



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Estado de México

Inteligencia Artificial Avanzada para la Ciencia de Datos I

Reto

TC3006C

Grupo: 101

Profesores

Víctor Adrián Sosa Hernández

Elisabetta Crescio

Julio Guillermo Arriaga Blumenkron

Jorge Adolfo Ramirez Uresti

David Higuera Rosales

Presenta

David Damián Galán - A01752785

Fecha de entrega: 16 de septiembre de 2023

Introducción

A diferencia de los algoritmos convencionales que suelen utilizarse para la resolución de problemas, la inteligencia artificial permite encontrar aproximaciones bastante útiles para aquellos problemas que antes se consideraban “inresolubles”, como lo es la clasificación de imágenes, por ejemplo. En este proyecto, se pretende utilizar modelos de inteligencia artificial para predecir el precio de venta de una casa dadas diversas características como el tamaño, número de habitaciones, localización, etcétera.

Este proyecto representa una gran oportunidad de aprendizaje sobre técnicas de machine learning y también para demostrar cómo utilizar este conocimiento aplicado a problemas del mundo real.

Caso de negocio y objetivos comerciales

La predicción de valor de una casa es un logro bastante interesante para las compañías de bienes raíces. Para el proyecto, vamos a tomar en cuenta una agencia de compra-venta de casas que ofrece a los clientes vender su casa de manera fácil y rápida. Con un modelo entrenado de machine learning, la agencia puede obtener valor al saber la información de qué monto es el que le puede ofrecer a su cliente con el fin de maximizar su ganancia, de acuerdo a los datos recopilados del dataset. Una vez que se ha implantado el modelo, se espera que la agencia reduzca costos de tener que calcular el valor de cada casa manualmente, ya que este procedimiento será automatizado.

Expertise del desarrollador

Dado el trabajo en proyectos anteriores, el desarrollador ha adquirido habilidades de manejo de los lenguajes de programación C++ y Python principalmente, aunque también en una menor medida Javascript, Typescript y lenguajes funcionales como Clojure. Para el trabajo del frontend, se planea realizar una página web sencilla utilizando HTML5, que permita interactuar con el modelo, mediante la arquitectura de cliente-servidor.

Requerimientos funcionales

Los requerimientos funcionales para el proyecto son los siguientes:

1. El sistema tendrá una interfaz gráfica, la cual consta de un formulario en el cual se introducen los valores para generar una predicción.
2. El sistema desplegará un mensaje con la predicción.
3. El sistema desplegará la predicción en dólares americanos (\$USD).
4. El sistema solicitará la información de área en la medida de pies cuadrados (sf²).
5. El modelo de machine learning deberá tener una métrica de coeficiente de determinación de al menos 0.75 en el dataset de validación para asegurar que es efectivo.

Requerimientos no funcionales

El sistema debe contar con los siguientes requerimientos no funcionales:

1. El frontend se realizará con el framework React.
2. El backend se realizará con el framework Flask.
3. La página web se conectará al servidor a través de una petición HTTP POST.
4. El servidor deberá tener habilitado CORS para permitir a la página web conectarse.

Notebook de entrenamiento y análisis

<https://colab.research.google.com/drive/1RxsqMKhg5mjDPYJd9h2xIGeXO7eqjAxU?usp=sharing>

Preprocesamiento de la información

En el notebook de entrenamiento se presenta a detalle cómo es que se realiza el preprocesamiento del modelo. En resumen, se utiliza la librería de pandas para analizar primeramente los datos faltantes y realizar la imputación correspondiente. Posteriormente, se analiza si es necesario escalar los datos y se determina que sí, ya que las distribuciones no son normales y además, la variable que se quiere intentar predecir presenta un sesgo. Entonces se realiza dicho escalamiento con las técnicas combinadas de estandarización y logaritmo de la variable.

