DETERMINACIÓN DE CARACTERISTICAS CONSTRUCTIVAS DE UN PREDIO MEDIANTE EL ANÁLISIS DE FACHADA POR MEDIO DE UNA RED NEURONAL ARTIFICIAL PARA UNA UPZ DE BOGOTÁ.

DANIEL FELIPE VANEGAS AVILA

20141025081

PROYECTO DE GRADO PARA OPTAR AL TÍTULO DE INGENIERO CATASTRAL Y GEODESTA BAJO LA MODALIDAD DE INVESTIGACIÓN – INNOVACIÓN

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

FACULTAD DE INGENIERÍA

INGENIERÍA CATASTRAL Y GEODESIA

BOGOTÁ, COLOMBIA

2020

DETERMINACIÓN DE CARACTERISTICAS CONSTRUCTIVAS DE UN PREDIO MEDIANTE EL ANÁLISIS DE FACHADA POR MEDIO DE UNA RED NEURONAL ARTIFICIAL PARA UNA UPZ DE BOGOTÁ.

PROYECTO DE GRADO PARA OPTAR AL TÍTULO DE INGENIERA CATASTRAL Y GEODESTA BAJO LA MODALIDAD DE INVESTIGACIÓN – INNOVACIÓN

DANIEL FELIPE VANEGAS AVILA

20141025081

DIRECTOR

INGENIERO CATASTRAL Y GEODESTA

EDWIN ROBERT PEREZ CARVAJAL

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

FACULTAD DE INGENIERÍA

INGENIERÍA CATASTRAL Y GEODESIA

BOGOTÁ, COLOMBIA

2020

Tabla de contenido

1. Resumen ejecutivo 6
2. Introducción 7
3. Descripción del problema 7
4. Objetivos 8
   4.1. General 8
   4.2. Específicos 9
5. Estado del arte. 9
   5.1. Propuesta metodológica para calcular el avalúo catastral de un predio utilizando redes neuronales artificiales. 9
   5.2. Metodología de precios hedónicos vs. Redes neuronales artificiales como alternativas a la valoración de inmuebles. Un caso real. 12
   5.3. Diseño y evaluación de un sistema experto difuso para la calificación de construcciones de tipología media y baja de uso residencial en el reconocimiento predial urbano. 13
6. Marco teórico 16
   6.1. Inteligencia Artificial 16
   6.2. Aprendizaje automático 18
   6.3. Aprendizaje profundo 19
   6.4. Redes neuronales artificiales 21
      6.4.1. Funciones de activación 24
      6.4.2. Algoritmo del gradiente descendiente (AGD) 28
      6.4.3. Función de coste 28
      6.4.4. Optimizadores 30
      6.4.5. Rango de aprendizaje 32
      6.4.6. Funciones de métrica 33
   6.5. Árbol de decisiones. 35
   6.6. Python 37
      6.6.1. Numpy 39
      6.6.2. Pandas 39
      6.6.3. SciKit-Learn 40
      6.6.4. Tensorflow 41
      6.6.5. Keras 42
   6.7. Reconocimiento predial urbano 43

6.7.1. Calificación de construcción. 43
7. Marco Normativo 44
7.1. Ley 14 de 1983 44
7.2. Decreto 3496 de 1983 44
7.3. CONPES 3641 de 2010 45
7.4. Resolución 070 de 2011 45
7.5. Ley 1753 de 2015 46
7.6. CONPES 3859 de 2016 47
7.7. CONPES 3958 de 2019 47
7.8. Ley 1955 de 2019 48
7.9. Decreto 148 de 2020 48
8. Marco Espacial – UPZ 65 Arborizadora 49
9. Metodología 53
9.1. Adquisición del conocimiento 53
9.2. Red neuronal artificial 54
9.2.1. Identificación de las variables de entrada y variables de salida. 55
9.2.2. Limpieza de los datos. 55
9.2.3. Entrenamiento de los datos. 56
9.2.4. Evaluación del modelo 56
9.2.5. Resultados y selección del modelo 56
9.3. Árbol de decisiones 62
10. Análisis de Resultados 63
11. Evaluación y Cumplimiento de Objetivos 69
12. Conclusiones y Recomendaciones 70
13. Referencias 71
14. Anexos 80

Lista de Figuras

Figura 1. Propuesta metodológica para calcular el avalúo catastral de un predio utilizando redes neuronales artificiales……11
Figura 2. Elementos de una neurona artificial…….23
Figura 3. Mapa conceptual, estructura de una Red Neuronal Artificial ……...……..23
Figura 4. Función de activación sigmoide……..25
Figura 5. Función de activación Tangente hiperbólica……...26
Figura 6. Función de activación ReLU……...27
Figura 7. Función de activación Softmax……...28
Figura 8. Tipos de construcción según su calificación……...44
Figura 9. Fases del aprendizaje supervisado……..54
Figura 10. Ajuste de las redes neuronales……..60
Figura 11. Ajuste de las redes neuronales……..60
Figura 12. Parámetros de las redes neuronales……..61
Figura 13. Parámetros de la red neuronal seleccionada…….63
Figura 14. Pérdida de entrenamiento Vs Pérdida de validación…….64
Figura 15. Ajuste de entrenamiento Vs Ajuste de validación…….64
Figura 16. Perdida y ajuste de la red neuronal en la evaluación…….65
Figura 17. Datos de entrada para la predicción de la red neuronal…….65
Figura 18. Vector de probabilidad de la predicción……66
Figura 19. Nodo del árbol de decisiones…….67
Figura 20. Ajuste de entrenamiento vs Ajuste de evaluación…….67
Figura 21. Datos de entrada para la predicción árbol de decisiones……...……67
Figura 22. Predicción árbol de decisiones…….…68
Figura 23. Importancia de las variables árbol de decisiones…….…68

1. Resumen ejecutivo

Los procesos catastrales de formación y actualización, reglamentados en la resolución 070 de 2011 del instituto geográfico Agustín Codazzi contempla una etapa de reconocimiento predial masivo mediante el cual se realiza una descripción detallada de las características constructivas del predio. Esta calificación está ligada a la subjetividad del reconocedor predial que por medio del conocimiento previamente adquirido puede afectar la calidad de la información.

Para la inicialización del proyecto es necesario generar una base de datos debidamente diligenciada con la información de las unidades constructivas de una unidad de planeación zonal de la ciudad de Bogotá plasmada en las fichas prediales que mediante el trabajo de los reconocedores prediales serán procesadas en una herramienta de inteligencia artificial con el objetivo de optimizar el proceso y generar una manera más rápida y optima de generar la descripción de los predios con variables de fácil acceso para el reconocedor.

El objetivo final del documento es presentar el desarrollo y resultado de una serie de experimentos de inteligencia artificial, red neuronal artificial y árbol de decisiones, que está dirigido a la calificación de las construcciones de uso residencial con miras a optimizar los procesos catastrales en el marco de la implementación del catastro multipropósito en el que el país se encuentra inmerso.

2. Introducción

El desarrollo de la inteligencia artificial y las herramientas que esta propone ha significado un gran avance en el avance de nuevas metodologías y tecnologías que han sido aplicadas en diferentes campos y aspectos de la sociedad que en han desembocado en una revolución en sí misma y de este modo en las ciudades y ambientes en los que se aplica.

Las redes neuronales artificiales nacen en el área de la inteligencia artificial, evocando la estructura biológica de las redes neuronales presentes en el cerebro, buscando que, mediante un aprendizaje previo, el sistema se pueda adaptar a las características que el medio le determina y pueda realizar el debido análisis y procesamiento de la información para brindar soluciones a algún problema.

Es por esto que en el presente estudio se propone desarrollar una metodología para la calificación y descripción predial, basada en redes neuronales artificiales utilizando como base las variables y atributos que existen en la ficha predial, la cual es utilizada en los procesos catastrales por un reconocedor predial, se pretende hacer óptimo el proceso de reconocimiento mejorando tiempos, precisión y disminuyendo la subjetividad existente en el procedimiento.

3. Descripción del problema

El artículo 3 de la ley 14 de 1983 establece que las autoridades catastrales serán las responsables de los procesos de formación, actualización y conservación cuyo objetivo es la identificación de información predial, esta información predial definida en el artículo 3 de la resolución 070 del 2011 consiste en el levantamiento de la información físico jurídica del predio que será en inscrita en la ficha predial. Esta información es levantada por los

reconocedores prediales que mediante capacitaciones y manuales adquiere los conocimientos necesarios para evaluar los aspectos de la construcción que describe la ficha predial.

Partiendo de lo anterior, la confianza y robustez de los datos se ve comprometida al hallarse ligada a la subjetividad del reconocedor que puede debilitar los datos recolectados en campo; La inteligencia artificial brinda herramientas que mediante la minería de datos, produce modelos que partiendo del aprendizaje automático y el aprendizaje profundo mejora exponencialmente el proceso de adquisición de la información predial mediante la generación de predicciones que parten de algoritmos entrenados y evaluados que en este caso, se espera generen la tipificación de las construcciones de uso residencial partiendo de información fácilmente identificable para el funcionario y que permita disminuir el tiempo del levantamiento predial, evitando la entrada del funcionario al predio y disminuyendo el tiempo que se toma cada funcionario por vivienda al recolectar la información; la finalidad del experimento es determinar qué tan efectivo es la implementación de herramientas como las rede neuronales artificiales y los árboles de decisiones en los procesos catastrales realizando comparaciones entre los valores obtenidos por cada tipo de algoritmo, enfrentándolo con el valor obtenido en campo por los reconocedores prediales.

4. Objetivos

4.1. General

Diseñar una red neuronal artificial para la determinación masiva de las características constructivas de los predios de uso residencial basados en la fachada.

4.2. Específicos

- Diseñar una red neuronal artificial para el apoyo a los procesos masivos de toma de información predial en lo relacionado con calificación de la construcción.
- Comparar diferentes tipos de redes y su desempeño en el proceso de pronóstico para seleccionar la más eficiente.
- Evaluar la calidad y precisión de los pronósticos resultantes de la red neuronal artificial seleccionada frente a las calificaciones realizadas en reconocimiento predial.

5. Estado del arte.

Al momento de realizar la búsqueda de artículos y literatura relacionada con la aplicación de redes neuronales en el catastro y específicamente en el reconocimiento predial, se encontró que su desarrollo es escaso, la bibliografía encontrada muestra la metodología de las redes neuronales artificiales en sectores ambientales y otras herramientas de la inteligencia artificial relacionadas con la toma de decisiones.

5.1. Propuesta metodológica para calcular el avalúo catastral de un predio utilizando redes neuronales artificiales.

Casas, A (2014)

El estudio tiene como objetivo la determinación del avalúo de un predio urbano empleando la metodología de Redes Neuronales Artificiales (RNA) que es el método econométrico menos tradicional del campo de la inteligencia artificial. Para el desarrollo del modelo se

usó las variables que el IGAC emplea tradicionalmente para el cálculo de los avalúos catastrales y se incorporaron variables de ubicación espacial de los predios apoyados en un Sistema de Información Geográfica. Los datos que se emplearon para la estimación del modelo fueron obtenidos en la investigación de mercado directa e indirecta realizada en el proceso de Actualización de la Formación Catastral de la zona urbana del Municipio de Fusagasugá que entró en vigencia el primero de enero de 2013. Para obtener el mejor modelo de RNA, en el que se seleccionen las variables que más contribuyen a la determinación del valor de un inmueble, se realizaron, numerosas pruebas, utilizando el algoritmo Perceptrón Multicapa de tipo supervisado MLP y el software estadístico SPSS V. 21.

Las variables explicativas incluidas en el modelo final para calcular el avalúo catastral de un predio han sido: la Área Terreno, Área C Privada, USOPOT, X Centroide, Distancia, Comuna, Puntaje Estructura, Puntaje Total, No Pisos, No Habitaciones, Edad, Y Centroide, Puntaje Cocina, Puntaje Acabados, Puntaje Baños y No Baños. La prueba paramétrica t de Student para datos pareados o muestras relacionadas, corroboró el mayor poder de la Red Neuronal Artificial para estimar los avalúos respecto de resultados obtenidos con la metodología del IGAC. Tras esto, se considera que las variables utilizadas para caracterizar el valor de un predio en el municipio de Fusagasugá, fácilmente se pueden adaptar a otras ciudades de Colombia.

La metodología que siguieron los autores (Figura 3) se visualiza a continuación.

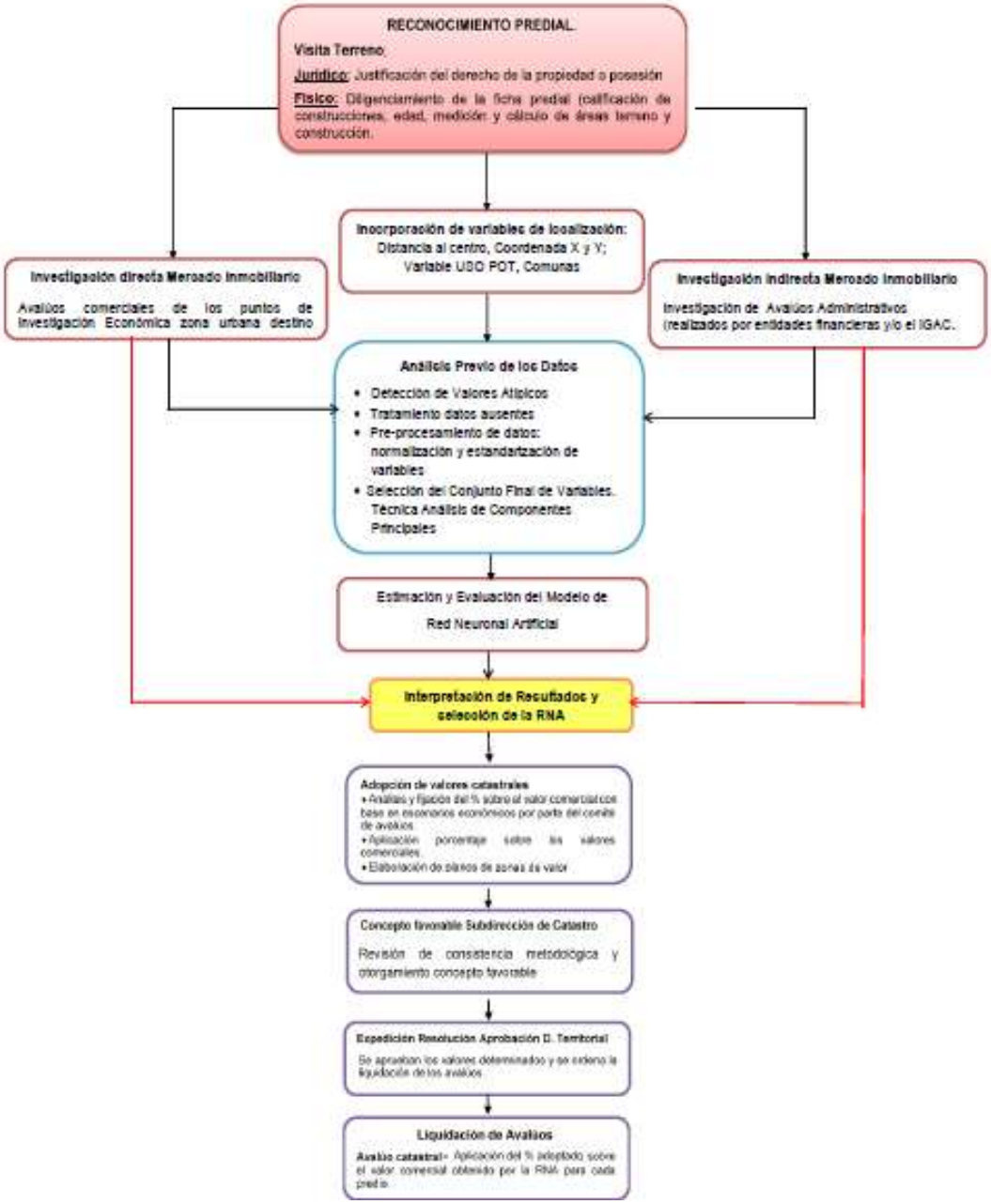


Figura 1. Propuesta metodológica para calcular el avalúo catastral de un predio

1utilizando redes neuronales artificiales. Fuente: Casas, 2014

5.2. Metodología de precios hedónicos vs. Redes neuronales artificiales como alternativas a la valoración de inmuebles. Un caso real.

Caridad, J., Núñez, J., & Ceular, M. (2008).

En este trabajo se examinan dos métodos de valoración aplicables al mercado inmobiliario: la Metodología de Precios Hedónicos, que analiza el precio del bien vivienda en función de sus principales características, frente a las Redes Neuronales Artificiales, que tratan de superar la inflexibilidad y linealidad de los modelos hedónicos tradicionales.

