos 1 y 2 respectivamente. En el mismo sentido, para el caso de *Cross-Validation* se tiene un 98,19% en la base de datos 1, y 56,02% en la base de datos 2, y para el modo de prueba *Supplied Test Set* se cuenta con un 90% y un 65% en las bases de datos 1 y 2. Finalmente, el mejor aprendizaje para las dos bases de datos se da en el modo de prueba *Use Training Set*, en el cual se presenta 100% de instancias correctamente clasificadas en la base de datos 1 y 92,46% en la base de datos 2.

Al realizar la comparación de los resultados obtenidos en los dos casos (bases de datos 1 y 2), el algoritmo *BayesNet* aplicado a la base de datos que contiene las fotografías, alcanza los mejores porcentajes de clasificación en todos los modos de prueba. En el primer caso, se presentan resultados con una clasificación mayor al 90%, y en el caso de la base de datos 2 alcanza un 60% aproximadamente en 3 modos de prueba. De esta manera, es posible establecer que para los dos casos se encuentra una amplia diferencia de instancias correctamente clasificadas.

7 CONCLUSIONES Y PERSPECTIVAS

7.1 Conclusiones

Para llevar a cabo este experimento fue necesario realizar la depuración de la información contenida en la base de datos inicial, lo anterior para determinar las variables que garantizaran un buen aprendizaje del algoritmo. También para seleccionar una cantidad limitada de predios que permitieran efectuar el proceso de adquisición de las fotografías, y posterior asociación con cada inmueble en la base de datos.

El entrenamiento de los algoritmos se realizó empleando una base de datos híbrida con información alfanumérica y de imágenes. Por consiguiente, el aprendizaje se dio mediante el reconocimiento de datos no solamente de carácter fotográfico, sino datos adicionales que describen las características de los inmuebles. Dicho lo anterior al contar con datos adicionales a las fotografías, la capacidad de aprendizaje del algoritmo mejora permitiendo alcanzar mayores niveles de precisión en la clasificación realizada.

La exploración de los acabados de las fachadas en cada una de las fotografías permitió adicionar una variable denominada *acab_fachada*, que contribuyó al aprendizaje de los algoritmos aplicados en la base de datos con imágenes, especialmente en el algoritmo BayesNet. Así mismo, el preprocesamiento del conjunto de datos se consolidó como un procedimiento importante en este experimento. Lo anterior debido a que el tratamiento previo realizado a las variables mediante los filtros de Weka permitió estandarizar los datos, reducir los errores y procesar las variables con datos ocultos como lo son las fotografías.

El aplicativo Mapas Bogotá con la extensión de Street View resultó ser una herramienta útil para el trabajo realizado en la toma de fotografías de las fachadas de cada uno de los predios, facilitando su búsqueda con el identificador CHIP y verificando de la nomenclatura asociada a cada construcción. Además, mediante este procedimiento se optimizaron los tiempos y costos que hubiera requerido una visita en campo para obtener este insumo.

El software Weka permitió realizar desde el tratamiento de los datos hasta el proceso de entrenamiento y validación en cada experimento. Así mismo, facilitó la aplicación de múltiples algoritmos disponibles en los grupos de Bayes, Lazy, Funciones y Árboles de Decisión. Además, consolida los errores generados en el algoritmo mediante instrumentos como la matriz de confusión, estadístico Kappa y los errores estadísticos a nivel general e individual.

Los algoritmos seleccionados: BayesNet, IBK y RandomForest, en general consiguieron un porcentaje de más del 90% de instancias correctamente clasificadas en todos los modos de prueba. Aunque se obtuvieron resultados satisfactorios en la evaluación de los algoritmos seleccionados, se escogió el BayesNet por tener el mayor acierto en la clasificación de las instancias. Esto permitió hacer énfasis en resultados como la matriz de confusión y errores estadísticos entregados en cada método de evaluación.

De acuerdo con la evaluación realizada mediante los modos de prueba en el algoritmo BayesNet, se tuvieron errores bajos, los cuales fueron determinados por criterios estadísticos como el *error medio absoluto*, el *error cuadrático medio*, el *error absoluto relativo* y el *error de raíz cuadrada relativo*. Así mismo, el *estadístico Kappa* de cada método de validación permitió establecer una alta correlación entre la predicción esperada y la resultante. Por último, los *errores de predicción* mostraron valores que indicaron que el grado de pureza de cada clasificación fue alta.

