Machine Learning

CSCE 5215

Group Project Proposal

HOUSE PRICE PREDICTION USING MACHINE LEARNING

TEAM MEMBERS

Name	EUID
Naga Sai Sivani, Tutika	11703058
Sai Chandana Kadala	11639945
Sarika Kannikanti	11715703
Keerthi, Lekkala Dilli	11703021

ABSTRACT

In this project, our central objective is to develop a machine learning model designed specially to predict house prices. The significance of a home purchase, often the most substantial financial commitment in a person's life, underscores the need for homeowners to have a dependable means of monitoring their investments. Additionally, prospective buyers would greatly benefit from the ability to forecast a home's future price before making a purchase, enabling them to make well-informed decisions.

First-time buyers frequently encounter challenges in identifying the crucial factors influencing home purchases and their impact on property costs. This lack of understanding can result in bad choices, potentially exposing them to unethical real estate agents who may lead them into overpaying for a property.

Our objective is to provide an effective machine learning model capable of predicting house prices in specific areas of interest, thereby empowering users to make better decisions when considering property acquisitions. Furthermore, this house price prediction model will serve a dual purpose, benefiting not only potential buyers but also property owners seeking to assess their real estate's current market value. Armed with this data, property owners can make informed choices regarding refinancing, selling, or leveraging their property's current market worth.

Throughout this project, our primary focus is on the meticulous development and testing of machine learning and artificial intelligence techniques to accurately assess property values based on historical data. Ensuring the model's precision and reliability in predicting house prices is our utmost priority.

Goals & Objectives:

Motivation:

Buying a house for many people is their biggest purchase of a lifetime and a dream they might have worked on for many years. The motivation behind pursuing this project lies in its potential to address a real-world problem with significant impact. Real-estate is not transparent like other assets such as gold, stocks or bonds. No one tracks it so closely and it's not same for everyone everywhere. It depends on area, people, seller, buyer and many other invisible factors a person could think of. The project is driven by the desire to assist people making well-considered choices when it comes to buying or valuing a home, thus enhancing their lives and financial well-being. The prospect of developing a machine-learning model having ability to provide useful insights into housing prices, potentially preventing buyers from making costly errors and aiding property owners in effectively managing their investments, serves as a compelling driving force. The opportunity to leverage artificial intelligence and machine learning to address practical challenges and ultimately, the project's potential to bring about a meaningful difference in people's lives is a powerful motivator.

Objectives

1. <u>Data Preparation & Quality Assurance:</u> Gathering and cleaning comprehensive housing data. Improving data quality by addressing missing values and outliers.
2. <u>Feature Engineering and Selection:</u> Identifying key influencing features, creating informative variables and optimizing feature set.
3. <u>Machine Learning Model Development:</u> Developing an accurate predictive models and fine-tuning them for optimal performance. Repeating until satisfactory results are met.
4. <u>Model Evaluation and Validation:</u> Assessing the developed models using different metrics and generalization and validating real-world performance.
5. <u>Documentation and Reporting:</u> Document project details and generate comprehensive performance reports and user guides.
6. <u>Transparency and Interpretability:</u> Maintain transparency in the process and enable users to interpret predictions effectively.
7. <u>Continuous Improvement and Maintenance:</u> Plan for ongoing model updates and incorporate feedback and market insights.

The goal is to deliver an accurate, adaptable house price prediction tool that empowers informed real estate decisions while maintaining transparency and facilitating continuous improvement.

Significance:

- An accurate machine learning model can possibly save or make a significant sum of monetary value.
- Real estate professionals and agencies can use this tool to aid there decision-making processes and develop better strategies.
- Helps in urban planning and infrastructure development by understanding where real estate prices are heading.
- Helps mortgage houses and financial institutions to refine their lending criteria based on predicted real estate values.

Features:

- Data collection: Collecting data from open-source datasets and resources. Finding reliable sources and making them available for further process of developing machine learning models for prediction.
- Data Preprocessing: Handling missing values, detecting outliers, normalization & standardization and any other data inconsistencies. The housing dataset also consists of many categorical features that need to be encoded.
- Exploratory Data Analysis: Understanding data through data visualizations and drawing out statistical and logical observations from results. It's especially important as the dataset has nearly 80 features. For building effective models we need to do extensive EDA to identify useful features.
- Model Development: Applying different machine learning models like regression, decision trees, random forests etc. And decide which one works best.

- Evaluation: Testing the accuracy of models with different metrics and comparing them to find the best model for deployment.
- Hyper parameter Tuning: Changing hyper parameters to find their optimal value for the model until maximum possible accuracy is achieved.
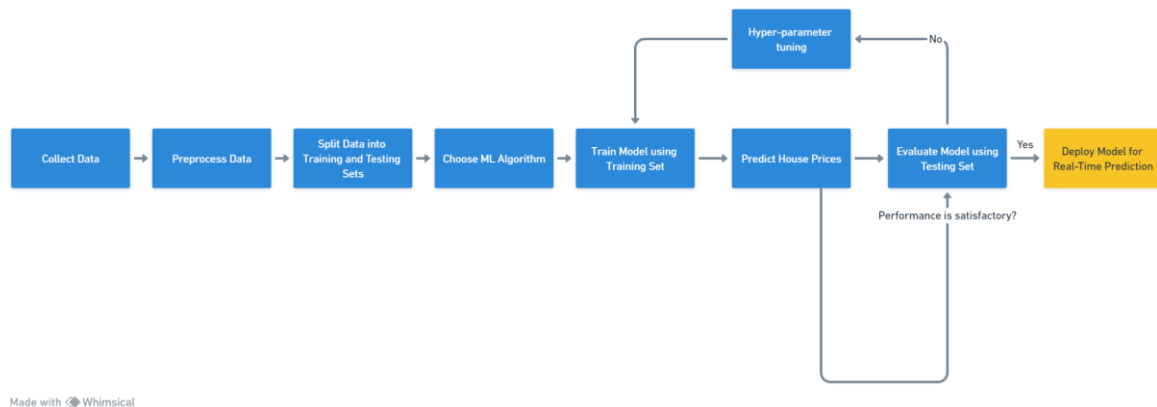

fig. project workflow

References

- Image in first page: https://www.reduceloans.com.au/wpcontent/uploads/2020/02/House-prices-rising-in-2020.jpg
- Dataset: House Prices - Advanced Regression Techniques | Kaggle
- Source for Kaggle dataset: Anna Montoya, DataCanary. (2016). House Prices - Advanced Regression Techniques. Kaggle. https://kaggle.com/competitions/house-prices-advanced-regression-techniques
- Workflow image made using whimsical: https://whimsical.com/ml-project-proposal-UwSAY3ug89FjywDymMrzU5