En el apartado empírico se trata de determinar el precio de la vivienda en Córdoba comparando ambas metodologías, concluyendo que la red presenta un mayor poder de predicción que las estimaciones obtenidas a través del modelo hedónico. El presente trabajo aborda el tema de la determinación de los precios dentro del mercado inmobiliario. Los precios de la vivienda en propiedad en España han experimentado desde mediados de los años ochenta y, especialmente, en la última década un crecimiento espectacular. Esto ha provocado que los consumidores destinen una parte muy significativa del presupuesto familiar a la adquisición del bien vivienda. También es preciso tener presente que existe una evidente preferencia sociológica de los españoles hacia la compra de viviendas frente al alquiler.

Por consiguiente, un mecanismo objetivo de valoración resulta de interés para numerosos colectivos, entre los que cabe citar los propios propietarios, constructores, agentes de la propiedad inmobiliaria, inversores, entidades tasadoras, financieras, aseguradoras o la propia Administración. A nivel internacional podrían situarse los orígenes del estudio del

mercado inmobiliario en la década de los cincuenta, siendo a partir de los estudios de Rosen, a mediados de los setenta, cuando comienza a analizarse el precio del inmueble en los núcleos urbanos aportando un marco metodológico completo utilizando modelos hedónicos. Este tipo de análisis ha sido relativamente frecuente fuera de nuestro país, pero comparativamente existen muy pocos estudios al respecto en España, donde en cualquier caso éstos se realizan desde el punto de vista macroeconómico, explicando eminentemente el comportamiento del agregado correspondiente a la inversión en construcción residencial o a las inversiones en vivienda. Por otra parte, existe un limitado número de trabajos en este ámbito que aplican la técnica de Redes Neuronales Artificiales.

Este trabajo utiliza la metodología hedónica y las redes para determinar empíricamente los precios de los inmuebles en la ciudad de Córdoba y efectuar una comparación entre el poder de predicción de ambas técnicas para terminar sugiriendo cuál es más apropiada para la valoración de una vivienda.

5.3. Diseño y evaluación de un sistema experto difuso para la calificación de construcciones de tipología media y baja de uso residencial en el reconocimiento predial urbano.

Cortés L. Laura V. (2020).

Este documento presenta el desarrollo y los resultados de un sistema experto difuso orientado a la determinación de las características constructivas de tipología baja y media para el uso residencial con el objetivo de crear una base de conocimientos plasmados en reglas difusas para que aquellos que realizan la función de reconocimiento predial bajo el análisis que seguiría un experto. Para realizar este experimento el autor generó una base de conocimiento surgida de entrevistas con expertos y revisión bibliográfica, siguiente a esto,

con ayuda de la base de datos se diseñó el sistema difuso donde se identifican las variables de entrada y salida, el tipo de sistema y cómo va a determinar las decisiones para al final evaluar los resultados comparando el método tradicional con el obtenido mediante el sistema de lógica difusa.

El autor define la lógica difusa como una rama de la inteligencia artificial que le permite a una computadora analizar la información tomada del mundo real y evaluarlo en una escala de falso y verdadero; la lógica difusa también permite tratar información imprecisa en términos de conjuntos difusos, que es una colección de objetos que verifican el grado de pertenencia o no a un conjunto, que se combinan para generar reglas que determinan acciones por lo que se simplifica el modelo de la siguiente manera: variables de entrada que se definen en conjuntos difusos que mediante un grupo de reglas generar uno o varios valores de salida.

La metodología aplicada por el autor en el desarrollo del proyecto fue dividida en varias fases que son acordes a los objetivos planteados, en primera estancia se realiza la obtención del conocimiento generando una base de datos bibliográfica y una base de datos predial mediante entrevistas a expertos para instruir el sistema experto difuso, el siguiente paso es el diseño del sistema mediante el cual se identifican las variables de entrada y las de salida, las funciones de pertenencia, la determinación de los conjuntos y las reglas del sistema para resultar en la elaboración del código en el software de elección, una vez diseñado el sistema se procede a desarrollarlo para generar una base de hechos que muestra los resultados del sistema que pasaran a análisis y evaluación para concluir sobre el desarrollo del proyecto.

Entre los resultados mostrados por el autor tenemos la base de conocimiento que almacena en forma de reglas de estructura" Sí.Entonces" el conocimiento adquirido mediante la revisión bibliográfica y las entrevistas a los expertos en el reconocimiento predial urbano en el sector de interés que son las construcciones de tipología baja y media de uso residencial. La entrevista realizada a los expertos se dividió en dos partes, la primera determinó los valores de las variables que determinan las funciones del sistema bajo los criterios que tienen en cuenta los reconocedores prediales, la segunda parte de la entrevista presenta las diferentes combinaciones posibles de las variables presentadas anteriormente para instruir al sistema experto sobre las decisiones a tomar.

El diseño del sistema inicia con la identificación de las variables las cuales divide en cuatro grandes capítulos con trece subgrupos. Las variables están dispuestas del siguiente modo: Estructura, con los capítulos de armazón, muros y cubierta. Acabados principales, con los capítulos de fachada, cubrimiento de los muros y pisos. Baño, con los capítulos de tamaño, enchapes y mobiliario, por último, Cocina, con los capítulos de tamaño, enchapes y mobiliario. A estas variables se le adicionó estado de conservación de las construcciones. Siguiente a esto se definen las funciones de membresía y los conjuntos difusos relacionado con aquellos datos que presentan mayor frecuencia en el conjunto, siguiente a esto, se definen las reglas del sistema a partir de las diferentes combinaciones que pudieran darse entre los parámetros de entrada en las construcciones. Una vez diseñado el sistema se procede a desarrollarlo en la aplicación Fuzzy Logic Designer que brinda mayores facilidades visuales y de interpretación en el desarrollo de los sistemas.

Una vez desarrollado el sistema se realizaron tratamientos diferentes en el experimento que corresponden a cada una de las combinaciones posibles mediante métodos de inferencia difusa y con ayuda de la base de hechos generada se evalúo de forma descriptiva e inferencial los resultados de cada uno de los sistemas del que se selecciona el que tuvo mejor desempeño al obtener los resultados más cercanos a los obtenidos mediante la consulta catastral en el portal del Instituto Geográfico Agustín Codazzi.

El autor concluye que calificar construcciones con el sistema experto difuso obtiene resultados cercanos a los obtenidos por los reconocedores prediales plasmados en el portal del Instituto Geográfico Agustín Codazzi en un 70% de las observaciones obtenidas en la base de hechos. También concluye que las variables plasmadas en la ficha predial, la cual evalúa las construcciones, tienen opciones reducidas en comparación con las que se deberían evaluar en cada característica según el manual de reconocimiento predial. Por último, resalta el hecho de que las variables recogidas por los usuarios en el sistema experto difuso son más que las recogidas mediante el sistema tradicional, pero de esta forma permite recoger información más precisa lo que permite un mayor acercamiento a la realidad de la construcción.

6. Marco teórico

6.1. Inteligencia Artificial

Los avances en la tecnología han permitido nuevos desarrollos en las diferentes áreas del conocimiento, entre estos conocimientos se encuentra la inteligencia artificial la que Márquez (2020) define como “Una disciplina perteneciente a las ciencias de la computación, que plantea modelos computacionales de aprendizaje basado en redes neuronales biológicas humanas. En este sentido, gracias a los avances en la tecnología

computacional, ha permitido desarrollos de sistemas “inteligentes” que facilitan procesar mayor cantidad de datos en un tiempo menor, agilizando la toma de decisiones.”.

La interdisciplinaridad de la inteligencia artificial que permite a las maquinas aprender, basándose en los datos de los que dispone, admite múltiples aplicaciones entre las que Ballesteros (2019) menciona “Predicción, recomendación, clasificación y reconocimiento, detección de objetos, procesamiento del lenguaje natural y sistemas autónomos entre otros”; La versatilidad de la inteligencia artificial ha generado proyectos en temas como medicina, robótica, medioambiente, sistemas de comunicación, análisis de grandes cantidades de datos y gestión del riesgo, etcétera.

La inteligencia artificial está estrechamente relacionada con la lógica y el aprendizaje, por esto, la inteligencia artificial es el diseño de herramientas que simulan la inteligencia humana, Avila et al, (2020) señalan que el objetivo de la inteligencia artificial es “diseñar herramientas informáticas que simulen procesos de inteligencia humana que incluyen el aprendizaje, el razonamiento y la autocorrección. Mediante diversos algoritmos las máquinas aprenden y son capaces de tomar decisiones”. También mencionan la teoría del procesamiento de la información de Mahoney que habla sobre el ser humano como una máquina que procesa información a partir de estímulos, de este modo comparan el funcionamiento humano y el procesamiento de la máquina de este modo. “El aprendizaje humano es progresivo y se desarrolla a partir de la exposición a situaciones de forma consecutiva, proceso que puede llevar años, en cambio, la adquisición de estos datos por parte de una máquina es mucho más rápida, ya que esta no se agota y no tiene funciones vitales” (Avila et al, 2020).

Es importante mencionar que la inteligencia artificial utiliza herramientas como el aprendizaje profundo y el aprendizaje automático cuya acción “converge en sistemas que trabajan con técnicas algorítmicas avanzadas para procesar diversos tipos de señales, optimizar modelos, procesar información masiva mediante Big Data y Clusters; todo ello mediante el uso algorítmicos de regresión, clasificación y clustering, entre otros” (Márquez, 2020) para llegar a ello utilizan procesos de aprendizaje supervisado, no supervisado y semi-supervisado que permiten la simulación de sistemas dinámicos.

Russell & Norvig (2020) nos hablan de 4 diferentes categorías en la inteligencia artificial: sistemas que piensan como humanos que emulan el pensamiento humano, sistemas que actúan como humanos que imitan el comportamiento humano, sistemas que piensan racionalmente que utilizan un pensamiento lógico racional y sistemas que actúan racionalmente los cuales imitan conductas humanas mediante el uso de un razonamiento lógico.

6.2. Aprendizaje automático

Zheng Li (2011) afirma que el procesamiento masivo de la información se ha convertido en un tópico urgente que se presenta tanto como un desafío tanto como una oportunidad para desarrollar técnicas de modelamiento masivo, que, se han vuelto de gran interés para las grandes compañías y han impactado en la vida cotidiana de las personas alrededor de todo el mundo; lo que a su vez ha generado que nuevas técnicas de procesamiento de la información surjan y sean aplicadas en múltiples campos que responden a las diferentes problemáticas generadas a nivel mundial, y el propósito de este tipo de técnicas es complementar y recopilar toda esta información y manejar el desafío que propone procesar la magnitud de información generada en el mundo globalizado.

Como nos cuenta Bagnato (2017), el aprendizaje automático o machine learning es un subcampo de la inteligencia artificial que busca resolver el problema sobre la construcción de programas que se mejoren automáticamente mediante la experiencia, por lo que por medio de esta definición se infiere que no es necesaria la intervención humana de una forma activa ya que el sistema por sí mismo mejora continuamente. De este modo, se infiere que la maquina usa algoritmos desarrollados para el análisis de datos para aprender de ellos y de este modo desarrollar predicciones; la maquina es entrenada por la gran cantidad datos que se le suministra brindado la oportunidad de que mejore los algoritmos por lo que se sintetiza que el aprendizaje automático no se trata de memorizar y recolectar información sino de crear modelos a partir de la información suministrada para generar conclusiones.

Entre las diferentes aplicaciones del aprendizaje profundo, Gonzalo (2019) menciona que se puede dividir en cuatro diferentes grandes campos como son el aprendizaje supervisado, que aborda problemas de clasificación y regresión, el aprendizaje no supervisado, que contempla problemas de clustering y asociación, el aprendizaje semi supervisado, que se aplica a casos de clasificación y clustering, y el aprendizaje reforzado, mayormente usado para problemas de clasificación y control; y la diferencia en la que se basa la diferencia entre cada uno de los cuatro campos del aprendizaje automático radica en la forma de los datos de entrenamiento así como la técnica mediante la cual se desarrolla el modelo.

6.3. Aprendizaje profundo

Para hablar de aprendizaje profundo, Akinosho et al (2020) aseguran que es necesario hablar primero de aprendizaje automático el cual definen como “La capacidad de los sistemas inteligentes para aprender y mejorar a través de la experiencia adquirida a partir de

un flujo de datos” , así, afirman que el mayor desafío para la inteligencia artificial es la representación apropiada de una serie de características en forma de datos que permitan al sistema a llegar a una decisión; este problema es el que aborda el aprendizaje profundo ya que a partir de múltiples capas, donde abstrae las características, para construir una representación compleja partiendo de bases simples.

Partiendo de lo anterior, Athanasiadou et al (2018) describen, “El aprendizaje profundo es un subcampo del aprendizaje automático que utiliza capas de redes neuronales artificiales para entrenar un sistema, identificar y clasificar los artículos”, de esta forma podemos afirmar que el aprendizaje profundo analiza patrones en la información disponible para poder distinguir e identificar características que han sido procesadas en diferentes capas del sistema para llegar a predicciones y clasificaciones.

El aprendizaje profundo ha tomado relevancia a partir de 2006 siendo principalmente impulsado por herramientas como el reconocimiento de voz, como describen Zhao et al (2019) el impulso del aprendizaje profundo se debe principalmente a tres factores “1. El surgimiento de grandes cantidades de información que puede ser entrenada para mostrar la gran capacidad de aprendizaje, 2. La gran velocidad de desarrollo de los sistemas de computación y 3. Los avances significativos en el diseño de las estructuras de red y las estrategias de entrenamiento “. Estos avances revolucionarios han demostrado la efectividad del aprendizaje profundo que resultan en el desarrollo de las redes neuronales profundas en las que cada capa, compuesta por neuronas, recibe la salida de la capa anterior como entrada generando un modelo estrechamente conectado que optimiza el proceso de aprendizaje.

La integración de las redes neuronales y el aprendizaje profundo han resultado en una mejora integral a las deficiencias que posee el modelo de redes neuronales como son el alto costo de mantenimiento, el movimiento masivo de datos y la capacidad limitada del modelo, con ayuda de dispositivos altamente optimizados reducir el tiempo de entrenamiento y acelerar el proceso de aprendizaje, este modelo implica dos fases como explican Venkatesan (2018) "1. Distribuir los datos de entrada entre los nodos del clúster en el que cada nodo entrena la parte de datos que recibe y 2. Reducción del modelo en la que se promedian los pesos y sesgos en cada nodo del clúster juntos". La herramienta que se utiliza para este proceso es la librería TensorFlow desarrollada por Google que permite a los usuarios construir sus redes propias con un entorno más agradable ya que cuenta con redes prefabricadas de fácil acceso para la comunidad.

6.4. Redes neuronales artificiales

Las Redes Neuronales Artificiales (RNA) son sistema de procesamiento de datos cuyo funcionamiento y estructura es similar al de las redes neuronales biológicas en las que un conjunto de nodos, cuyo nombre es neurona, están conectadas entre sí por un sistema de conexiones con un valor numérico determinado llamado peso; Luera & Mimin, (2014) afirman "Las redes neuronales artificiales están inspiradas en la arquitectura de los sistemas nerviosos biológicos y en la analogía matemática que consisten de un gran número de sistemas relativamente simples, que funcionan en paralelo para tomar decisiones rápidamente". De este modo, las RNA se definen como la adición e interacción de elementos simples supremamente interconectados que procesan información mediante sinapsis artificiales que asemejan el funcionamiento del cerebro humano.

Como afirma Matich (2001), el primero en concebir el estudio del cerebro humano como una máquina de computacional fue Alan Turing en 1936 pero fue en 1943 cuando el neurofisiólogo Warren McCulloch y el matemático Walter Pitts con su teoría "Un cálculo lógico de la inminente idea de la actividad nerviosa" quienes modelaron la primera red neuronal simple usando circuitos eléctricos, en 1957 Frank Rosenblat desarrollo una de las redes neuronales más antiguas con el concepto del perceptrón cuyo logro fue que el modelo fue capaz de reconocer una serie de patrones pero no era capaz de clasificar en clases. Desde la década de los ochenta el panorama ha ido mejorando respecto a las investigaciones y el desarrollo de las redes neuronales que en la actualidad han resultado en variados trabajos en diferentes áreas del conocimiento que se publican cada año y que ha permitido lanzamiento de nuevas herramientas como la simulación.

"La neurona es la unidad fundamental del sistema nervioso y en particular del cerebro. Cada neurona es una simple unidad procesadora que recibe y combina señales desde y hacia otras neuronas; El axón (salida) de la neurona se ramifica y está conectada a las dendritas (entradas) de otras neuronas a través de uniones llamadas sinapsis" (Bosogain, n.d.). Continuando el símil de la neurona biológica con la neurona artificial, cada neurona u elemento procesador, recibe varias entradas y con una suma básica las combina llevándolas a una salida que puede estar conectada a otras neuronas artificiales mediante una ponderación (peso) que corresponde a la eficacia de la sinapsis. La figura 1. nos muestra una representación gráfica de una neurona artificial.