A partir de la comparación de los resultados alcanzados mediante los modos de prueba en el algoritmo *BayesNet*, se encontró que el conjunto de datos que incluye las fotografías presenta un porcentaje de aciertos superior al obtenido por la base de datos que no contiene imágenes. La diferencia de resultados es amplia considerando que la base de datos que incluye las fotografías presenta aciertos superiores al 90%, y el conjunto de datos numérico muestra porcentajes del 60% aproximadamente en 3 modos de prueba. Según lo anterior, la variable *imagen* aporta información importante en el aprendizaje del algoritmo.

Finalmente, se comprobó que la metodología propuesta resultó satisfactoria ya que permitió desarrollar de forma secuencial, clara y organizada los objetivos propuestos. Es válida, pues se alcanzaron los resultados esperados para la realización de la predicción del puntaje de construcción de los predios mediante datos numéricos y las fotografías de sus fachadas en el área de estudio definida.

7.2 Perspectivas

Es necesario que los entes territoriales desarrollen estrategias que permitan mejorar la metodología de adquisición y actualización de la información catastral. Lo anterior aseguraría el incremento de los recaudos obtenidos por el IPU, y un mayor crecimiento del PIB en un país como Colombia que cuenta con una tasa de urbanización de más del 80%.

Así mismo, se requiere la actualización periódica de las bases de datos catastrales asociadas a la calificación de las construcciones. Este insumo garantiza que el cálculo del puntaje de las edificaciones y a su vez del IPU, esté acorde con las características físicas actuales de los inmuebles.

Este proyecto corresponde a la implementación de la IA en la liquidación del IPU, uno de los procesos más importantes en el área del catastro. Se propone continuar y mejorar la exploración del recurso tecnológico enfocado en la IA, para ejecutar otros procesos catastrales operativos y aplicarlos en el territorio nacional.

Debido a la amplia cantidad de tiempo que se requiere para la asociación de cada fotografía con su predio en la base de datos, se considera importante agregar un complemento a las funcionalidades de Weka que permita optimizar este procedimiento. Por otro lado, una estrategia más ambiciosa podría ser la creación de un sitio web o el desarrollo de una aplicación para los dispositivos móviles. De esta manera se adelantaría todo el proceso de calificación de las construcciones de forma automática al insertar la fotografía de la fachada del predio.

Weka es un software fácil y rápido de usar, ofrece la posibilidad de aplicar múltiples algoritmos en tiempos de espera muy cortos con bases de datos muy robustas. Según lo anterior, este proyecto puede implementarse en áreas territoriales más grandes como una ciudad o un municipio. Además, sirve como base para aplicar nuevas variables que se usan en el método convencional (baño, cocina, pisos, entre otros), y que pueden estar contenidas en bases de datos actualizadas y disponibles.

A partir del análisis y comparación de los resultados obtenidos se determinó que las fotografías de las fachadas aportaron información importante para el aprendizaje de los algoritmos, especialmente el *BayesNet*. Se propone incluir variables de tipo imagen en futuras investigaciones y experimentos asociados a procesos catastrales desarrollados con IA.