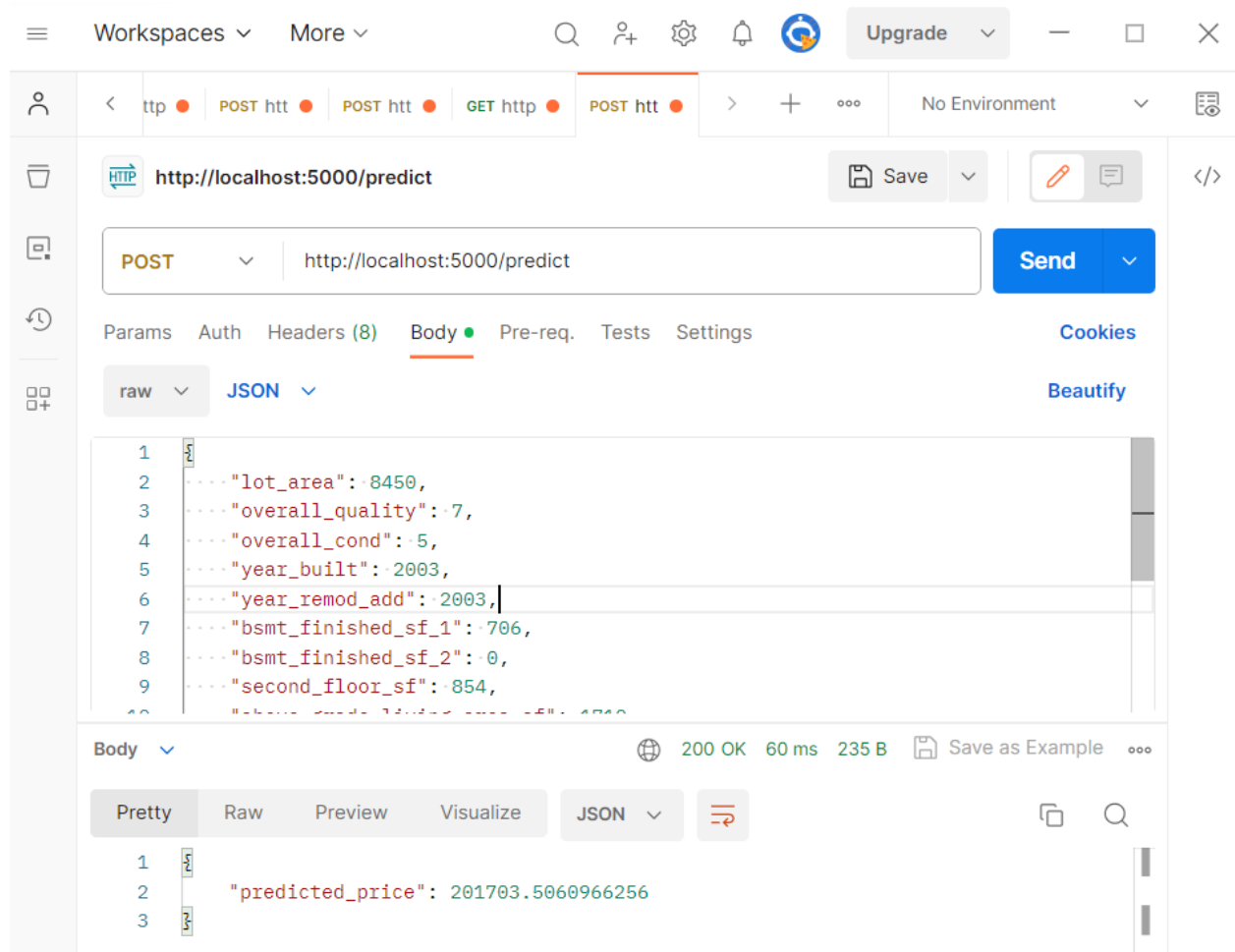
Generación del modelo

Una vez que se tienen los datos escalados, se observa que la variable que se intenta predecir tiene una forma mucho más parecida a la forma normal. Por lo tanto, empezamos a explorar un

modelo lineal. En el notebook se describe a detalle el procedimiento para encontrar las variables estadísticamente significativas para el modelo, y se van descartando conforme no lo son. Además, se intentan agregar variables categóricas para verificar si ayudan a mejorar el modelo. Sin embargo, se encuentra que no es así. El modelo final tiene un coeficiente de determinación de alrededor de 0.86. Posteriormente, se realiza el entrenamiento con el dataset de entrenamiento, y se prueban diferentes técnicas de regularización, las cuales son Lasso, Ridge y Elastic Net. Finalmente se exporta el mejor modelo.

Descripción del producto final

Se implementó un servidor utilizando Flask, que utiliza el modelo exportado para realizar predicciones mediante un API por HTTP POST. Para la vista de usuario, se utilizó el framework React, y se realizó un formulario que le permite al usuario final introducir los valores deseados.



Calcule el valor de su propiedad

Área del lote (pies cuadrados):
8450

Calidad del material de construcción (1 al 10):
7

Estado de la casa (1 al 10):
5

Año en que se construyó la casa:
2003

Año de última remodelación:
(mismo año en que se construyó si no aplica)
2003

Área del primer sótano (pies cuadrados):
706

Área del segundo sótano (pies cuadrados):
0

Área del segundo piso (pies cuadrados):
854

Área del inmueble por encima del nivel del piso
(pies cuadrados):
1710

Cantidad de baños en el sótano:
1

Cantidad de dormitorios (no sótano):

localhost:3000 dice
El precio de su propiedad es de \$201703.51 USD
Aceptar

Área del segundo piso (pies cuadrados):
854

Área del inmueble por encima del nivel del piso
(pies cuadrados):
1710

Cantidad de baños en el sótano:
1

Cantidad de dormitorios (no sótano):
3

Cantidad de habitaciones de cocina:
1

Cantidad de chimeneas:
0

Año de construcción del garage
(0 si no tiene garage):
2003

Cantidad de autos del garage:
1

Área del garage (pies cuadrados):
230

Calcular Precio

Conclusiones

De este trabajo, concluimos que las técnicas de Machine Learning pueden ser muy útiles para los problemas del tipo “realizar una estimación teniendo ciertos parámetros”. En este caso, la estimación del precio de una propiedad. Esto es algo importante no solamente desde el punto de vista ingenieril, sino también desde el punto de vista del negocio. Existe una gran oportunidad de crecimiento y de generar ganancias al utilizar este tipo de tecnologías, ya que se está

contribuyendo a resolver una necesidad identificada del cliente. Este es solo un ejemplo de los muchos en los cuales se pueden entrenar modelos de Machine Learning para la solución de una problemática.