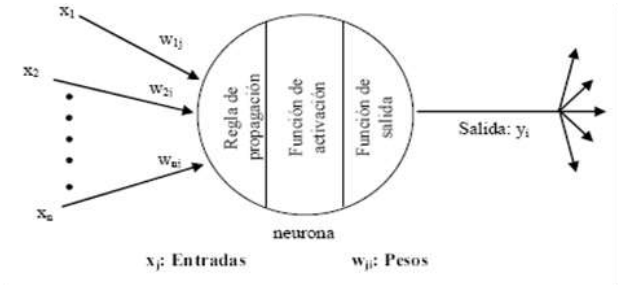


Figura 2. Elementos de una neurona artificial. Fuente: Loxa Ecuador, 2007.

La estructura de una red neuronal artificial (Figura 2), como mencionamos anteriormente, viene dada por una red multicapa de nodos interconectados que se puede distribuir en tres diferentes capas: la capa de entrada, la capa oculta y la capa de salida, el número de capas ocultas puede varias dependiendo del modelo y este se determina por el método de ensayo y error, en el, Luera & Peña (2014), afirman que a través del aprendizaje la red modifica los pesos como respuesta a la información de entrada que producen destrucción, modificación y creación de conexiones en la capa de salida para determinar el diseño óptimo.

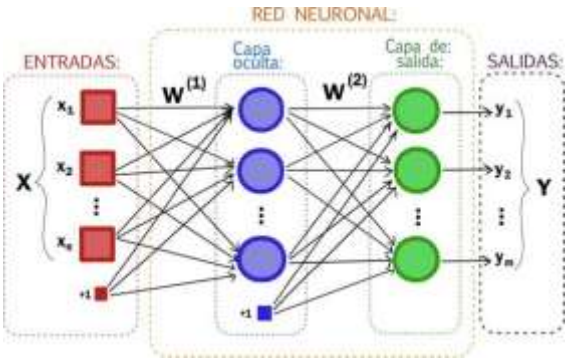


Figura 3. Mapa conceptual, estructura de una Red Neuronal Artificial. Fuente: Pinzón, 2018.

Matich (2001), resume las ventajas de las redes neuronales artificiales en cinco diferentes aspectos que son: La capacidad de aprender a realizar tareas basadas en un entrenamiento o una experiencia inicial (Aprendizaje adaptativo), la capacidad de crear su propia

organización o representación de la información recibida en su etapa de aprendizaje (Autoorganización), la Tolerancia a fallos que permite retener una parte de la estructura reduciendo la afectación al sistema, la operación en tiempo real y su fácil implementación dentro de la tecnología disponible actualmente.

6.4.1. Funciones de activación

Las funciones de activación de las capas intermedias son importantes para que la red comprenda el carácter de la información que se le está suministrando por lo que por medio de estas funciones se filtra la información antes de suminístrasela a la neurona en la que cada neurona está inactiva (-1, 0) o totalmente activa (1). (Wang et al, 2020).

6.4.1.1. Función sigmoide

La función sigmoide transforma los valores introducidos a una escala de (0, 1) donde los valore altos tienen de manera asintótica a 1 y los más bajos a 0. Esta función tiene mejor rendimiento en la última capa, pero tiene una lenta convergencia y una saturación baja del gradiente. Esta función está dada por la ecuación 1.

$$f(x) = \frac{1}{1 - e^{-x}}$$

Donde:

$f(x)$: $activación$

x: $entradas$

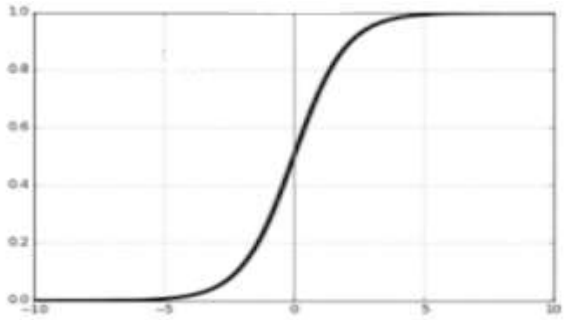


Figura 4. Función de activación sigmoide. Fuente: Bolton, 2018.

6.4.1.2. Función Tangente Hiperbólica.

La función tangente hiperbólica transforma los valores introducidos a una escala (-1,1), donde los valores altos tienen de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a -1. De manera similar a la función sigmoide esta también tiene baja saturación del gradiente y una lenta convergencia y tiene buenos rendimientos para redes recurrentes o redes en las que se tiene que elegir entre dos opciones. Esta función está dada por la ecuación 2.

$$f(x) = \frac{2}{1 + e^{-2x}} - 1$$

Donde:

$f(x)$: $activación$

x: $entradas$

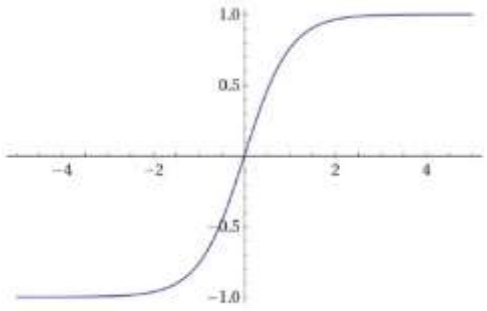


Figura 5. Función de activación Tangente hiperbólica. Fuente: Bolton, 2018.

5.4.1.3. Función unidad lineal rectificada (ReLU)

La función ReLU transforma los valores introducidos anulando los valores negativos y dejando los positivos tal y como entran a la red, este tipo de activación no se encuentra acotada por lo que tiene buen rendimiento con imágenes y redes convolucionales. Esta función está dada por la ecuación 3.

$$f(x) = \max(0, x) = \begin{bmatrix} 0 & si & x < 0 \\ x & si & x \geq 0 \end{bmatrix}$$

Donde:

$f(x)$: $activación$

x: $entradas$

max: $valor\ más\ grande\ de\ la\ función$

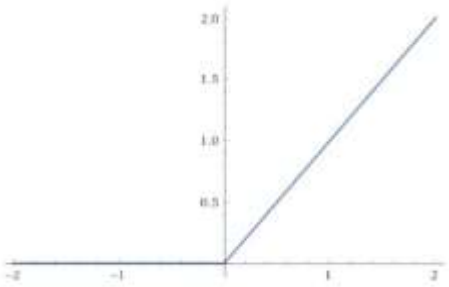


Figura 6. Función de activación ReLU. Fuente: Bolton, 2018.

6.4.1.4. Función Softmax

La función softmax transforma las salidas a una forma probabilística de tal manera que al sumar todas las probabilidades de las salidas da 1. Este tipo de función de activación se utiliza cuando queremos dar como resultado de la predicción una representación probabilística en el que puedas normalizar múltiples capas por lo que se tiene buen rendimiento en las últimas capas. Esta función está dada por la ecuación 4.

$$f(x) = \frac{e^{X}}{\sum_{J=1}^{K} e^{x_j}}$$

Donde:

$f(x)$: $activación$

x: $entradas$

k: $cantidad\ de\ entradas$

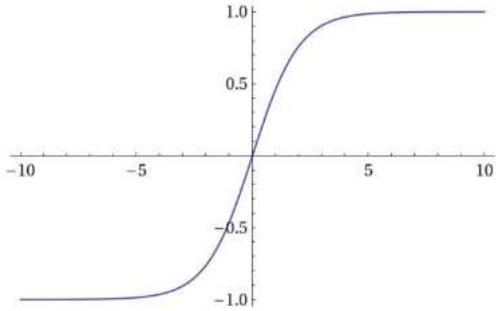


Figura 7. Función de activación Softmax. Fuente: Bolton, 2018.

6.4.2. Algoritmo del gradiente descendiente (AGD)

El AGD es un cálculo que permite saber cómo ajustar los parámetros de una red de tal modo que los valores de la desviación en su salida sean mínimos. Para hacer este cálculo realizamos un gradiente a la función de coste, con una derivada multivariable, con respecto a los parámetros de la red haciendo un énfasis en los pesos actuales. Con ayuda del método de propagación hacía atrás (Back-propagation), que facilita el cálculo de estos mínimos entre las capas de la red, calcula las derivadas parciales de la función de coste a la salida respecto a los parámetros de la capa anterior hasta llegar al inicio de la red del que se obtiene un vector gradiente. (Durán, 2020)

6.4.3. Función de coste

Este tipo de función mide el rendimiento del modelo para la información dada. La función de coste mide el error entre los valores predichos y los valores esperados mostrándolo en un único número real. Dependiendo del problema la función de coste se puede observar en diferentes formas. El propósito de la función de coste puede ser: minimizar, en el que el valor de retorno será la perdida, error o coste cuyo objetivo es ser el menor número posible,

o maximizar, donde el retorno será la ganancia donde el objetivo es obtener el número mayor posible (Krzyk, 2020)

Existen diferentes tipos de funciones que son usadas dependiendo del tipo de red que se esté desarrollando, tenemos funciones de coste de tipo probabilístico, de regresión y clasificadores de tipo máximo. A continuación, mencionaremos algunas de las funciones más usadas.

6.4.3.1. Entropía cruzada binaria (Binary Crossentropy)

Computa la perdida de entropía cruzada entre las etiquetas reales y las predichas, el uso de este tipo de función se realiza cuando solo tenemos dos clases dentro del sistema con características binarias de 0 y 1 en el que solo debe haber un punto por predicción.

6.4.3.2. Entropía cruzada categórica (Categorical Crossentropy)

Calcula la perdida de entropía cruzada entre las etiquetas y las predicciones, este tipo de función es usada cuando se tienen dos o más etiquetas, sin embargo, se espera que las etiquetas tengan una codificación categórica (One-hot) y debe existir un tipo determinado de clases por cada característica.

6.4.3.3. Entropía cruzada categórica dispersa (Sparse Categorical Crossentropy)

Este tipo de función, calcula la perdida de la entropía cruzada entre las etiquetas y las predicciones. Este tipo de función es más usada cuando las etiquetas recibidas son números enteros.

6.4.3.4. Error medio cuadrático (Mean Squared Error)

Calcula la media de los cuadrados de los errores entre las predicciones y las etiquetas; la perdida se produce debido a la aleatoriedad o porque el estimador no tiene en cuenta la información que podría producir una estimación más precisa.

6.4.3.5. Error medio absoluto (Mean Absolute Error)

Computa la media absoluta de la diferencia absoluta entre las predicciones y las etiquetas; este tipo de medida se usa para cuantificar la precisión de la predicción siendo la distancia vertical u horizontal entre los valores reales y las predicciones.

6.4.4. Optimizadores

Los optimizadores son usados en la propagación hacia atrás en el AGD y estos son los encargados de generar pesos cada vez mejores en la red neuronal, en el cual calculan el gradiente de la función de coste para cada peso de la red con base al rango de aprendizaje. (Velazco, 2020)

A continuación, se mencionan los optimizadores más importantes y usados más frecuentemente en procesos de aprendizaje profundo.

6.4.4.1. Descenso del gradiente estocástico (Stochastic Gradient Descent)

Este método de optimización es el más básico entre los sistemas de redes neuronales y la actualización de pesos se define en la ecuación 5, en este el rango de aprendizaje se mantiene fijo

$$W = W - \propto \Delta W$$

Donde:

W: pesos

∝: rango de aprendizaje

6.4.5.2. RMSprop

Este tipo de optimizador busca corregir los efectos negativos de la acumulación global de los caches mediante un promedio movible exponencial de los pesos, esta función se define mediante la ecuación 6 y la ecuación 7.

$$C = C + d * C(1 - d) * \Delta W^2$$

$$W = W - \frac{\propto * w^2}{\sqrt{C + \varepsilon}}$$

Donde:

C: suma de los cuadrados de los gradientes por cada parámetro

d: tasa de decadencia

W: pesos

∝: rango de aprendizaje

C: suma de cuadrados de los gradientes por cada parámetro

ε: valor cercano a cero.

6.4.5.3. Adam

El nombre real de este optimizador es Estimación adaptativa de momentos el cual tiene elementos similares al RMSprop pero añadiendo momentum y este se define en las ecuaciones 8, 9 y 10.

$$m = \beta_1 * m + (1 - \beta_1) * \Delta W$$

$$v = \beta_2 * v + (1 - \beta_2) * \Delta W^2$$

$$W = W - \frac{\propto * m}{\sqrt{v + \varepsilon}}$$

Donde:

m: momento que modela la media de los gradientes a lo largo del tiempo

v: momento que modela la varianza de los gradientes a lo largo del tiempo

β_1: valor fijo 0.9

β_2: valor fijo 0.999

W: pesos

$\propto$: rango de aprendizaje

ε: valor cercano a cero.

6.4.5. Rango de aprendizaje

El rango de aprendizaje o tasa de aprendizaje es un parámetro que controla cuanto cambia el modelo en relación al error estimado cuando los pesos del modelo son actualizados.

Entre más pequeño sea este valor más lento será el aprendizaje, caso contrario, al elegir un valor mayor la velocidad también aumentara provocando un proceso de aprendizaje inestable, esta toma valores entre 0 y 1. (Understand the Impact of Learning Rate on Neural Network Performance, 2020)

6.4.6. Funciones de métrica

Similar a las funciones de perdida, las funciones de métrica son usadas para evaluar la eficacia del modelo con excepción de que los resultados de estas funciones no van a ser usados dentro del modelo; Depende del objetivo del modelo se puede usar diferente tipo de función de métrica: las de exactitud miden que tanto las funciones coinciden con las etiquetas, las de probabilidad miden la entropía cruzada entre las etiquetas y las predicciones, las de regresión cuantifican el error medio cuadrado entre las predicciones y las esperadas, entre otras. (Keras, n.p.)

A continuación, se mencionan las funciones de métrica mayormente usadas en proyectos de redes neuronales y sistemas de aprendizaje profundo.

6.4.6.1. Ajuste

Calcula la frecuencia en la que las predicciones son iguales a las etiquetas. Esta métrica crea dos variables locales que son usadas para calcular la frecuencia con la que las etiquetas coinciden con las etiquetas.

6.4.6.2. Ajuste binario

Computa la frecuencia con la que las predicciones coinciden con las etiquetas binarias, similar a la anterior, crea dos variables locales que se usan para medir la frecuencia con que las predicciones coinciden con las etiquetas.

6.4.6.3. Entropía cruzada binaria

Calcula la entropía cruzada entre las predicciones y las etiquetas, mayormente usada cuando solo se tienen dos tipos de clases en el sistema (0 y 1).

6.4.6.4. Entropía cruzada categórica

Computa la entropía cruzada categórica entre las predicciones y las etiquetas, este tipo de métrica es usada cuando se tiene una clase con múltiples etiquetas (2 o más).

6.4.6.5. Error medio cuadrático

Como el nombre de la métrica lo dice, esta clase calcula el error medio cuadrático existente entre las predicciones y las etiquetas.

6.4.6.6. Error medio absoluto.

Similar a la clase anterior, la clase calcula el error medio absoluto entre las etiquetas y las predicciones de la red.

6.5. Árbol de decisiones.

Los algoritmos de decisión basados en árboles mejor conocidos como arboles de decisiones se consideran como uno de los mejores métodos utilizados en el aprendizaje automático de carácter supervisado ya que “Los métodos basados en árboles potencian los modelos predictivos con alta precisión, estabilidad y facilidad de interpretación” (Árbol de Decisión En Machine Learning (Parte 1) - Sitiobigdata.Com, n.d.). Su nombre se debe a que su estructura gráfica se asemeja a las ramas de un árbol ya que tenemos múltiples alternativas que se irán dibujando y estas nuevas alternativas a su vez se dividen en más. Como mencionan Ravina et al (2018) “Los árboles de decisión recogerán las diferentes alternativas con las que vaya a tener que lidiar un determinado sujeto para decantarse por aquella que considere más útil”.

Entre las aplicaciones que tienen los árboles de decisión se encuentran principalmente dos tipos de problemas: de regresión, en los que se predice el valor de una variable continua, y los de clasificación, en los que se predice valores de una variable categórica. Los árboles de decisión constan de una serie de nodos u hojas, que en su interior contienen una prueba lógica que considera un único atributo, este resultado arroja una clase que le es asignada a cada hoja que representa un valor según sea adecuado. Continuando el símil con un árbol biológico, el árbol también consta de ramas que corresponde a rangos de valores que deben ser mutuamente excluyentes y completos. Rokach & Maimon (2017) definen así: “Estas dos propiedades de desarticulación e integridad son importantes, ya que garantizan que cada instancia de datos se asigna a una categoría. Comenzamos con una raíz de un árbol; consideramos la característica que corresponde a la raíz y definimos a qué rama corresponde el valor observado de la característica dada. A continuación, consideramos el

nodo en el que aparece la rama dada. Repetimos las mismas operaciones para este nodo hasta que lleguemos a una hoja".