BIBLIOGRAFÍA

- [1] B. KAHRAMAN, “Qué efectos a nivel de principios y derechos constitucionales producen las disposiciones contempladas en el artículo 6 de la Ley 1742 de 2014,” vol. 147, pp. 11–40, 2016.
- [2] Alcaldía de Medellín, “Glosario Certificado Catastral,” p. 5, 2017.
- [3] J. Mora-Esperanza, “La inteligencia artificial aplicada a la valoración de inmuebles: un ejemplo para valorar Madrid,” *CT Catastro*, pp. 51–68, 2004.
- [4] Unidad Administrativa Especial de Catrasto, “Catastro Bogotá,” 2020. [Online]. Available: catastrobogota.gov.co.
- [5] S. Quete and D. Cuéllar, “XIII Seminario sobre sobre ‘ Catastro como instrumento p para la tributacion inmobiliaria,”” pp. 1–30, 2010.
- [6] J. F. Ramos, A. B. Iregui, and L. B. Melo, “Ana María Iregui B., Ligia Melo B,” *Rev. Econ. del Rosario*, pp. 1–35, 2005.
- [7] J. J. Romero, C. Dafonte, Á. Gòmez, and F. J. Penousal, *Inteligencia Artificial Y Computación Avanzada*. 2007.
- [8] T. D. E. Contenido, “Elaboración del estudio de zonas homogéneas físicas y geoeconómicas y determinación del valor unitario por,” 2010.
- [9] Catastro Bogotá, “¿Cómo se establece el Avalúo Catastral de cada predio?” [Online]. Available: <https://www.catastrobogota.gov.co/pregunta/como-se-establece-el-avaluo-catastral-de-cada-predio>.
- [10] Secretaría Distrital de Hacienda, “Impuesto predial unificado,” 2020. [Online]. Available: <https://www.shd.gov.co/shd/node/28579>.
- [11] H. E. y D. Erba, “Catastro, Valoración Inmobiliaria Y Tributación Municipal,” 2020.
- [12] H. Banda, “Inteligencia Artificial: Principios y Aplicaciones,” no. April, 2014.
- [13] J. M. Vargas, C. T. Pinzón, and C. R. Patiño, “Técnicas de inteligencia artificial para la solución de laberintos de estructura desconocida.,” *Técnicas Intel. Artif. Para La Solución Laberin. Estruct. Desconocida.*, vol. 2, no. 39, pp. 135–140, 2008.
- [14] J. M. C. García-Retuerta, David, Román Andrés Bondía, Javier Prieto Tejedor, “Inteligencia artificial para la asignación automática de categorías constructivas.”
- [15] A. Emilcen and C. Fajardo, “Propuesta Metodológica para Calcular el Avalúo Catastral de un Predio Utilizando Redes Neuronales Artificiales,” 2014.
- [16] J. J. Rodrigo and J. J. Rosales, “Extracción de conocimiento mediante la aplicación de inteligencia artificial a la información espacial Spatial artificial intelligence knowledge extraction,” vol. 26, pp. 40–47, 2018.

- [17] J. M. Núñez Tabales, F. J. Rey Carmona, and J. M. Caridad y Ocerin, "Redes neuronales (RN) aplicadas a la valoración de locales comerciales," *Inf. la Constr.*, vol. 69, no. 545, 2017.
- [18] M. Michoacan, "Redes neuronales artificiales , una aplicación al mercado inmobiliario habitacional de segmento medio de la ciudad de Morelia Michoacán , México," vol. XIV, no. 1, pp. 49–68, 2019.
- [19] C. Of and S. A. N. Francisco, "Valuación masiva de la tierra urbana mediante inteligencia artificial. El caso de la ciudad de San Francisco, Córdoba, Argentina.," pp. 90–112, 2019.
- [20] W. R. Osorio, G. Alexander, R. Patiño, C. Director, and M. A. Álvarez, "Predicción del avalúo catastral de los predios urbanos en la ciudad de Pereira mediante aprendizaje de máquina," 2017.
- [21] L. M. C. SALAZAR and Y. P. BARRIGA, "Cultura de la polación colombiana con la constribución del impuesto predial," pp. 1036–1037, 2012.
- [22] F. Reyes Bueno, D. Miranda Barros, and R. Crecente Maseda, "Situacion de la valoracion catastral rural," *CT. Catastro. Rev. la Dir. Gen. del Catastro*, no. 75, pp. 41–54, 2012.
- [23] DANE, "Metodología Índice de Valoración Predial," *Dane*, 2009.
- [24] IGAC, "Resolución 070 de 2011," vol. 11, no. 2, pp. 10–14, 2011.
- [25] I. G. A. C. (IGAC), "¿En qué consiste el proceso de actualización catastral?," 2018. [Online]. Available: <https://www.igac.gov.co/es/contenido/en-que-consiste-el-proceso-de-actualizacion-catastral>.
- [26] A. M. Iregui-Bohórquez, L. A. Melo-Becerra, and J. E. Ramos-Forero, "El impuesto predial en Colombia : evolución reciente, comportamiento de las tarifas y potencial de recuado," *Ensayos sobre Política Económica*, no. 36, pp. 186–248, 2004.
- [27] Catastro Bogotá, "Reconocimiento predial." [Online]. Available: <https://www.catastrobogota.gov.co/glosario/reconocimiento-predial>.
- [28] D. E. S. D. E. Informaci, "Manual de calificación de las construcciones," pp. 1–134.
- [29] A. M. de Bogotá, "Ley 44 de 1990," no. January, p. 43, 2001.
- [30] M. J. V. Ledo, A. M. González, and D. V. Santiago, "Inteligencia artificial en la docencia médica," *Rev. Cuba. Educ. Medica Super.*, vol. 33, no. 3, 2019.
- [31] J. Méndez Palma and R. Morales Marín, *Inteligencia Artificial*, vol. 2, no. 6. 2007.
- [32] H. Rondón Troncoso, L. Ladino Gaspar, and P. Orduz Camacho, "Acerca de la enseñanza del teorema de Bayes," *Rev. Educ. y Desarro. Soc.*, vol. 9, no. 1, pp. 144–159, 2015.
- [33] H. Díaz-Barrios, Y. Alemán-Rivas, L. Cabrera-Hernández, A. Morales-Hernández, M. Chávez-Cárdenas, and G. Casas-Cardoso, "Algoritmos de aprendizaje automático para clasificación de Splice Sites en secuencias genómicas," *Rev. Cuba. Ciencias Informáticas*, vol. 9, no. 4, pp. 155–170, 2015.
- [34] C. L. Corso, M. F. Pereyra, G. Martínez, and F. Gibellini, "Sistema de Soporte de Decisión para