Para este tipo de algoritmo, Echavarría et al (2013), describen 4 tipos de nodos: 1. Nodo raíz que inicia el árbol de decisión que recibe la muestra total que plantea una decisión con varias alternativas y resultados desconocidos, 2. Nodo de decisión, que toma diferentes caminos que son seleccionados a partir de decisiones previas, 3. Nodos de eventos que son los posibles resultados con su debida probabilidad de ocurrencia y 4. Los nodos finales que muestran los resultados generados por la serie de decisiones tomadas previamente por el árbol.

Diferentes autores señalan que, aunque los árboles de decisión son una gran herramienta para la minería de datos y el aprendizaje automático estos también poseen diferentes desventajas que ponen en riesgo la precisión del modelo. "las principales debilidades de esta metodología se basan en la subjetividad, en la definición de las probabilidades que se asignan a cada uno de los escenarios y en el uso de la misma tasa de descuento para todas las ramas y para todos los periodos." (Echavarría et al, 2013). Sin embargo, el análisis que se realice a los resultados dados por el modelo junto con el valor esperado, la integridad del modelo y la facilidad de su interpretación, puede superar esas debilidades y ser una alternativa consistente a la hora de tomar decisiones.

6.6. Python

“Python es uno de los lenguajes de programación más populares para el procesamiento de datos y gracias a que es una comunidad de código abierto, una gran cantidad de librerías sobre aprendizaje automático y ciencias de la información han sido desarrolladas.” (Raschka & Mirjalili, 2019). Como mencionan los autores, el lenguaje de programación Python es un lenguaje de programación multiparadigma cuya filosofía se ha basado en un tipado dinámico y fluido que facilite su interpretación y usabilidad que pretende ser entendido tanto por desarrolladores experimentados como principiantes que mediante su facilidad de lectura y una sintaxis cercana al lenguaje natural se ha popularizado a gran escala.

Entre las principales características que posee el lenguaje de programación Python, Oliphant (2007), se encuentra que al ser un lenguaje de código abierto permite compartir y difundir de una manera totalmente libre los programas y aplicaciones desarrollados mediante este lenguaje de programación sin necesidad de pedir ningún tipo de licenciamiento y es que el hecho de que Python puede ser ejecutado en variadas plataformas significa que no debemos limitarnos al momento de crear una aplicación sin miedo de tener limitaciones o restricciones; La sintaxis de fácil acceso e interpretación permite escribir y compartir líneas de código con bondad hacia el usuario y el desarrollador que permite optimizar el tiempo de desarrollo y aumenta la productividad y gracias a esto se puede implementar código de Python en sistemas ya existentes y gracias a ser un lenguaje de código abierto, posee un gran número de librerías que permite construir programas sofisticados con un balance automático que permite manejar grandes cantidades de información con recursos limitados y al ser una comunidad tan activa, se ha desarrollado

de múltiples y diversas formas que han desarrollado un repositorio disponible en la web dónde se recibe soporte y ayuda sobre los recursos de Python.

“Aunque Python no fue desarrollado por satisfacer las necesidades de la comunidad científica, captó rápidamente el interés de los científicos e ingenieros que debido al syntaxis sencillo del lenguaje y la versatilidad de manejo de diferentes tipos de dato, se convirtió en el espacio de trabajo ideal para las ciencias de la comutación” (Millman & Aivazis, 2011); el año 1995 se formó la comunidad de Python alrededor de un grupo especializado interesado en el desarrollo de un nuevo tipo de dato desarrollaron una extensión del lenguaje de programación C llamado Numeric, basado en un objeto matricial que fue incorporando más ideas. Durante esta primera etapa, la interacción entre la comunidad científica y estándar de Python empezó a crecer gradualmente y aumentar su interacción, el resultado de este proceso fue la creación de nuevas características y una nueva sintaxis que fue requerida especialmente por la comunidad científica, mientras estos múltiples cambios fueron desarrollados la comunidad fue guiando Python hacía una sintaxis más sencilla y fácil de leer y manipular. En el año 2000 hubo un número creciente de módulos desarrollados que creó un mayor interés de crear un ambiente de desarrollo para las ciencias de datos dentro de Python lo que resultó en la implementación de un ambiente apto para la implementación de múltiples librerías aptas para la ciencia de datos, entre esas librerías se han orientado algunas hacía la minería de datos y la inteligencia artificial entre las cuales tenemos Numpy, Pandas, Matplotlib, tensorflow, entre otras; a continuación mencionaremos las librerías más populares de Python, en el desarrollo de programas de inteligencia artificial.

6.6.1. Numpy

Numpy o Numerical Python, de donde toma su nombre, es la librería principal para la informática científica que proporciona potentes estructuras de datos, mediante el manejo de matrices que garantizan cálculos eficientes. Es una biblioteca de Python que proporciona matrices y una variedad de rutinas que incluye procesos matemáticos, lógicos, interpretación y manipulación de formas, álgebra lineal básica, simulación aleatoria y operaciones estadísticas básicas, entre otras; El núcleo de funcionamiento de NumPy es el objeto ndarray, que encapsula matrices n-dimensionales de tipos de datos homogéneos. (What is numpy- numpy.org, 2008). Entre las ventajas de Numpy se encuentra la vectorización de la información que trae múltiples virtudes al código entre las cuales está: el código vectorizado es más conciso y fácil de leer, permite describir funciones en menos líneas de código que generalmente conlleva menos errores, la codificación se parece más a la notación matemática estándar que facilita su implementación e interpretación que brinda flexibilidad que convierte a Numpy en el lenguaje más sencillo para el intercambio multidimensional de datos.(Introduccion a Numpy- aprendeia.com, 2008).

6.6.2. Pandas

Pandas es una biblioteca mayormente usada para manipulación y análisis de datos en Python, esta librería ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales. El tipo de dato usado en esta librería es conocido como DataFrame que es “una colección ordenada de columnas con nombres y tipos, parecido a una tabla de base de datos, donde una sola fila representa un único caso y las columnas

representan atributos particulares.” (Introducción a pandas-programación.net, 2012), que permite una manipulación de las estructuras de datos en formatos de archivos variados, gracias a esto permite la alineación de dato y manejo integrado de datos fallantes, la reestructuración y segmentación de los datos, la inserción y eliminación de columnas en estructuras de datos, entre otros; resumiendo concluimos que gracias a la sencillez de Pandas y la versatilidad de su uso, se considera una herramienta eficaz y apta para usuarios poco expertos en programación y una de las librerías más demandadas en el lenguaje de programación Python. (Pandas: herramienta básica para el data science en Python, Universidad de Alcalá, 2021).

6.6.3. SciKit-Learn

La librería Scikit-Learn es un recurso de código abierto que soporta y brinda herramientas para el aprendizaje supervisado y no supervisado, también provee herramientas para el diseño de modelos, preprocesamiento de la información, selección de modelos y evaluación entre otros. (Scikit-learn.org, 2007) Esta librería está basada en NumPy, SciPy y Matplotlib. que con una interfaz fácil de aprender ha sido ampliamente usada para problemas tanto de clasificación como de regresión que, además, brinda ventajas con los hiperparámetros que traen consigo valores por defecto que son adecuados para su uso en varios casos. La gran variedad de usos de Scikit-Learn la convierten en una herramienta básica para empezar a programar y estructurar sistemas de análisis de datos y modelado estadístico.

La librería de Scikit-Learn toma elementos de la mayoría de algoritmos de aprendizaje automático o machine learning , desde modelos lineales como la regresión lineal hasta métodos bayesianos como los árboles de decisión; también existen varios métodos para

verificar la precisión de los modelos como la validación cruzada que la librería provee para implementar el método demasiadas líneas de código (Librería Scikit-Learn-aprendeia.com, n.d., 2018). Para concluir una de las principales razones por las que la librería se ha popularizado y ha estado en constante desarrollo es debido a su código abierto que con más de 35 colaboradores que contribuyen a construir una gran comunidad entorno a ella.

6.6.4. Tensorflow

La librería de Tensorflow, desarrollada por Google, es mayormente usada para realizar cálculos mediante diagramas de flujo, de esta forma, al momento de procesar no manipulamos un programa sino grafos en los que los nodos de los grafos representan operaciones matemáticas y las aristas representas tensores o matrices n-dimensionales, es por esto que la librería es mayormente usada para aprendizaje profundo y aplicaciones de cálculo científico (iartificial.net, 2015). Una de las características más relevantes de Tensorflow es que es una plataforma de código abierto que cuenta con un ecosistema flexible con recursos y herramientas creadas por la misma comunidad qye permite a los investigadores impulsar proyectos de aprendizaje automático con facilidad de aplicación para los desarrolladores.

Entre las principales ventajas de Tensorflow se encuentra la compilación sencilla de modelos, que permite una iteración de modelos inmediata y de fácil depuración, su alojamiento en la nube permite mayor facilidad para el desarrollador o programador ya que le permite entrenar modelos desde la nube, navegador o cualquier otro dispositivo sin importar el lenguaje que se esté utilizando(tensorflow.org, 2015); por último, la comunidad de Tensorflow resalta la experimentación y la investigación como uno de los pilares de la librería, esto se refleja en que Tensorflow compila modelos sin sacrificar velocidad ni

rendimiento lo que brinda mayor flexibilidad y control de las funciones para una fácil depuración que permite una ejecución inmediata.

6.6.5. Keras

La librería Keras es una interfaz de alto nivel diseñada para el desarrollo de redes neuronales, esta librería escrita en Python se basa en Tensorflow, Microsoft Cognitive Toolkit o Theano, que fue diseñada con el objetivo de facilitar el proceso de experimentación rápida que por medio de Keras permite un acceso rápido e intuitivo de las herramientas de las demás librerías. Además de lo mencionado anteriormente, Keras es una biblioteca de código abierto que mejora la usabilidad de las herramientas por lo que se considera como una herramienta orientada al uso por personas principalmente y por maquinas en segundo lugar. (Keras.io, n.d.).

Las acciones de los usuarios se reducen al mínimo lo que mejora la productividad que sumado con la conexión a los sistemas de aprendizaje profundo brinda múltiples ventajas entre las que tenemos la amplia compatibilidad entre las diversas plataformas para los modelos desarrollados, también, tenemos soporte para múltiples unidades de procesamiento gráfico (GPU) que permite que los recursos necesarios para desarrollar los procesos de aprendizaje profundo sean distribuidos con mayor facilidad a varios chips o tarjetas gráficas y debido a su popularidad y facilidad de desarrollo múltiples compañías han brindado mantenimiento y desarrollo de Keras lo que ha convertido a la biblioteca en una interfaz universal para múltiples plataformas de aprendizaje automático y de gran utilidad en el sector de la inteligencia artificial (Keras.io, n.d.).

6.7. Reconocimiento predial urbano

El manual de reconocimiento predial define el reconocimiento predial como un proceso efectuado sobre predios que han sido debidamente localizados dentro del perímetro urbano que tiene como objetivo identificar sus aspectos físico, que es la identificación de los linderos del terreno y edificaciones del predio, y jurídico, en el que se pretende indicar la relación entre el propietario o poseedor y el bien inmueble en los procesos de formación y actualización predial. (Instituto Geográfico Agustín Codazzi, IGAC, 1997).

6.7.1. Calificación de construcción.

El manual reconoce la calificación de la construcción como la puntuación de la unidad de construcción que define como "La edificación dentro de un predio que tiene unas características específicas en cuanto a uso y elementos constitutivos físicos de la misma, en un predio se pueden presentar unidades residenciales, comerciales o industriales que se diferencian en sus características físicas y uso que determinan un puntaje de calificación" (IGAC, 1997).

"El proceso de calificación consiste en convertir las características observadas del predio y anotadas en el formulario respectivo, el puntaje final del predio es el resumen del comportamiento representado en los valores asignados" (IGAC, 1997). El manual de reconocimiento predial reconoce cuatro variables que en su conjunto representan la construcción que son: estructura, acabados principales, baño y cocina, el cual, una vez reconocido se sumaran los puntajes obteniendo un total que será la calificación total del predio.

La calificación total del predio da un puntaje que nos ayuda a clasificar las unidades residenciales en diferentes categorías que han sido reunidas según características similares obteniendo seis tipologías del modo mostrado en la Figura 6.

Tipo	Rango de Puntaje
1. Tipo 1	0-10
2. Tipo 2	11-28
3. Tipo 3	29-46
4. Tipo 4	47-64
5. Tipo 5	65-82
6. Tipo 6	Mayor a 82

Figura 8. Tipos de construcción según su calificación. Fuente: IGAC, 1997.

7. Marco Normativo

7.1. Ley 14 de 1983

La ley 14 expedida el 6 de julio de 1983 cuyo objeto es fortalecer los fiscos de las entidades territoriales y se dictan otras disposiciones, en el artículo 3 del capítulo 1 dicta que las autoridades catastrales serán responsables de los procesos catastrales de formación, actualización y conservación del catastro, con el objetivo de que la identificación referente a los aspectos físico, jurídico, fiscal y económico, sea correcta. Es el articulo 4 el que establece que el avaluó de cada predio se obtendrá a partir de sumar los avalúos parciales de terreno y construcción.

7.2. Decreto 3496 de 1983

El decreto reglamenta parcialmente la la ley 14 de 1983 en el que en su artículo 2 define el catastro como “el inventario o censo, debidamente actualizado y clasificado, de los bienes inmuebles pertenecientes al Estado y a los particulares, con el objeto de lograr su correcta identificación física, jurídica, fiscal y económica”. El articulo 3 define que el aspecto físico

contiene la identificación de las edificaciones sobre documentos gráficos, descripción y clasificación.

7.3. CONPES 3641 de 2010

El objeto de este documento del Consejo Nacional de Política Económica y Social (CONPES) expedido en 2010 es consolidar la relación entre catastro y registro en el cual se identifica principalmente la dificultad de intercambio, mantenimiento y acceso a la información catastral. En él se crea una estrategia para consolidar el Sistema de Información Catastral (SIC) en el que dispone que la información de la ficha predial y de los registros 1 y 2 será incluida dentro del SIC con el que se complementa la información alfanumérica.

7.4. Resolución 070 de 2011

La resolución 070 expedida por el IGAC el 4 de febrero de 2011 tiene como objeto reglamentar técnicamente los procesos catastrales: formación, actualización y conservación, garantizando la seguridad jurídica y el uso de tecnologías para avalar el intercambio y aprovechamiento de la información catastral por parte de los usuarios y de las autoridades catastrales. Entre sus artículos se definen gran cantidad de instrumentos para las labores catastrales como los siguientes:

Artículo 21: Construcción o edificación. - Es la unión de materiales adheridos al terreno, con carácter de permanente, cualesquiera sean los elementos que la constituyan. (IGAC, 2011)

Artículo 31: Identificación predial. - Es el levantamiento de la información y la verificación de los elementos físico y jurídico del predio, mediante la práctica de la inspección catastral

y demás medios probatorios para identificar en documentos cartográficos y/o catastrales su ubicación, linderos, extensión, construcciones y/o edificaciones, y precisar el derecho de propiedad o posesión. (IGAC, 2011)

Artículo 33: Ficha predial. - Es el documento, en medio análogo o digital en el cual se consigna la información correspondiente a cada uno de los predios de la unidad orgánica catastral según el modelo que determine el Instituto Geográfico “Agustín Codazzi”. (IGAC, 2011)

Artículo 47.- Tipología de las construcciones y/o edificaciones. - Son las características de diseño y constructivas especiales de las construcciones y/o edificaciones. La tipología de las construcciones y/o edificaciones aplicables a nivel nacional, será la que establezca el Instituto Geográfico “Agustín Codazzi”. (IGAC, 2011).

7.5. Ley 1753 de 2015

La ley 1753 del 9 de junio de 2015 por el cual se expide el plan nacional de desarrollo 2014-2018 en su artículo 104 dice lo siguiente: Se promoverá la implementación del catastro nacional con enfoque multipropósito, entendido como aquel que dispone información predial para contribuir a la seguridad jurídica del derecho de propiedad inmueble, al fortalecimiento de los fiscos locales, al ordenamiento territorial y la planeación social y económica. El Gobierno nacional, a través del Instituto Geográfico Agustín Codazzi (IGAC) con el apoyo de los catastros descentralizados, podrá realizar las actividades necesarias para la formación y actualización catastral de manera gradual e integral, con fines adicionales a los fiscales señalados en la Ley 14 de 1983, logrando plena coherencia entre el catastro y el registro, mediante levantamientos por barrido y predial

masivo, en los municipios y/o zonas priorizadas con el Departamento Administrativo Nacional de Estadística, el Departamento Nacional de Planeación, el Ministerio de Agricultura y Desarrollo Rural y el Ministerio de Ambiente y Desarrollo Sostenible, conforme con la metodología definida para el efecto (Ley 1753, 2015).