- la Gestión de fallos en equipos industriales , basado en métodos de ensamble .,” pp. 186–197.
- [35] F. Gomariz-castillo, “Modificación del algoritmo Random Forest para su empleo en clasificación de imágenes de teledetección,” pp. 359–368, 2016.
- [36] R. De Telecomunicación, C. G. Cambronero, and I. G. Moreno, “Algoritmos de aprendizaje: knn & kmeans.”
- [37] M. M. GARCÍA, “Modelo De Un Sistema De Razonamiento Basado En Casos Para El Análisis En La Gestión De Riesgos,” *Ser. Científica La Univ. Las Ciencias Informáticas, No 11*, vol. 4, 2011.
- [38] M. L. Group, “WEKA The workbench for machine learning.” [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/>.
- [39] D. G. Morate, *Manual de Weka*. 2002.
- [40] M. Hall, E. Frank, and I. H. Witten, “Practical Data Mining Tutorial 1 : Introduction to the WEKA Explorer,” 2012.
- [41] F. europeo de desenvolvimiento rexional; xunta de G. consellería de Sanidade, “Concordancia y consistencia,” *Epidat*, 2014.
- [42] P. E. Paucar Uyaguari and L. A. Yubi Quito, “Pronostico de perfiles de consumo de electricidad a través de modelos estadísticos y desarrollo de una metodología comparativa,” 2020.
- [43] A. del P. A. Robles, “Análisis de efectividad al implementar la técnica de árboles de decisión del enfoque de aprendizaje de máquina para la determinación de avalúos masivos para las UPZ 97 Calandaima, 65 Arborizadora y 73 Garcés Navas,” 2017.
- [44] I. Corso and C. Lorena, “Aplicación de algoritmos de clasificación supervisada usando Weka.,” p. 11, 2007.
- [45] A. De Datos, C. Técnicas, D. E. A. D. E. Datos, and E. N. Weka, “Técnicas de análisis de datos en WEKA.”
- [46] IDECA, “Mapas Bogotá.” [Online]. Available: <https://mapas.bogota.gov.co/#>.
- [47] U. de Waikato, “WEKA The workbench for machine learning.”
- [48] C. R. Manzanares, “Análisis de los métodos de predicción aplicados a los desvíos en el sistema eléctrico Ibérico.”

ANEXOS

• **ANEXO 1: RESULTADOS DEL ALGORITMO BAYESNET CON EL MODO DE PRUEBA USE TRAINING SET**

ACTUAL	PREDICCIÓN	ERROR
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	1
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,997
Tipo 3	Tipo 3	0,998
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	1
Tipo 1	Tipo 1	0,996
Tipo 1	Tipo 1	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	1
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 1	Tipo 1	0,991
Tipo 3	Tipo 3	1
Tipo 3	Tipo 3	0,997
Tipo 3	Tipo 3	0,996
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	1
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	0,996
Tipo 3	Tipo 3	0,998
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	1

Tipo 3	Tipo 3	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	1
Tipo 2	Tipo 2	0,995
Tipo 3	Tipo 3	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,996
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 1	Tipo 1	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	1
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,996
Tipo 3	Tipo 3	0,995
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,995
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,998
Tipo 3	Tipo 3	0,989
Tipo 3	Tipo 3	0,997
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	1
Tipo 3	Tipo 3	0,986
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,989
Tipo 2	Tipo 2	1

Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,993
Tipo 2	Tipo 2	1
Tipo 4	Tipo 4	0,915
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,997
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	0,997
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,997
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	1
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,997
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,998
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,992
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	0,989
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1

Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,992
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,906
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	0,999
Tipo 4	Tipo 4	0,993
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 4	Tipo 4	0,967
Tipo 1	Tipo 1	0,999
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 4	Tipo 4	0,98
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 4	Tipo 4	0,984
Tipo 4	Tipo 4	0,906
Tipo 1	Tipo 1	1

Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 1	Tipo 1	0,996
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,998
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,998
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	1
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	0,995
Tipo 4	Tipo 4	0,991
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	1
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	1
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	1

Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,995
Tipo 3	Tipo 3	0,996
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,995
Tipo 3	Tipo 3	0,995
Tipo 2	Tipo 2	0,999
Tipo 4	Tipo 4	0,997
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	0,991
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	0,998
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,996
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	1

- ANEXO 2: RESULTADOS EN CADA PREDICCIÓN DEL ALGORITMO BAYESNET CON EL MODO DE PRUEBA SUPPLIED TEST SET

ACTUAL	PREDICCIÓN	ERROR
Tipo 1	Tipo 1	0,451
Tipo 1	Tipo 3	0,675
Tipo 1	Tipo 4	0,919
Tipo 1	Tipo 4	0,985
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	0,999
Tipo 1	Tipo 1	0,966
Tipo 1	Tipo 1	0,994
Tipo 1	Tipo 1	0,99
Tipo 1	Tipo 2	0,425
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,994
Tipo 2	Tipo 2	0,999

Tipo 2	Tipo 2	0,661
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	0,998
Tipo 2	Tipo 2	0,992
Tipo 2	Tipo 2	0,993
Tipo 3	Tipo 3	0,983
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,571
Tipo 3	Tipo 3	0,997
Tipo 3	Tipo 3	0,919
Tipo 3	Tipo 3	0,977
Tipo 3	Tipo 3	0,999

Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,904
Tipo 3	Tipo 3	0,998
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	1

- ANEXO 3: RESULTADOS DEL ALGORITMO BAYESNET CON EL MODO DE PRUEBA PERCENTAGE SPLIT

ACTUAL	PREDICCIÓN	ERROR
Tipo 2	Tipo 2	0,993
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,832
Tipo 4	Tipo 4	1
Tipo 2	Tipo 2	0,987
Tipo 4	Tipo 4	0,999
Tipo 2	Tipo 2	0,986
Tipo 4	Tipo 4	0,992
Tipo 3	Tipo 3	0,937
Tipo 3	Tipo 3	0,994
Tipo 2	Tipo 2	0,999
Tipo 4	Tipo 4	0,999
Tipo 2	Tipo 2	0,985
Tipo 2	Tipo 2	0,987
Tipo 3	Tipo 3	0,998
Tipo 2	Tipo 2	0,992
Tipo 3	Tipo 3	0,989
Tipo 4	Tipo 4	1
Tipo 3	Tipo 3	0,971
Tipo 3	Tipo 3	0,989
Tipo 2	Tipo 2	0,99
Tipo 3	Tipo 3	0,989
Tipo 2	Tipo 2	0,938
Tipo 3	Tipo 3	0,793
Tipo 4	Tipo 4	0,999
Tipo 3	Tipo 3	0,698
Tipo 3	Tipo 3	0,989
Tipo 1	Tipo 1	0,712
Tipo 2	Tipo 2	0,988
Tipo 3	Tipo 3	0,863
Tipo 3	Tipo 3	0,962
Tipo 2	Tipo 2	0,944
Tipo 3	Tipo 3	0,989
Tipo 1	Tipo 1	0,998
Tipo 3	Tipo 3	0,973
Tipo 2	Tipo 2	0,991