7.6. CONPES 3859 de 2016

El documento CONPES del 13 de junio de 2016 propone una política para la implementación del catastro multipropósito que sea moderno, completo, confiable, consistente con el sistema de registro de la propiedad inmueble, integrado con los sistemas de planeación e información del territorio, y en concordancia con estándares internacionales. En este también se menciona que el IGAC debe renovar la metodología para que se estimen valores integrales de los predios acordes con la dinámica del mercado inmobiliario, y no a través de valores parciales de terreno y construcción como se había realizado desde hace tres décadas. (CONPES 3859, 2016)

7.7. CONPES 3958 de 2019

El documento CONPES 3958 del 26 de abril de 2019 traza una nueva estrategia para la implementación del catastro multipropósito como política pública, esta se basa y recoge la fase piloto implementada en el CONPES 3859, En el documento se contemplan estrategias para definir el nuevo modelo institucional para la implementación del catastro multipropósito, asegurar recursos para la gestión catastral con el objetivo de tener actualizada la base de datos catastral.

7.8. Ley 1955 de 2019

Esta ley expedida el 25 de mayo de 2019 tiene como objetivo dictar el plan nacional de desarrollo 2018-2022, En este documento se define que la gestión catastral será un servicio público en su artículo 79 que dice lo siguiente:

La gestión catastral es un servicio público que comprende un conjunto de operaciones técnicas y administrativas orientadas a la adecuada formación, actualización, conservación y difusión de la información catastral, así como los procedimientos del enfoque catastral multipropósito que sean adoptados. (Congreso de Colombia, 2019).

Las actividades catastrales estarán reguladas por el IGAC, en el cual los gestores catastrales estarán encargados de los procesos catastrales entre los que se incluye los de ámbito multipropósito y los operadores catastrales que estarán encargados de desarrollar las tareas operativas de la gestión catastral. También se contempla que la Superintendencia de Notariado y Registro (SNR) tendrá una labor de inspección y control de la gestión catastral que llevan los gestores y operadores catastrales.

7.9. Decreto 148 de 2020

Este decreto expedido por la presidencia de la república el 4 de febrero de 2020 reglamenta los artículos de la ley 1955 de 2019 por lo cual se reglamenta la gestión catastral como un servicio público, en este se cambia la definición de avalúo catastral adaptándola al concepto multipropósito. También en su artículo 2.2.2.2.1. cambia la definición de la información catastral quedando de esta forma:

La información catastral. Corresponde a las características físicas, jurídicas y económicas de los bienes inmuebles. Dicha información constituirá la base catastral y deberá ser

reportada por los gestores catastrales en el Sistema Nacional de Información Catastral (SINIC) o en la herramienta tecnológica que haga sus veces, de acuerdo con los estándares y especificaciones técnicas definidas por la autoridad reguladora (Decreto 148, 2020).

8. Marco Espacial – UPZ 65 Arborizadora

La Unidad de Planeación Zonal (UPZ) Arborizadora está ubicada en la localidad de Ciudad Bolívar, tiene un área total de 306.3 hectáreas con un área equivalente al 9% de la localidad de Ciudad Bolívar y está ubicada en la región nororiental de la localidad (Anexo 1). La UPZ limita al norte con la avenida Ferrocarril del Sur y la UPZ Apogeo de la localidad de Bosa, por el oriente con el río Tunjuelo y la UPZ Venecia de la localidad de Tunjuelito, por el sur con la avenida Ciudad de Villavicencio y la UPZ San Francisco de la localidad de Ciudad Bolívar, y por el occidente con la avenida Ciudad de Villavicencio y las UPZ Ismael Perdomo y Jerusalén. (Unidad Administrativa Especial de Catastro Distrital, UAECD, 2013).

En la actualidad Ciudad Bolívar cuenta con vías principales que facilitan el acceso de las personas a la localidad, como la Autopista, la Avenida Ciudad de Villavicencio, la Avenida Boyacá, y la Avenida Jorge Gaitán Cortés; La topografía de la localidad es un 90% montañosa y tiene una gran extensión en la zona rural que corresponde a un 72% de la localidad y cuenta con una temperatura promedio de 14° C. ((*Conociendo Mi Localidad*, n.d.)

La UPZ 65 está constituida por los siguientes barrios: Atlanta, Arborizadora Baja, La Playa, Madelena, Rafael Escamilla, Santa Helena, Santa Rosa Sur, Villa Helena, Casa Linda, La

Coruña y las urbanizaciones Protecho, Casa Linda, La Coruña y Atlanta (Anexo 2) ((*Conociendo Mi Localidad*, n.d.).

El uso del suelo en la localidad se distribuye de la siguiente manera; (Anexo 3) el uso de suelo en vivienda NPH representó la mayor proporción en el año 2002, registrando 8.894 unidades de uso con 700.168 m2, lo que representó un 63,9% de participación frente al total de área construida de todos los usos de la UPZ. Para el año 2012, aunque el porcentaje de participación disminuyó al 59,1%; sin embargo, el número de unidades subió a 9.763, y un aumento a 895.101 m2 construidos, y manteniéndose en el primer lugar de participación. En segundo lugar, se ubica el uso de vivienda en PH, que figuró en el año 2002 con 3.561 unidades de uso con 199.466 m2 construidos y una participación de 18,2%, el cual aumento a un 23,8% de participación en el año 2012, con un incremento en unidades de uso a 6.669 e igualmente aumentó a 360.360 m2 construidos, así mismo, los usos de industria, bodega y almacenamiento, se ubicaron en tercer lugar en el año 2002; los cuales registraron 13 y 56 unidades con 62.877 m2 y 58.328 m2 construidos y un 5,7% y 5,3% de participación respectivamente (UAECD, 2013).

El manual de procedimiento para la determinación de zonas homogéneas físicas urbanas y rurales define área de actividad como la destinación establecida a zonas del suelo urbano para la asignación de usos en función de la estructura urbana propuesta por el modelo territorial. (UAECD, 2005) En la UPZ se encuentran 5 tratamientos y zonas sin ningún tipo de clasificación.

El área de actividad dotacional corresponde al 0.64% del total de la UPZ, estas zonas de actividad dotacional pueden ser: zonas para el desarrollo de instalaciones educativas,

culturales, de salud, de bienestar social y de culto, de instalaciones deportivas y recreativas, de servicios relacionados con seguridad ciudadana, defensa y justicia, abastecimiento de alimentos, recintos feriales, cementerios y servicios funerarios, servicios de administración pública, servicios públicos y de transporte, zonas definidas para la provisión de parques públicos; El área de actividad residencial constituye el 26.66% del total de la UPZ y estas zonas pueden ser: zonas de uso residencial con presencia limitada de comercio y servicios sin superar el 5% del área normativa, zonas de uso residencial con zonas delimitadas de comercio y servicios y zonas de uso residencial en las que se delimitan zonas de uso exclusivo residencial y zonas de comercio y servicios sin superar el 30% del área del sector normativo; El área de actividad industrial representa el 8.33% del total de la UPZ y se define como aquella en la que se permite la localización de establecimientos dedicados a la producción, elaboración, fabricación, preparación, recuperación, transformación, tratamiento y manipulación de materias primas para la producción de bienes y servicios; El área de actividad múltiple que representa el 18.50%, constituye zonas que por su localización y características adquiridas a través de la formación y consolidación de sus estructuras, presenta una mezcla de usos urbanos y la intensificación de alguno de ellos especialmente los comerciales; por último tenemos áreas de desarrollo que representa el 6.83% de la UPZ, en la que se presentan áreas de desarrollo mixto con tratamiento urbanístico de desarrollo.

El manual de procedimiento para la determinación de zonas homogéneas físicas urbanas y rurales (UAECD, 2005) define tratamientos urbanísticos como el manejo de los sectores con distintas normas sobre los aspectos del comportamiento físico y se establecen a partir de la valoración de los elementos de la estructura urbana y se manifiestan a través de la

adopción de normas específicas, en la UPZ 65 se tiene presencia de 7 diferentes tratamientos y una zona sin clasificación que se distribuye de la siguiente forma:

a. Consolidación básica, con una participación de 42.28% se define consolidación básica a los sectores residenciales con tendencia de desarrollo principalmente de autoconstrucción, donde se pretende consolidar permitiendo la modificación moderada o ampliación de las edificaciones existentes.

b. Mejoramiento integral, con un total de 12.49% de la zona de estudio, y se define como aquel que rige las actuaciones de planeamiento para la regularización de los asentamientos urbanos de origen informal para su integración a la estructura de la ciudad.

c. Consolidación máxima, representa el 12.70% de la UPZ y se define como sectores residenciales donde se pretende mantener el patrón urbanístico existente debido a la calidad y estado de conservación.

d. Desarrollo sin planes parciales, con un total de 2.60% en la UPZ, es aquel que se aplica a predios urbanizables localizados en el suelo urbano.

e. Desarrollo por planes parciales, contiene el 13.31% del área total de la UPZ y similar que el anterior, se aplica a predios urbanizables localizados en suelo urbano o de expansión mediante plan parcial previo al proceso de urbanización.

f. Renovación urbana reactivación, con un total de 59.11% del área de la zona de estudio se aplica a sectores donde se requiere la habilitación y mejoramiento parcial del espacio público con sustitución parcial y paulatina del espacio edificado.

g. Renovación urbana redensificación, representa el 7.56% y busca la transformación de las zonas desarrolladas mediante la cual aumenta su capacidad de albergar nueva población e intensificar usos del suelo mediante la recualificación de los soportes urbanos.

9. Metodología

La metodología utilizada en este experimento pretende describir los pasos y procesos necesarios para diseñar y evaluar una red neuronal para la descripción de características constructivas de una edificación mediante el análisis de la fachada para una UPZ de la ciudad de Bogotá, este proceso se puede implementar a un nivel más amplio con los predios a nivel localidad, ciudad, región; expandiendo su área de aplicación según los datos con los que se disponga.

9.1. Adquisición del conocimiento

Para adquirir la información predial se solicita a la UAECD la información física de los bienes inmuebles de la UPZ 65 (Arborizadora), con interés especial en los datos correspondientes a la fachada de la edificación, obtenida mediante los procesos catastrales de formación y/o actualización catastral, plasmados en la ficha predial con el objetivo de buscar tendencias que permitan hacer una descripción de las características constructivas del predio mediante una red neuronal artificial y un árbol de decisiones.

La red neuronal artificial y el árbol de decisiones son modelos de aprendizaje supervisado, ya que los datos que usan para su entrenamiento y posterior aprendizaje contienen pares de datos que son la entrada de la red y una salida con datos esperados; cuyo objetivo es crear

una función que tenga la capacidad de predecir el valor correspondiente a cada entrada después de haber visto ejemplos, el resultado de esta predicción puede ser un valor numérico o una etiqueta resultantes de un proceso de regresión y de clasificación respectivamente. La figura 8 muestra la metodología mediante la cual se generan modelos en aprendizaje supervisado.

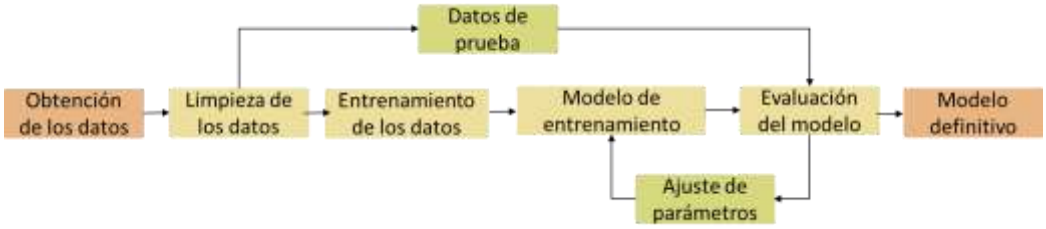


Figura 9. Fases del aprendizaje supervisado. Fuente: Elaboración propia, 2020.

9.2. Red neuronal artificial

Una red neuronal artificial se compone de una capa de entrada, compuesta por neuronas, que recibe los datos, capas de procesamiento, que varían en su número según se disponga y capas de salida, donde se muestran los resultados de la predicción. La ecuación 11 nos describe de forma matemática el proceso de predicción de una red neuronal.

$$y = f(\sum w_i x_i)$$

Donde:

y: valor de predicción

w_i: peso sináptico i

x_i: valor de entrada i

9.2.1. Identificación de las variables de entrada y variables de salida.

Con la previa elección de los parámetros de entrada se comprende cómo será el funcionamiento de la red y como a partir de estas se puede llegar a los valores deseados. En este caso los parámetros de entrada de la red corresponden a los primeros 8 campos de la ficha predial; de la estructura: armazón, muros, cubierta y conservación, de los acabados: fachada, cubrimiento de los muros, pisos y conservación, la novena variable de entrada corresponde a si el predio constituye o no propiedad horizontal, las características anteriormente descritas son variables categóricas ya que se les asigna un valor numérico como resultado de la abstracción de una serie de características por lo que intuimos que la variable de salida será categórica por lo que el tipo de red que manejaremos será de clasificación cuya variable respuesta será una categoría que describe una tipificación de la construcción.

9.2.2. Limpieza de los datos.

La base de datos suministrada por la UAECD contiene información de todas las construcciones de la UPZ, para procesar la información es necesario eliminar aquellas unidades cuya información se encontrara incompleta y con usos diferentes al residencial, después de realizar está limpieza se realiza el preprocesamiento de los datos que nos permitirá adaptar los datos al sistema establecido por la red, por último se realiza la segmentación del total de datos en un grupo para entrenamiento y otro para evaluación.

9.2.3. Entrenamiento de los datos.

Con la red ya establecida se realiza el entrenamiento de los datos que mediante una cantidad de iteraciones genera el aprendizaje esperado de la red que resulta en la generación de los pesos que son utilizados en la siguiente iteración, en esta se genera otra segmentación de la base de datos ingresada en un segmento de validación que es utilizada por la red para evaluar el asertividad del aprendizaje.

9.2.4. Evaluación del modelo

El resultado de las iteraciones generadas durante el entrenamiento entrega parámetros que ayudan a analizar el funcionamiento y aprendizaje de la red, al observar estos parámetros se puede reiterar el modelo ajustándolo como sea pertinente así, cuando se obtengan resultados acertados se asienta el modelo que ya está debidamente entrenado para generar las predicciones.

9.2.5. Resultados y selección del modelo

Para desarrollar el experimento se realizaron diferentes modelos en los que se variaba el tipo de neurona y red con el objetivo de encontrar aquella que tuviera mejor desempeño a la hora de clasificar los datos de entrada. Así es importante reconocer como se ingresaron las entradas a la red, el rendimiento del entrenamiento y su salida.

Las entradas de la red son 9 variables que están descritas a continuación:

1. Estructura

   El capítulo de la estructura está dividido en cuatro variables que varían según el material empleado o el estado de conservación.

a. Armazón: madera, prefabricado, ladrillo o bloque, concreto hasta 3 pisos, concreto cuatro o más pisos

b. Muros: materiales de desecho o esterilla; bahareque, adobe o tapia; madera; concreto prefabricado; bloque o ladrillo.

c. Cubierta: materiales de desecho o telas asfálticas; zinc, teja de barro o Eternit rustico; entrepiso (cubierta provisional); prefabricado.

d. Conservación: malo, regular, bueno o excelente.

2. Acabados principales:

   Este capítulo está dividido en cuatro variables que dependen del material empleado o el estado de conservación.

   a. Fachadas: pobre, sencilla, regular, buena o lujosa.

   b. Cubrimiento de los muros: sin cubrimiento; pañete, papel común o ladrillo prensado; estuco, cerámica o papel fino; medra o piedra ornamental; mármoles u otros lujosos.

   c. Pisos: tierra pisada; cemento o madera burda; baldosa común de cemento o tablón ladrillo; listón machi-embrado; tableta, caucho, acrílico, granito o baldosa fina; parquet, alfombra o retal de mármol (grano pequeño); retal de mármol, mármol u otros lujosos.

   d. Conservación: malo, regular, bueno o excelente.