Tipo 3	Tipo 3	0,922
Tipo 2	Tipo 4	0,367
Tipo 1	Tipo 1	0,943
Tipo 3	Tipo 3	0,973
Tipo 3	Tipo 3	0,863
Tipo 4	Tipo 4	0,993
Tipo 4	Tipo 4	0,995
Tipo 4	Tipo 4	0,996
Tipo 3	Tipo 3	1
Tipo 3	Tipo 3	0,989
Tipo 3	Tipo 3	0,989
Tipo 3	Tipo 3	0,981
Tipo 4	Tipo 4	0,999
Tipo 2	Tipo 2	0,991
Tipo 3	Tipo 3	0,92
Tipo 3	Tipo 3	0,982
Tipo 2	Tipo 2	0,991
Tipo 2	Tipo 2	0,984
Tipo 2	Tipo 2	0,964
Tipo 2	Tipo 2	0,978
Tipo 3	Tipo 3	0,998
Tipo 1	Tipo 1	0,999
Tipo 3	Tipo 3	0,989
Tipo 2	Tipo 2	0,925
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,979
Tipo 3	Tipo 3	0,994
Tipo 1	Tipo 1	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,986
Tipo 2	Tipo 2	0,964
Tipo 4	Tipo 4	0,997
Tipo 4	Tipo 4	0,829
Tipo 3	Tipo 3	0,993
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,99
Tipo 2	Tipo 2	0,997
Tipo 4	Tipo 4	0,996

Tipo 3	Tipo 3	0,995
Tipo 2	Tipo 2	0,451
Tipo 3	Tipo 3	0,994
Tipo 2	Tipo 2	0,996
Tipo 2	Tipo 2	0,993
Tipo 2	Tipo 2	0,989
Tipo 3	Tipo 3	0,997
Tipo 4	Tipo 4	0,628
Tipo 3	Tipo 3	0,977
Tipo 2	Tipo 2	0,961
Tipo 4	Tipo 4	0,993
Tipo 4	Tipo 4	0,999
Tipo 2	Tipo 2	0,964
Tipo 3	Tipo 3	0,989
Tipo 4	Tipo 4	1
Tipo 1	Tipo 4	0,576
Tipo 4	Tipo 4	0,985
Tipo 2	Tipo 2	0,945
Tipo 2	Tipo 2	0,873
Tipo 2	Tipo 2	0,979
Tipo 3	Tipo 3	0,991
Tipo 4	Tipo 4	0,829
Tipo 3	Tipo 3	1
Tipo 4	Tipo 4	1
Tipo 3	Tipo 3	0,937
Tipo 4	Tipo 4	0,869
Tipo 3	Tipo 3	0,966
Tipo 1	Tipo 1	0,985
Tipo 3	Tipo 3	0,989
Tipo 2	Tipo 2	0,994
Tipo 1	Tipo 1	0,971
Tipo 3	Tipo 3	0,989
Tipo 1	Tipo 2	0,384
Tipo 2	Tipo 2	0,996
Tipo 2	Tipo 2	0,994
Tipo 4	Tipo 4	0,998
Tipo 1	Tipo 1	0,998
Tipo 3	Tipo 3	0,823

- ANEXO 4: RESULTADOS DEL ALGORITMO BAYESNET CON EL MODO DE PRUEBA CROSS- VALIDATION

ACTUAL	PREDICCIÓN	ERROR
Tipo 3	Tipo 3	0,981
Tipo 3	Tipo 3	0,993
Tipo 3	Tipo 3	0,98
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,967
Tipo 3	Tipo 3	0,949
Tipo 3	Tipo 3	0,988
Tipo 3	Tipo 3	0,978
Tipo 3	Tipo 3	0,866
Tipo 3	Tipo 3	0,995
Tipo 3	Tipo 3	0,974
Tipo 3	Tipo 3	0,996
Tipo 3	Tipo 3	0,924
Tipo 2	Tipo 2	0,986
Tipo 2	Tipo 2	0,992
Tipo 2	Tipo 2	0,995
Tipo 2	Tipo 2	0,996
Tipo 2	Tipo 2	0,835
Tipo 2	Tipo 4	0,411
Tipo 2	Tipo 2	0,992
Tipo 2	Tipo 2	0,951
Tipo 2	Tipo 2	0,999
Tipo 1	Tipo 1	0,974
Tipo 1	Tipo 1	0,975
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	0,926
Tipo 4	Tipo 4	0,957
Tipo 4	Tipo 4	0,83
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,988
Tipo 4	Tipo 4	0,995
Tipo 4	Tipo 4	0,995
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,991