3. Propiedad horizontal: esta variable menciona si el predio está bajo el régimen de propiedad horizontal o no.

Luego de determinar las variables de entrada, se declara el objetivo de la clasificación en el cual se determinaron 6 categorías que comparten características constructivas similares que se describen a continuación:

a. Tipología 1: Las viviendas que pertenecen a este grupo están construidas sin planeación alguna y generalmente con materiales de desecho tanto en las paredes como en los techos, construido o acondicionado para alojamiento, aunque no reúna las condiciones sanitarias indispensables. Generalmente carecen de algún servicio público básico y/o están ubicadas en zonas marginales no urbanizadas ni urbanizables, siendo de invasión y con escasez de vías de comunicación y transporte urbano. (IGAC, 2010)

b. Tipología 2: Las viviendas que pertenecen a este grupo están construidas con o sin planeación y generalmente con materiales como madera, prefabricado, ladrillo o bloque, cubiertas en zinc o teja de barro y con acabados sencillos, pueden disponer de algunos servicios públicos básicos, están situados por lo general en la periferia y disponen de algunas vías públicas sin pavimentar y escasos medios de transporte. (IGAC ,2010)

c. Tipología 3: Las viviendas que pertenecen a este grupo generalmente están constituidas por edificaciones técnicamente planeadas, con fachadas calificadas como regular, con muros de ladrillo o bloque, cubiertas de placa de concreto o teja en asbesto cemento. Generalmente cuentan con servicios públicos básicos, poseen vías de comunicación pavimentadas y medio de transporte urbano. (IGAC ,2010)

d. Tipología 4: Las viviendas que pertenecen a este grupo generalmente están

constituidas por edificaciones técnicamente planeadas con fachada calificada de regular a buena, estructura en concreto, muros de ladrillo o bloque revocados, estucados y pintados, cubiertas de placa de concreto o teja en asbesto cemento; baños y cocinas calificados como medianos enchapados. Generalmente están conectados con todos los servicios públicos cuentan con vías pavimentadas y facilidad de transporte. (IGAC ,2010)

e. Tipología 5: Las viviendas que pertenecen a este grupo generalmente están constituidas por edificaciones totalmente terminadas y técnicamente planeadas con fachada calificada como buena, así como sus acabados, estructura en concreto. Generalmente cuentan con muros de ladrillo o bloque, cubiertas de placa de concreto o con materiales de buena calidad; baños y cocinas calificados como buenos y tamaño entre mediano y grande. Cuentan con todos los servicios públicos básicos y complementarios, con grandes zonas verdes, sociales y de recreación, parqueaderos y buenas vías de acceso. (IGAC ,2010)

f. Tipología 6: Las viviendas que pertenecen a este grupo generalmente están constituidas por edificaciones con grandes áreas, amplios jardines o apartamentos lujosos, paredes de bloque o ladrillo, cubiertas de placa de concreto o con materiales de excelente calidad y cuentan con los servicios públicos básicos y complementarios. Estas viviendas están ubicadas en sectores exclusivos y disfrutan de amplias zonas verdes, sociales y de recreación, parqueaderos y excelentes vías de acceso. (IGAC ,2010)

Una vez definida las entradas y las salidas, se pre-procesan para que al momento de entrar cada una de estas categorías se definan en una variable dummie con solo dos valores

posibles que determinan si el individuo pertenece o no a la categoría determinada. Se hace la segmentación de los datos en entrenamiento y evaluación para el desarrollo del modelo y se hace una transformación de los datos en una forma vectorial que permitan mayor facilidad de su procesamiento.

El lenguaje de programación elegido para realizar este experimento fue Python ya que dispone de múltiples herramientas y librerías orientadas a la inteligencia artificial que facilitan el desarrollo de los algoritmos. Una vez elegido el lenguaje escogemos la interfaz para desarrollar el algoritmo que es Google Colaboratory que es un entorno gratuito alojado totalmente en la nube que no requiere de configuraciones adicionales.

Se desarrollaron múltiples modelos que evaluamos con el ajuste que apreciamos en la figura 10 y la perdida en la figura 11, mediante los cuales se seleccionara el mejor modelo.



Figura 10. Ajuste de las redes neuronales. Fuente: Elaboración propia, 2020.



Figura 11. Pérdida de las redes neuronales. Fuente: Elaboración propia, 2020.

La red seleccionada fue la red número 20 que tuvo los mejores parámetros encontrados en los diversos experimentos con un rendimiento de la red optimo contra el tiempo de ejecución que podemos evaluar en el siguiente capítulo.

N° de modelo	Pérdida	Ajuste
1	0,6272	72,0145%
2	1,1057	72,6258%
3	1,1070	72,3455%
4	0,6250	72,8125%
5	1,1867	69,7015%
6	0,9182	72,1578%
7	0,7271	71,6225%
8	0,7125	71,8166%
9	0,6778	71,4877%
10	1,1375	71,8215%
11	1,1511	71,5621%
12	1,1853	72,0423%
13	0,8175	71,7308%
14	0,2441	85,2315%
15	0,2407	83,2721%
16	0,2448	84,3690%
17	0,2416	87,1621%
18	0,2395	86,7522%
19	0,2513	86,7912%
20	0,2404	89,5115%

Figura 12. Parámetros de las redes neuronales. Fuente: Elaboración propia, 2020.

9.3. Árbol de decisiones

Para el árbol de decisiones manejamos la misma estructura que se utilizó dentro de la red neuronal definiendo los parámetros de entrada y la misma respuesta objetivo con la diferencia de la forma en que estos datos entran en el algoritmo y su procesamiento dentro de él. Al ingresar los valores los datos conservan su tipo de dato original que son llevados por el algoritmo cuyo resultado es la categoría a la que pertenece el elemento.

El funcionamiento de un árbol de decisiones se da al pasar los datos de nodo a nodo según las condiciones establecidas por el modelo. Así, existen dos tipos de nodos: de decisión, que tienen una condición y nodos debajo de él, y de predicción, que no tiene una condición ni nodos debajo de él. Los nodos tienen la siguiente información: condición, que determina la división de los datos; criterion, que mide la impureza del nodo; simples, que muestra el número de muestras que satisfacen la condición; value, que dice cuántas muestras de cada clase llegan a ese nodo y class, presente en los nodos de predicción que le asigna la clase a las muestras que lleguen a ese nodo.

La principal ventaja de los árboles de decisiones es la representación gráfica que permite una fácil interpretación. Los nodos de decisión están claramente identificados ya que tienen una condición en su interior y de él se desprenden más nodos mientras que los de predicción no tienen nodos debajo de él y tienen un valor de clase asignado para los elementos que lleguen a él, cada nodo tiene asignado un color que representa una clase, entre más intenso sea el color más seguro está el árbol de la decisión por lo cual los tonos más blancos y pálidos representan que hay falta de certeza en ella.

10. Análisis de Resultados

Los resultados obtenidos en cada algoritmo serán analizados por separado, pero como ambas tienen el mismo objetivo podremos comparar directamente los resultados obtenidos mediante la metodología tradicional, el algoritmo de una red neuronal artificial y el algoritmo de árbol de decisiones para describir las características físicas de la construcción condensadas en la tipificación.

La forma en la que ingresamos los datos a la red neuronal es de una forma transformada ya que al ser una clasificación dada por 6 valores objetivos se optó por crear una red neuronal cuyo valor respuesta es un vector de probabilidades con 6 valores en el que el componente que tenga mayor valor corresponderá a aquella categoría, a la cual, por las características de entrada, y el aprendizaje realizado el individuo pertenezca.

Para llegar a ese valor la red neuronal artificial tuvo que pasar un proceso de aprendizaje en el cual la red neuronal va cambiando sus parámetros de acuerdo a los resultados de la iteración anterior. Para este modelo se utilizó una red neuronal con la capa de entrada, 3 capas ocultas y la capa de salida, antes de cada capa oculta se tiene una capa de dropout, que apaga cierto porcentaje de neuronas que permite a la red aprender diferentes caminos para llegar a la variable respuesta. Los parámetros obtenidos en el modelo elegido son los siguientes:

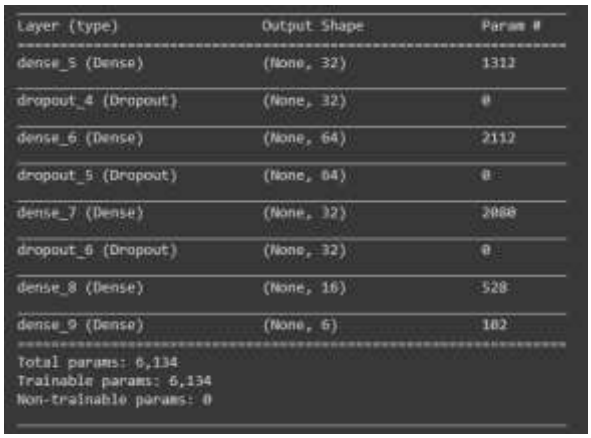


Figura 13. Parámetros de la red neuronal seleccionada. Fuente: Elaboración propia, 2020.

Como podemos observar cada capa densa tiene como salida una cantidad definida de parámetros correspondiente al número de neuronas que tiene por capa, al final observamos que la red posee un total de 6134 parámetros entrenados cuyo valor nos ayuda a realizar la clasificación. El grado de aprendizaje seleccionado es de 0.0005 por lo que nos aseguramos que al realizar la propagación hacía atrás estará siendo evaluada cada 0.0005 momentos con el que el optimizador generará los pesos de la red mejorándolos cada iteración que con ayuda de la entropía cruzada ajusta los pesos. La figura 14 nos muestra como la perdida de la red va descendiendo conforme el número de iteraciones aumentan tanto en el entrenamiento como en la validación. Mientras que la figura 15 nos muestra como el ajuste del modelo mejora.

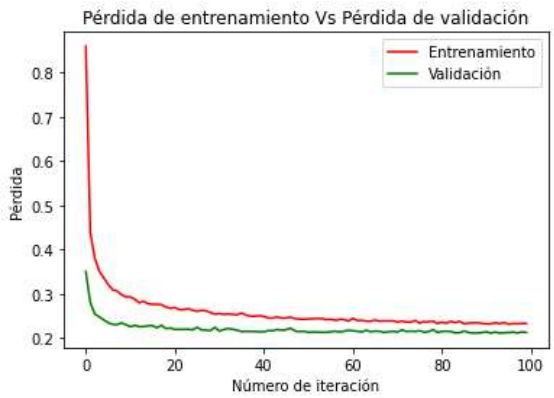


Figura 14. Pérdida de entrenamiento Vs Pérdida de validación. Fuente: Elaboración propia, 2020.

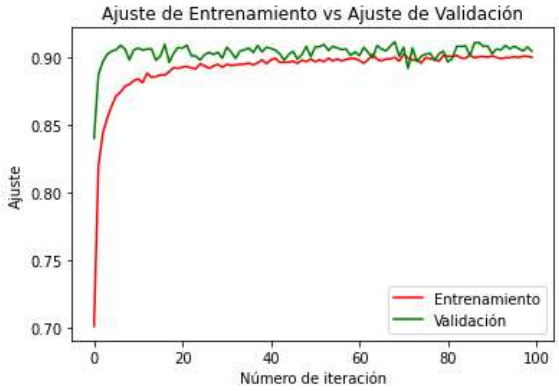


Figura 15. Ajuste de entrenamiento Vs ajuste de validación. Fuente: Elaboración propia, 2020.

Como podemos observar los valores de la validación se encuentran sobre los valores de entrenamiento por lo que podemos observar que hay sobreajuste, pero al ser está diferencia mínima podemos pasarla por alto. Al realizar la evaluación del modelo con los datos de entrenamiento obtenemos como resultado el valor de perdida y de ajuste para la evaluación que permite conocer con certeza el desempeño del algoritmo tenemos los siguientes valores

Pérdida de la evaluación	Ajuste de la evaluación
0,2163	90,61%

Figura 16. Perdida y ajuste de la red neuronal en la evaluación. Fuente: Elaboración propia, 2020.

Con el resultado anterior podemos inferir que de cada 10 predicciones que se hagan con la red va a acertar en 9 de ellas y en la restante hace la predicción con un grado menor de certeza, con la perdida, que está dada para funciones cuyo valor sea probabilístico nos muestra cuanto le cuesta a la red identificar a que clase pertenece un elemento. Para ejemplificar este modelo tomamos un elemento de la muestra con los parámetros descritos en la Figura 17.

Estructura	
Armazón	4
Muros	6
Cubierta	9
Conservación	0
Acabados principales	
Fachada	4
Cubrimiento de muros	2
Pisos	6
Conservación	2
PH	Sí
Tipología	4

Figura 17. Datos de entrada para la predicción de la red neuronal. Fuente: Elaboración propia, 2020.

Con estos valores realizamos una predicción dentro de la red neuronal para predecir la categoría a la cual este predio pertenece obteniendo como respuesta un vector de probabilidad en el que aquel campo con mayor valor será la clase a la que pertenece el elemento.

```
array([[9.3219878e-21, 3.3110577e-09, 5.9763195e-05, 9.9667054e-01,
        3.2697006e-03, 2.5088398e-10]], dtype=float32)
```

Figura 18. Vector de probabilidad de la predicción. Fuente: Elaboración propia, 2020.

Como observamos aquel campo de mayor probabilidad en el vector es número 4 que corresponde a la tipología seleccionada mediante el modelo tradicional cuyo valor podemos

observar que es bajo, esto se debe a que principalmente las variables de entrada se quedan cortas al clasificar los elementos, esto se debe a que los campos de la ficha predial no describen en su totalidad la construcción pero este valor puede mejorarse al establecer nuevas variables correspondientes a la fachada así como una reasignación de los puntajes de cada material dentro de la ficha ya que podemos observar que si encontramos un predio cuyo puntaje sea el máximo en todas las variables, este nunca alcanzara un puntaje total de 100.

Con el árbol de decisión que está totalmente descrito en el documento anexo, tenemos un análisis más visual que permite una mayor facilidad de interpretación de su desempeño, en las primeras líneas de código podemos observar que las etiquetas de destino son descritas desde un principio en la que se determinan que serán 6 posibles categorías a las que los elementos de la muestra tendrán que ser asignados, siguiendo su diseño, observamos que también se define las entradas como categorías mediante las cuales el árbol realizara la debida clasificación siguiendo la condición que dispongan los nodos. En la figura 19 se muestra un nodo del árbol de decisiones entrenado para ejemplificar su estructura

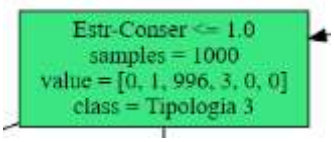


Figura 19. Nodo del árbol de decisiones. Fuente: Elaboración propia, 2020.

Al entrenar el algoritmo realizamos de nuevo un análisis de ajuste con los datos de evaluación y con los datos de entrenamiento con lo que obtenemos lo siguiente

Ajuste de entrenamiento	Ajuste de la evaluación
91,3058%	90,7335%

Figura 20. Ajuste de entrenamiento vs Ajuste de evaluación. Fuente: Elaboración propia, 2020.

Igual a como se observó en el algoritmo de la red neuronal, observamos que el ajuste de la evaluación es inferior al del entrenamiento, por lo cual se infiere que en el modelo se encuentra un sobreajuste, pero al ser una diferencia mínima se considera que el modelo está dentro de los parámetros normales que nos permite realizar predicciones dentro de ella, por lo cual realizaremos una predicción con los datos descritos a continuación:

Estructura	
Armazón	4
Muros	6
Cubierta	9
Conservación	0
Acabados principales	
Fachada	4
Cubrimiento de muros	2
Pisos	6
Conservación	2
PH	Sí
Tipología	4

Figura 21. Datos de entrada para la predicción árbol de decisiones. Fuente: Elaboración propia, 2020.

El resultado de la predicción al ingresar estos valores dentro del árbol de decisión es el siguiente:

```
array(['Tipologia 4'], dtype=object)
```

Figura 22. Predicción árbol de decisiones. Fuente: Elaboración propia, 2020.

Este tipo de algoritmo también permite reconocer con una mayor facilidad cuales de sus variables tienen mayor peso o importancia en el momento de hacer la predicción, de este

modo podemos determinar que la variable que tiene mayor peso en la clasificación de las tipologías constructivas es la Fachada en el capítulo de acabados principales. En la figura 23 se describe la importancia de cada variable en el algoritmo.

Variables	importancía
Acab-Fachad	0.545
Estr-Cubier	0.139
Acab-Conserv	0.094
Acab-Pisos	0.088
Estr-Conser	0.074
Acab-CubMur	0.040
Estr-Muros	0.016
PH	0.003
Estr-Armazon	0.001

Figura 23. Importancia de las variables árbol de decisiones. Fuente: Elaboración propia, 2020.

Como observamos en la red neuronal artificial, el algoritmo del árbol de decisiones también optimiza los valores que recibe promoviendo una clasificación de las características constructivas de los predios, pero se sigue teniendo dificultades al procesar la información disponible por lo que es recomendable reevaluar la metodología de la ficha predial ya que sus variables y características parecen estar incompletas.

11. Evaluación y Cumplimiento de Objetivos

Leyendo los resultados anteriormente descritos nos muestran que los objetivos fueron exitosamente alcanzados dentro del experimento. Al aplicar el algoritmo de la red neuronal artificial dentro de los datos de la UPZ 65 Arborizadora se pudo comprobar que la herramienta funciona de una manera eficaz para los procesos masivos como la formación y actualización catastral con el énfasis de las características constructivas del predio.

Durante el experimento se realizaron múltiples modelos que fueron comparados y seleccionados para llegar a aquel que tuvo mejor desempeño que se entrega como el modelo seleccionado para ser parte del proceso masivo, además se desarrolló un modelo con un algoritmo diferente, el árbol de decisiones, que nos permitió realizar un análisis desde diferentes perspectivas al problema planteado al inicio del experimento que es la determinación eficaz y rápida de las características constructivas de un predio.

12. Conclusiones y Recomendaciones

La inteligencia artificial ha evolucionado de una manera acelerada con la tecnologías disponible que permite a sus herramientas ser implementadas dentro de múltiples áreas, entre las cuales, el catastro puede aprovecharlas para optimizar sus procesos que dentro del marco del catastro multipropósito generan oportunidades a los gestores y operadores catastrales para mantener debidamente actualizada la información predial sin la necesidad de realizar un despliegue grande de recursos físicos, humanos y económicos.

La información disponible en la ficha predial es de gran ayuda para determinar características constructivas para las edificaciones pero debe ser actualizada con sus variables y características que se han quedado atrasadas ante la evolución de la forma en la que se construye en el país, La tipificación es una manera sencilla de abstraer estas variadas dinámicas constructivas pero debe ser analizada y construida de una forma regional ya que la manera de construir en la costa caribe no es la misma manera que se construye en la zona andina o los llanos orientales por lo que se recomienda construir una base de datos con esta información para implementar estas herramientas en el catastro nacional.

La calidad y precisión que se obtiene al realizar predicciones por medio de los algoritmos de red neuronal artificial o árbol de decisiones son altas y evita errores que puede cometer el personal humano en el proceso de reconocimiento predial pero para obtener modelos rígidos y sustentados es necesario contar una gran cantidad de datos que pueda ser entrenados y para eso estos datos deben estar bien clasificados y aunque los modelos optimizan la información recibida se espera que al tener una base de datos solida la calidad y precisión de las predicciones será mayor.

13. Referencias

A Systematic Review of Deep Learning Methods Applied to Ocular Images. Ciencia e Ingeniería Neogranadina, 30(1), 9–26. https://doi.org/10.18359/rcin.4242, & González Osorio, F. A. (2019). A Systematic Review of Deep Learning Methods Applied to Ocular Images. *Ciencia e Ingeniería Neogranadina*, *30*(1), 9–26.

Árbol de decisión en Machine Learning (Parte 1) - sitiobigdata.com. (n.d.). Retrieved December 2, 2020, from https://sitiobigdata.com/2019/12/14/arbol-de-decision-en-machine-learning-parte-1/#

Dahan, H., Cohen, S., Rokach, L., & Maimon, O. (2014). Proactive Data Mining Using Decision Trees (pp. 21–33).

Adelaida, M., & Echeverri, E. (n.d.). *Aplicación de Árboles de Decisión para la valoración de compañías de minería aluvial*.

Álvarez Cedillo, J. A., Álvarez Sánchez, T., Sandoval Gómez, R. J., & Aguilar Fernández, M. (2019). La exploración en el desarrollo del aprendizaje profundo / The Exploration in the Develop Deep Learning. *RIDE Revista Iberoamericana Para La Investigación y*

El Desarrollo Educativo, *9*(18), 833–844.

Akinosho, T. D., Oyedele, L. O., Bilal, M., Ajayi, A. O., Delgado, M. D., Akinade, O. O., & Ahmed, A. A. (2020). Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering*, *32*, 101827.

Meyer, J. A., & Guillot, A. (2005). Editorial: Special issue on artificial rodents. *Adaptive Behavior*, *13*(2), 85–86.

Santos, D., Dallos, L., & Gaona-García, P. A. (2020). Algoritmos de rastreo de movimiento utilizando técnicas de inteligencia artificial y machine learning. *Información Tecnológica*, *31*(3), 23–38.

Brei, V. A., Nicolao, L., Pasdiora, M. A., & Azambuja, R. C. (2020). An integrative model to predict product replacement using deep learning on longitudinal data. In *BAR - Brazilian Administration Review* (Vol. 17, Issue 2).

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object Detection with Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212–3232.

J, M. (2020). Artificial Intelligence and Big Data as solutions to COVID-19 Intel · ligència Artificial i Big Data per fer front a la COVID-19. *Revista de Bioética y Derecho Perspectivas Bioéticas*, *c*, 315–331.

Avila-Tomás, J. F., Mayer-Pujadas, M. A., & Quesada-Varela, V. J. (2020). Artificial

intelligence and its applications in medicine I: introductory background to AI and robotics. *Atencion Primaria*, *xx*.

Rusell, S., & Norvig, P. (2004). Inteligencia Artificial. Un Enfoque Moderno. In *Inteligencia Artificial*. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Inteligencia+Artificial:+un+enfoque+moderno#0

Filho, E. M. de S., Fernandes, F. de A., Soares, C. L. de A., Seixas, F. L., Dos Santos, A. A. S. M. D., Gismondi, R. A., Mesquita, E. T., & Mesquita, C. T. (2020). Artificial intelligence in cardiology: Concepts, tools and challenges-"the horse is the one who runs, you must be the jockey." *Arquivos Brasileiros de Cardiologia*, *114*(4), 718–725.

Johnsirani Venkatesan, N., Nam, C. S., & Shin, D. R. (2019). Deep Learning Frameworks on Apache Spark: A Review. *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, *36*(2), 164–177.

Athanasiadou, E., Geradts, Z., & Van Eijk, E. (2018). Camera recognition with deep learning. *Forensic Sciences Research*, *3*(3), 210–218.

Ramos, D. L., & García, L. A. (2019). Aprendizaje profundo para la extracción de aspectos en opiniones textuales. *Revista Cubana de Ciencias Informáticas*, *13*(2), 105–145.

Función de activación - Redes neuronales - Diego Calvo. (n.d.). Retrieved December 3,

2020, from https://www.diegocalvo.es/funcion-de-activacion-redes-neuronales/

Redes Neuronales. (n.d.). Retrieved December 3, 2020, from https://ml4a.github.io/ml4a/es/neural_networks/

A Quick Introduction to Artificial Neural Networks (Part 2). (n.d.). Retrieved December 6, 2020, from http://krisbolton.com/a-quick-introduction-to-artificial-neural-networks-part-2

Coding Deep Learning for Beginners — Linear Regression (Part 2): Cost Function | by Kamil Krzyk | Towards Data Science. (n.d.). Retrieved December 6, 2020, from https://towardsdatascience.com/coding-deep-learning-for-beginners-linear-regression-part-2-cost-function-49545303d29f

Todo lo que Necesitas Saber sobre el Descenso del Gradiente Aplicado a Redes Neuronales | by Jaime Durán | MetaDatos | Medium. (n.d.). Retrieved December 6, 2020, from https://medium.com/metadatos/todo-lo-que-necesitas-saber-sobre-el-descenso-del-gradiente-aplicado-a-redes-neuronales-19bdbb706a78

Optimizadores en redes neuronales profundas: un enfoque práctico | by Luis Velasco | Medium. (n.d.). Retrieved December 6, 2020, from https://medium.com/@velascoluis/optimizadores-en-redes-neuronales-profundas-un-enfoque-práctico-819b39a3eb5

Understand the Impact of Learning Rate on Neural Network Performance. (n.d.). Retrieved December 6, 2020, from https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/

Metrics. (n.d.). Retrieved December 7, 2020, from https://keras.io/api/metrics/

Conociendo mi localidad. (n.d.). Retrieved December 13, 2020, from http://www.ciudadbolivar.gov.co/mi-localidad/conociendo-mi-localidad/historia

Lopez-Garcia, T. B., Coronado-Mendoza, A., & Domínguez-Navarro, J. A. (2020). Artificial neural networks in microgrids: A review. *Engineering Applications of Artificial Intelligence*, *95*(June 2019), 103894.

Matich, D. J. (2001). Redes Neuronales: Conceptos Básicos y Aplicaciones. *Historia*, 55. ftp://decsai.ugr.es/pub/usuarios/castro/Material-Redes-Neuronales/Libros/matich-redesneuronales.pdf

Montaño, J. J. (2002). Redes Neuronales Artificiales aplicadas al Análisis de Datos. *Network*, 275.

Wang, L., Wang, C., Li, Y., & Wang, R. (2021). Explaining the Behavior of Neuron Activations in Deep Neural Networks. *Ad Hoc Networks*, *111*, 102346.

Básica, E., Red, D. E. U. N. A., Tradicional, C., Neuronal, Y. C., Fundamentos, T., Redes, D. E. L. A. S., Artificiales, N., Red, T., Organizing, S., & Red, M. A. P. Y. (n.d.). *REDES NEURONALES ARTIFICIALES- Xabier Bosogain Olabe.pdf*. http://cvb.ehu.es/open_course_ware/castellano/tecnicas/redes_neuro/contenidos/pdf/libro-del-curso.pdf

Astray, G., Castillo, J. X., Ferreiro-Lage, J. A., Gálvez, J. F., & Mejuto, J. C. (2010). Artificial neural networks: a promising tool to evaluate the authenticity of wine Redes neuronales: Una herramienta prometedora para evaluar la autenticidad del vino. *CYTA*

- *Journal of Food*, *8*(1), 79–86.

DNP. (2015). *Evaluación de la capacidad institucional y los procesos realizados para la entrega de los productos y resultados asociados a la interrelación entre el catastro y el registro*. *iii*, 9. www.dnp.gov.co

Peña, W. E. L., & Minim, L. A. (2001). Aplicación De Redes Neuronales Artificiales En La Modelización Del Tratamiento Térmico De Alimentos Application of Neural Networks in the Modelling of the Thermal Treatment of Food Aplicación De Redes Neuronais Artificiais Na Modelización Do Tratamento Té. *Ciencia y Tecnologia Alimentaria*, *3*(2), 81–88.

IGAC, T. D. E. (2010). *Elaboración del estudio de zonas homogéneas físicas y geoeconómicas.*

Congreso de Colombia. (2015). Ley 1753 de 2015. *Diario Oficial*, *Iso 19152*, 1–283.

Congreso de la República de Colombia. (1983). Ley 14 de 1983. *Control*, *Julio 6*, 6–8.

Cogua, M., & Vargas, F. (2012). *Dinámica de la construcción por usos, localidad ciudad Bolívar*. https://www.catastrobogota.gov.co/sites/default/files/7_0.pdf

Departamento Administrativo de la Presidencia de la República, Departamento Nacional de Planeación, Departamento Administrativo Nacional de Estadística, Instituto Geográfico Agustín Codazzi, Superintendencia de Notariado y Registro, & Agencia Nacional de Tierras. (2016). Política Para La Adopción E Implementación De Un Catastro Multipropósito Rural-Urbano CONPES 3859. *Documento CONPES 3859*, 86. https://colaboracion.dnp.gov.co/CDT/Conpes/Económicos/3859.pdf

Congreso de la República de Colombia. (2011). *Resolución 070 de 2011*.

Presidencia de la República de Colombia. (2018). *Ley 1955 Plan Nacional de Desarrollo. "Pacto por Colombia."*

C. L. V, Reglamentario, D., Administrativo, S., Reglamentario, D., presidente, E., Nacional, P., Final, A., Estable, P., Espec, D., Agust, F., Reglamentario, D., Administrativo, S., Espec, D., & Agust, F. (2020). *Decreto 148 de 2020*. *51*, 1–10.

Presidente de la Republica. (1983). *Decreto 3496 de 1983*. 10.

Departamento Nacional de Planeación. (2019). Consejo Nacional de Politica Economica y Social - Republica de Colombia [CONPES 3958], ESTRATEGIA PARA LA IMPLEMENTACIÓN DE LA POLÍTICA PÚBLICA DE CATASTRO MULTIPROPÓSITO. *Conpes*, 79. https://colaboracion.dnp.gov.co/CDT/Conpes/Económicos/3919.pdf

Dong, Q., Chen, Y., & Zhou, S. (2011). A machine learning-based method for protein global model quality assessment. *International Journal of General Systems*, *40*(4), 417–425. https://doi.org/10.1080/03081079.2010.544869

Edition, T. (2021). *Book Review Python Machine Learning : Machine Learning and Deep Learning With Python ,*. *11*(1), 67–70.

Eliseo, D. P., & Fuentes, G. (2019). *Aplicación de algoritmos de Machine Learning para la predicción del beneficio por cliente a partir de métricas de Google Analytics*.

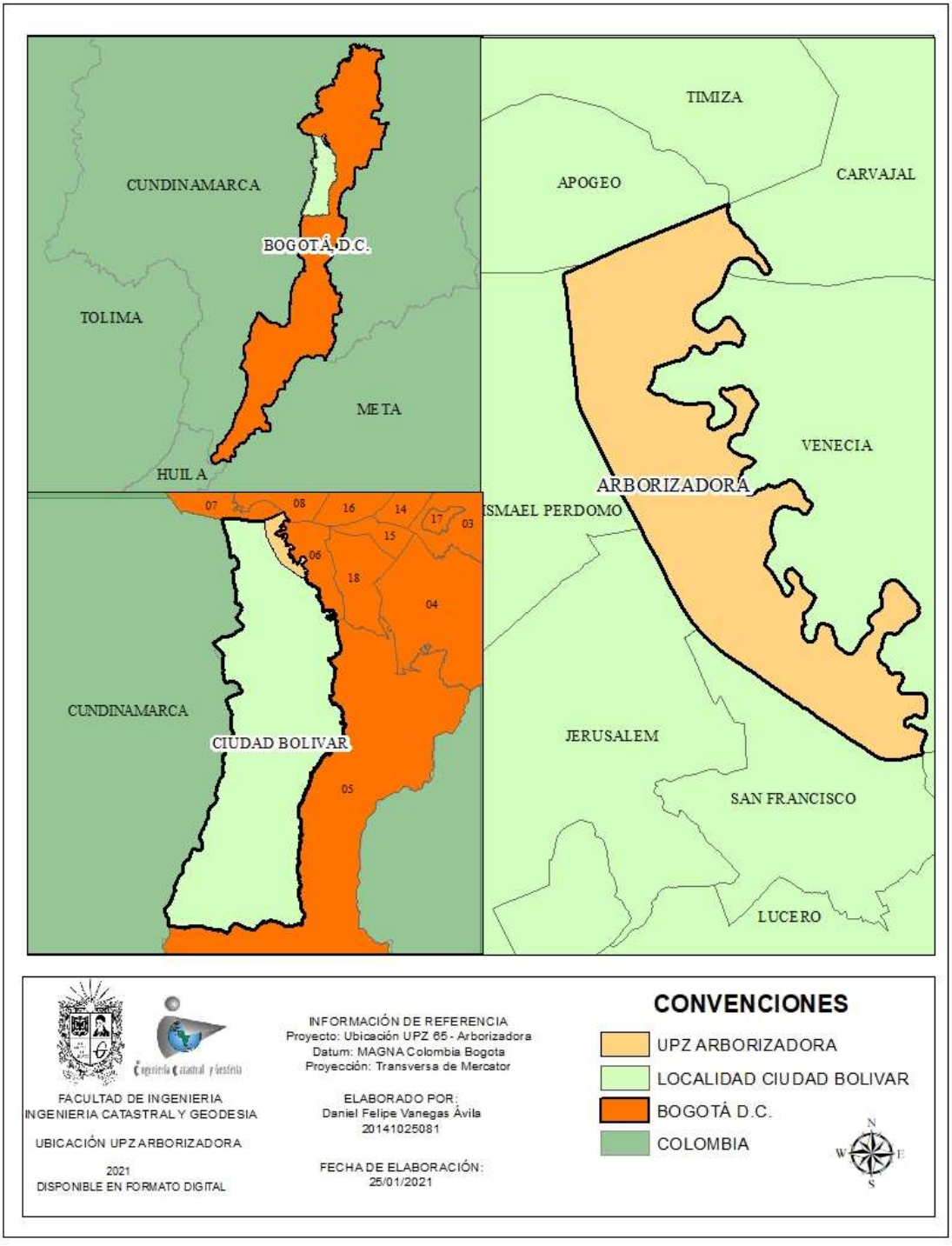
Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (2007). Python for Scientific Computing Python Overview. *Computing in Science and Engineering*, 10–20.