Tipo 3	Tipo 3	0,993
Tipo 3	Tipo 3	0,982
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,907
Tipo 3	Tipo 3	0,997
Tipo 3	Tipo 3	0,991
Tipo 3	Tipo 3	0,935
Tipo 3	Tipo 3	0,986
Tipo 3	Tipo 3	0,993
Tipo 2	Tipo 2	0,77
Tipo 2	Tipo 2	0,99
Tipo 2	Tipo 2	0,995
Tipo 2	Tipo 2	0,99
Tipo 2	Tipo 2	0,987
Tipo 2	Tipo 2	0,998
Tipo 2	Tipo 2	0,952
Tipo 2	Tipo 4	0,768
Tipo 2	Tipo 2	0,997
Tipo 1	Tipo 1	0,999
Tipo 1	Tipo 1	0,992
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,985
Tipo 4	Tipo 4	0,996
Tipo 4	Tipo 4	0,959
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	0,699
Tipo 4	Tipo 4	0,695
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 2	0,673
Tipo 3	Tipo 3	0,988
Tipo 3	Tipo 3	0,602
Tipo 3	Tipo 3	0,993
Tipo 3	Tipo 3	0,865
Tipo 3	Tipo 3	0,989

Tipo 3	Tipo 3	0,988
Tipo 3	Tipo 3	0,98
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,797
Tipo 3	Tipo 3	0,969
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,99
Tipo 3	Tipo 3	0,996
Tipo 2	Tipo 2	0,992
Tipo 2	Tipo 2	0,989
Tipo 2	Tipo 2	0,991
Tipo 2	Tipo 2	0,987
Tipo 2	Tipo 2	0,991
Tipo 2	Tipo 2	0,981
Tipo 2	Tipo 2	0,983
Tipo 2	Tipo 2	0,986
Tipo 2	Tipo 2	0,996
Tipo 1	Tipo 1	0,995
Tipo 1	Tipo 1	0,999
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,994
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,993
Tipo 4	Tipo 4	0,698
Tipo 4	Tipo 4	0,968
Tipo 4	Tipo 4	0,99
Tipo 4	Tipo 4	0,968
Tipo 3	Tipo 3	0,847
Tipo 3	Tipo 3	0,908
Tipo 3	Tipo 3	0,975
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,953
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,995
Tipo 3	Tipo 3	0,995
Tipo 3	Tipo 3	0,998

Tipo 3	Tipo 3	0,822
Tipo 3	Tipo 3	0,995
Tipo 3	Tipo 3	0,996
Tipo 3	Tipo 3	0,873
Tipo 2	Tipo 2	0,986
Tipo 2	Tipo 2	0,994
Tipo 2	Tipo 2	0,968
Tipo 2	Tipo 2	0,984
Tipo 2	Tipo 2	0,989
Tipo 2	Tipo 2	0,998
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	0,858
Tipo 2	Tipo 2	0,717
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	0,688
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,995
Tipo 4	Tipo 4	0,995
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,997
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 3	Tipo 3	0,937
Tipo 3	Tipo 3	0,996
Tipo 3	Tipo 3	0,965
Tipo 3	Tipo 3	0,948
Tipo 3	Tipo 3	0,967
Tipo 3	Tipo 3	0,914
Tipo 3	Tipo 3	0,979
Tipo 3	Tipo 3	0,946
Tipo 3	Tipo 3	0,949
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,998
Tipo 2	Tipo 2	0,98
Tipo 2	Tipo 2	0,999
Tipo 2	Tipo 2	0,995
Tipo 2	Tipo 2	0,986
Tipo 2	Tipo 2	0,992

Tipo 2	Tipo 2	0,986
Tipo 2	Tipo 2	0,991
Tipo 2	Tipo 3	0,441
Tipo 2	Tipo 2	0,991
Tipo 1	Tipo 1	0,999
Tipo 1	Tipo 1	0,999
Tipo 1	Tipo 1	0,999
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	0,992
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,999
Tipo 3	Tipo 3	0,978
Tipo 3	Tipo 3	0,87
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,993
Tipo 3	Tipo 3	0,62
Tipo 3	Tipo 3	0,946
Tipo 3	Tipo 3	0,93
Tipo 3	Tipo 3	0,887
Tipo 3	Tipo 3	0,99
Tipo 3	Tipo 3	0,968
Tipo 2	Tipo 2	0,986
Tipo 2	Tipo 2	0,995
Tipo 2	Tipo 2	0,996
Tipo 2	Tipo 2	0,988
Tipo 2	Tipo 2	1
Tipo 2	Tipo 2	0,969
Tipo 2	Tipo 2	0,998
Tipo 2	Tipo 2	0,992
Tipo 2	Tipo 2	0,987
Tipo 1	Tipo 1	1
Tipo 1	Tipo 1	1
Tipo 4	Tipo 4	0,995
Tipo 4	Tipo 4	0,952

Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,991
Tipo 4	Tipo 4	0,977
Tipo 4	Tipo 4	0,971
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,935
Tipo 4	Tipo 4	1
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,995
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,94
Tipo 3	Tipo 3	0,997
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,987
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,915
Tipo 3	Tipo 3	0,925
Tipo 3	Tipo 3	0,994
Tipo 2	Tipo 2	0,988
Tipo 2	Tipo 2	0,975
Tipo 2	Tipo 2	0,973
Tipo 2	Tipo 2	0,994
Tipo 2	Tipo 2	0,993
Tipo 2	Tipo 2	0,976
Tipo 2	Tipo 2	0,992
Tipo 2	Tipo 2	0,979
Tipo 2	Tipo 2	0,991
Tipo 1	Tipo 1	0,993
Tipo 1	Tipo 1	0,952
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,992
Tipo 4	Tipo 4	0,985
Tipo 4	Tipo 4	0,946
Tipo 4	Tipo 4	0,999

Tipo 3	Tipo 3	0,993
Tipo 3	Tipo 3	0,993
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,993
Tipo 3	Tipo 3	0,984
Tipo 3	Tipo 3	0,927
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,982
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,984
Tipo 3	Tipo 3	0,979
Tipo 3	Tipo 3	0,992
Tipo 2	Tipo 2	0,987
Tipo 2	Tipo 2	0,99
Tipo 2	Tipo 2	0,976
Tipo 2	Tipo 2	0,995
Tipo 2	Tipo 2	0,993
Tipo 2	Tipo 2	0,396
Tipo 2	Tipo 2	0,984
Tipo 2	Tipo 2	0,992
Tipo 2	Tipo 2	0,985
Tipo 1	Tipo 1	0,908
Tipo 1	Tipo 1	0,616
Tipo 4	Tipo 4	1
Tipo 4	Tipo 1	0,593
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,812

Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,961
Tipo 3	Tipo 3	0,995
Tipo 3	Tipo 3	0,903
Tipo 3	Tipo 3	0,911
Tipo 3	Tipo 3	0,992
Tipo 3	Tipo 3	0,969
Tipo 3	Tipo 3	0,997
Tipo 3	Tipo 3	0,994
Tipo 3	Tipo 3	0,974
Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,999
Tipo 2	Tipo 2	0,981
Tipo 2	Tipo 2	0,976
Tipo 2	Tipo 2	0,981
Tipo 2	Tipo 2	0,962
Tipo 2	Tipo 2	0,372
Tipo 2	Tipo 2	0,995
Tipo 2	Tipo 2	0,988
Tipo 2	Tipo 2	0,968
Tipo 2	Tipo 2	0,993
Tipo 1	Tipo 1	0,759
Tipo 1	Tipo 1	0,966
Tipo 4	Tipo 4	1
Tipo 4	Tipo 1	0,6
Tipo 4	Tipo 4	0,998
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,935
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,648
Tipo 4	Tipo 4	0,999

Tipo 3	Tipo 3	0,999
Tipo 3	Tipo 3	0,997
Tipo 3	Tipo 3	0,983
Tipo 3	Tipo 3	0,998
Tipo 3	Tipo 3	0,862
Tipo 3	Tipo 3	0,995
Tipo 3	Tipo 3	0,801
Tipo 3	Tipo 3	0,959
Tipo 3	Tipo 3	0,989
Tipo 3	Tipo 3	0,632
Tipo 3	Tipo 3	0,996
Tipo 3	Tipo 3	0,98
Tipo 2	Tipo 2	0,858
Tipo 2	Tipo 2	0,992
Tipo 2	Tipo 2	0,995
Tipo 2	Tipo 2	0,989
Tipo 2	Tipo 2	0,995
Tipo 2	Tipo 2	0,988
Tipo 2	Tipo 2	0,998
Tipo 2	Tipo 2	0,91
Tipo 2	Tipo 2	0,963
Tipo 1	Tipo 1	0,994
Tipo 1	Tipo 1	0,971
Tipo 4	Tipo 4	0,999
Tipo 4	Tipo 4	0,997
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	1
Tipo 4	Tipo 4	0,975
Tipo 4	Tipo 4	0,989
Tipo 4	Tipo 4	0,997
Tipo 4	Tipo 4	0,987
Tipo 4	Tipo 4	0,785
Tipo 4	Tipo 4	0,997