Grado, T. D. E. F. I. N. D. E., Boyarshinov, V., Sociedad Argentina de Informática e Investigación Operativa (SADIO), Basyarudin, Silvestre, L. J., de las Heras Mínguez, G., Ramiro, A., Alvarado, V., Javier Merchán Macías, F., Management Solutions, Tfg, T. D. E. L., Autor, T., Salam, E., Data, S. J., Pentakalos, O., Gago Utrera, R., Urcuqui, C., Osorio, J., Navarro, A., … González, F. A. (2018). Uso de algoritmos automático aplicado a bases de datos genéticos. *Высшей Нервной Деятельности*, *10*(2), 44. http://openaccess.uoc.edu/webapps/o2/bitstream/10609/65426/6/rgagoTFM0617memoria.pdf%0Ahttps://www.managementsolutions.com/sites/default/files/publicaciones/esp/machine-learning.pdf%0Ahttp://sedici.unlp.edu.ar/handle/10915/41722%0Ahttps://books.google.ca/

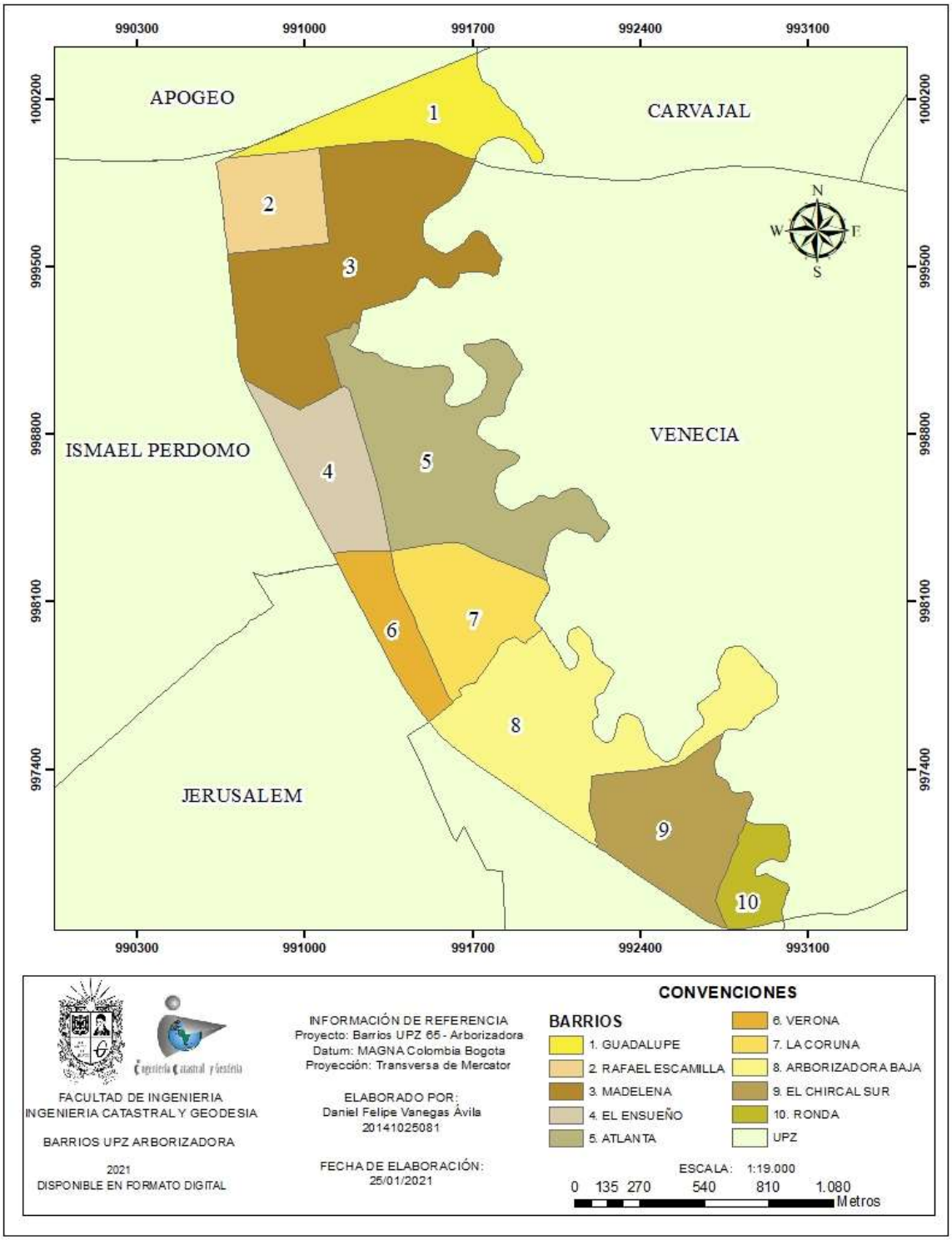
Li, G. Z. (2011). Special issue on massive data processing by using machine learning. *International Journal of General Systems*, *40*(4), 351–354. https://doi.org/10.1080/03081079.2010.530025

Weiss, E. (1974). Demand for scientists and engineers [2]. *Science*, *184*(4141), 1028. https://doi.org/10.1126/science.184.4141.1028-a

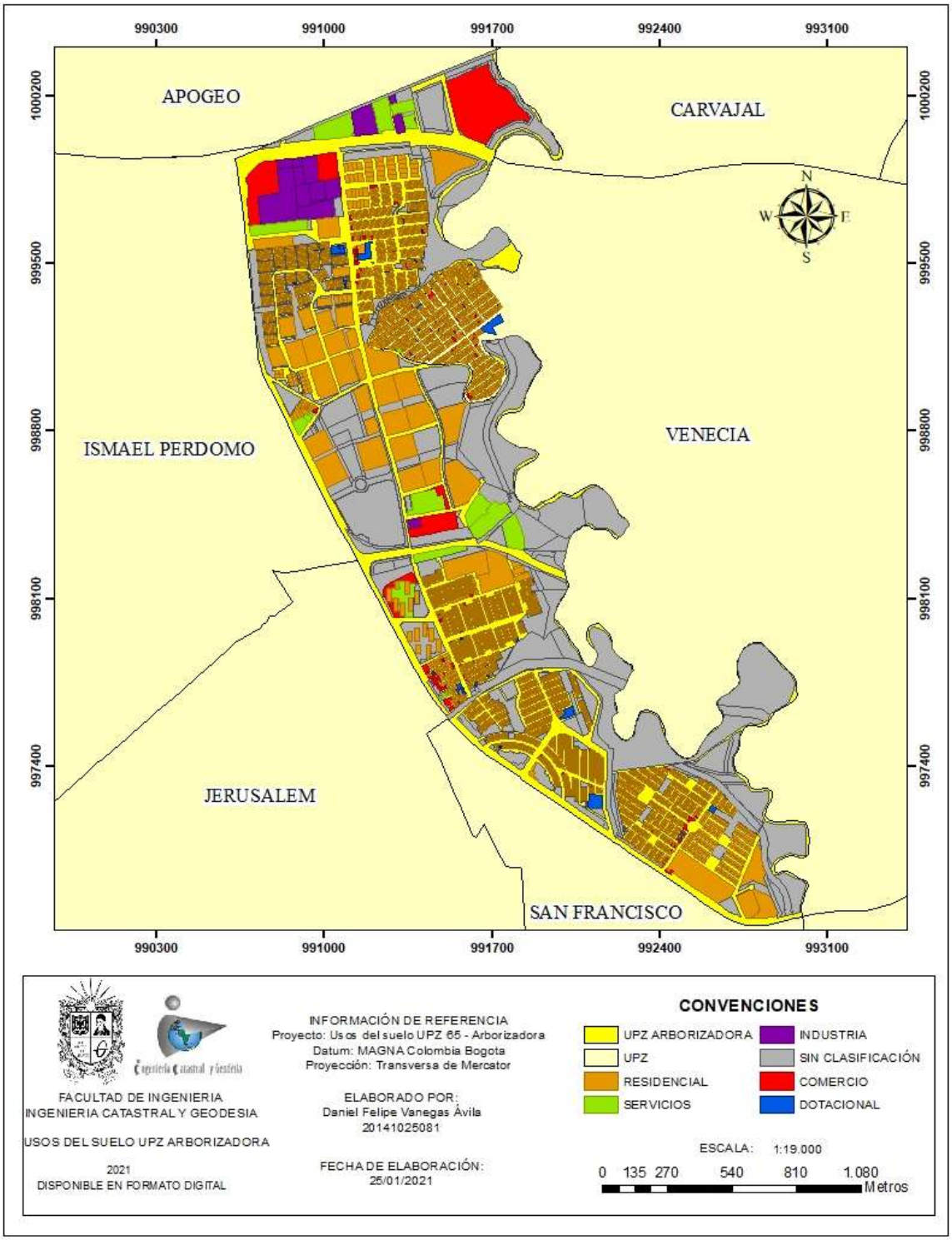
14. Anexos



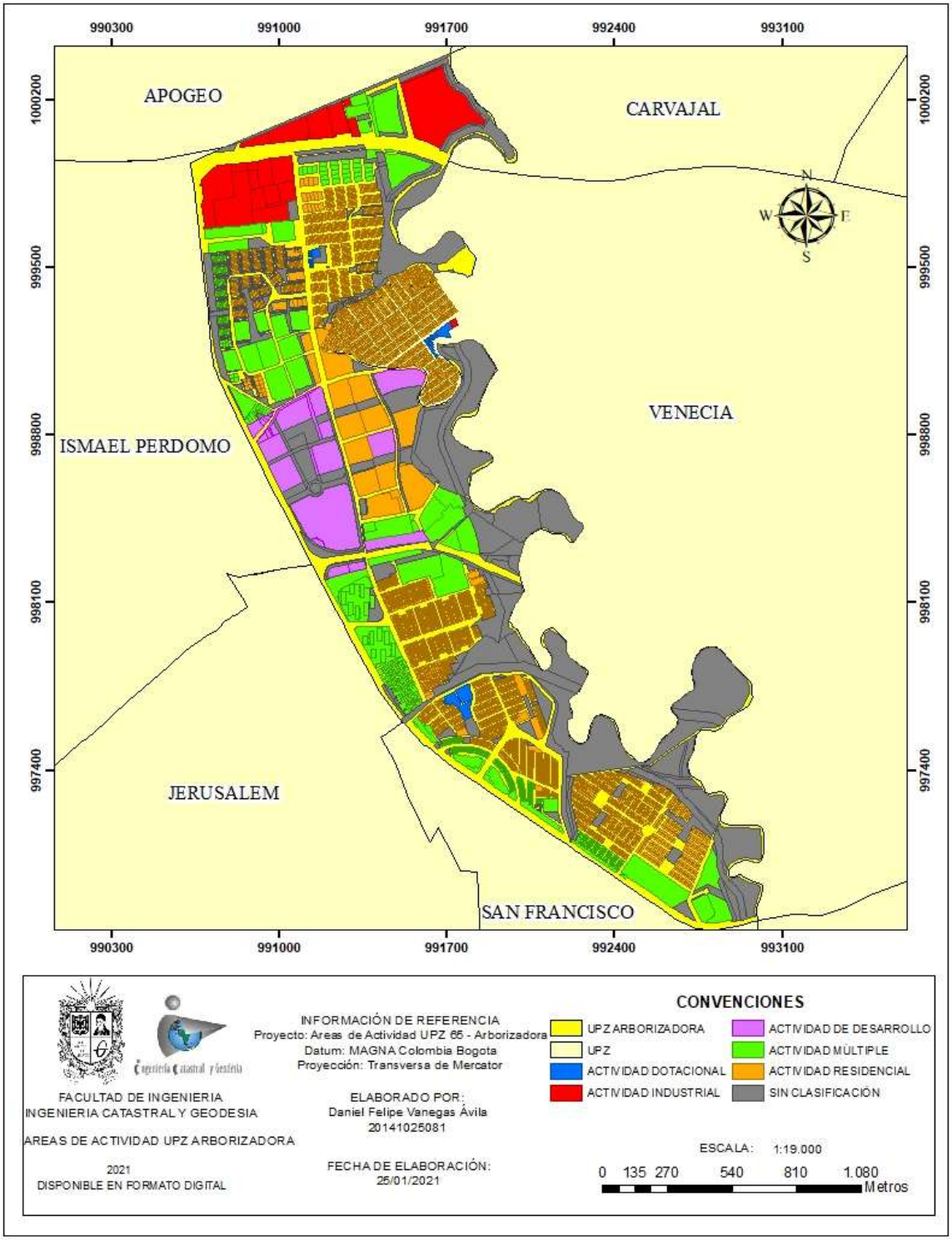
Anexo 1. Ubicación UPZ 65-Arborizadora. Fuente: Elaboración propia, 2021.



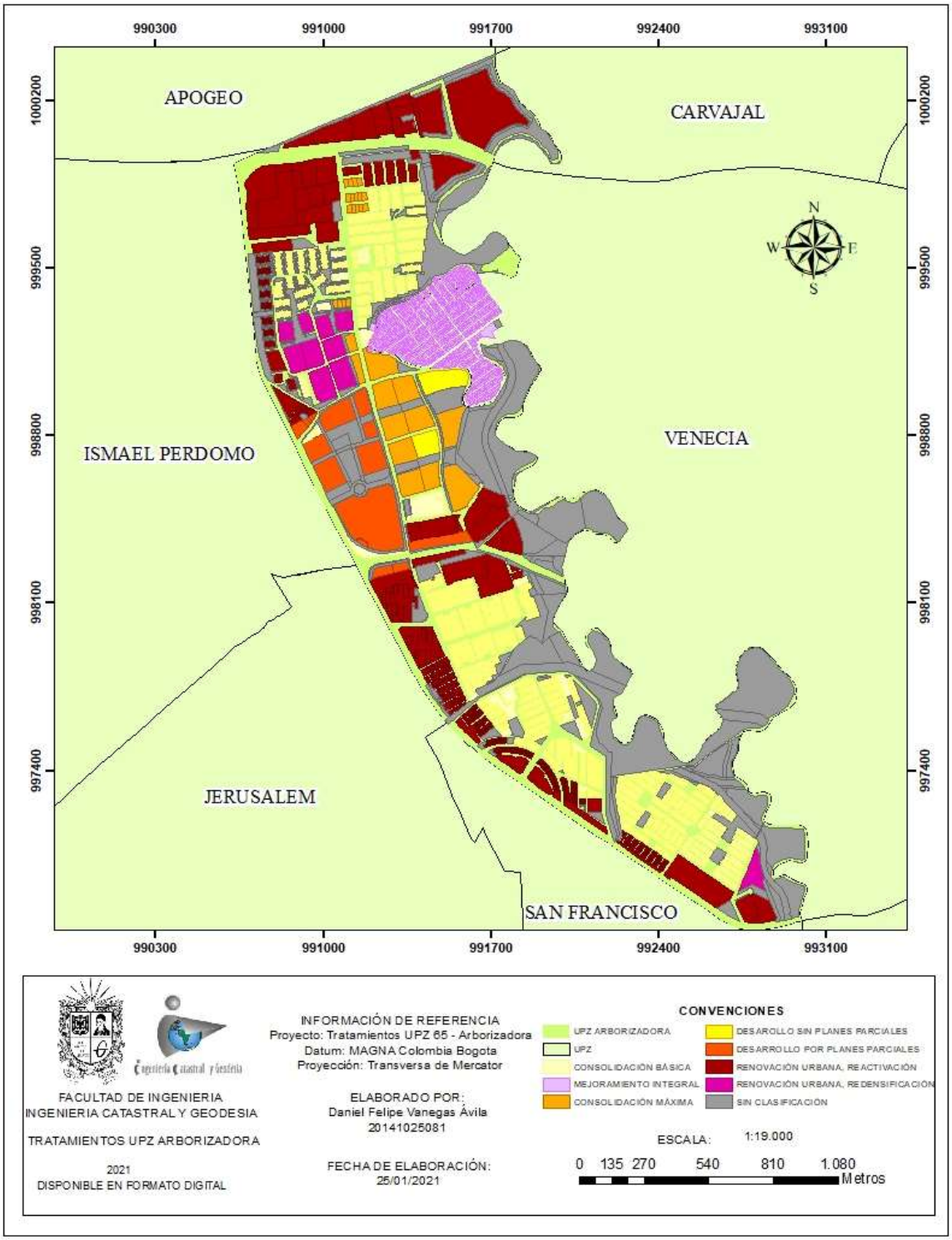
Anexo 2. Barrios UPZ 65 Arborizadora. Fuente: Elaboración propia, 2021.



Anexo 3. Usos de suelo UPZ 65 Arborizadora. Fuente: Elaboración propia, 2021.



Anexo 4. Áreas de actividad UPZ 65-Arborizadora. Fuente: Elaboración propia, 2021.



Anexo 5. Tratamientos urbanísticos UPZ 65-Arborizadora. Fuente: Elaboración propia, 2021.