Velammal Institute of Technology
Velammal Knowledge Park, Panchetti, Chennai.

Predicting House Prices using Machine Learning

IBM AI GROUP - 1

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

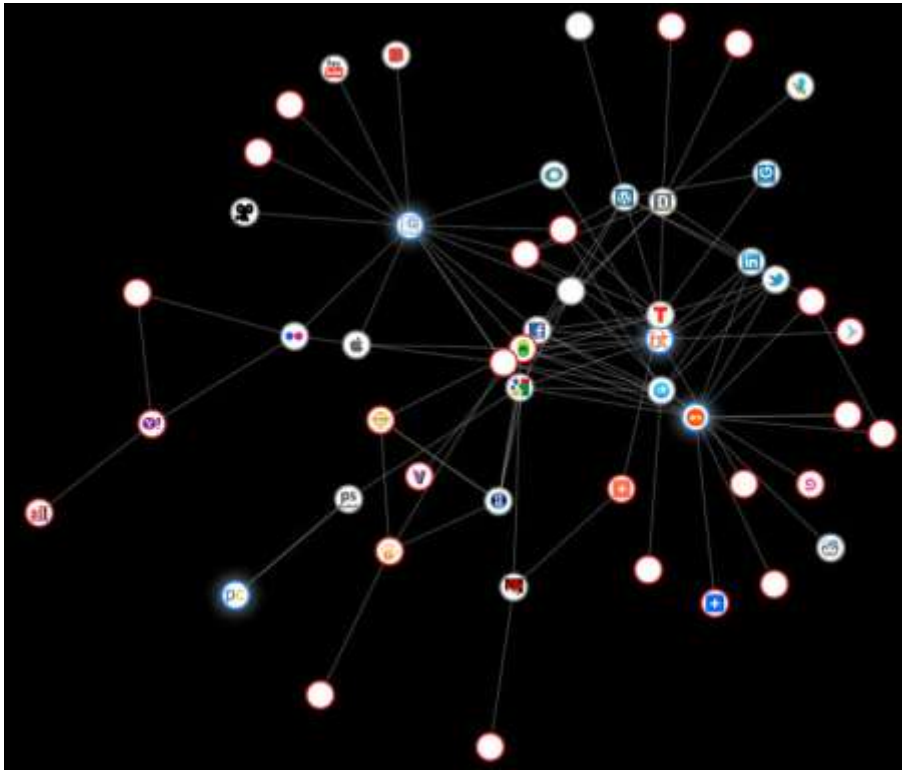
TEAM MEMBERS

SANDHIYA B(113321106080)
SRINIDHI M(113321106093)
SAI PREETHI K P(113321106078)
SHARADA S(113321106096)
PAPITA BISWAS(113321106066)

Introduction to the Problem

The Challenge

There's no one-size-fits-all formula to determine a home price. We will explore different algorithms and models to help solve this problem.



The Solution

By leveraging machine learning, we can eliminate guesswork — and the risk that comes with it — from pricing a home.

The Impact

Our approach ensures that potential homebuyers and sellers have access to accurate information that will enable them to make data-driven decisions.

Exploring the Dataset

Data Collection

We collected housing data from various sources including publicly available data, third-party agencies and private real estate companies.

Data Preprocessing

Using Python, we cleaned and merged our data through a number of techniques like data normalization and data transformation.

Data Visualization

To help us better understand key trends and patterns in the data, we explored them through visualizations like histograms and scatterplots.

Feature Selection and Engineeringture

1. Selection

Through trial and error, we identified the most significant features for predicting home prices such as location, square footage and number of bedrooms.

2. Feature Engineering

We also created new features like distance from nearest transportation hub, weather conditions and employment opportunities to improve model accuracy.

3. Normalization and Scaling

To ensure that features were on the same scale, we used various techniques like standardization and min-max normalization.

Model Selection and Training

- **Linear Regression**
  Simple yet powerful, linear regression is a good starting point for predicting home prices.

- **K-Nearest Neighbor**
  This algorithm is best suited for small datasets, where we can use distance metrics to calculate the similarity between data points.

- **Random Forest**
  One of the most popular algorithms for regression, random forest can help us overcome the challenges of overfitting and multicollinearity in a big dataset.

Model Evaluation and Fine-Tuning

Merits for performance evaluation

- **focus on the predictive capability of a model rather than how fast it takes to classify or built models,scalability,etc..**

		PREDICTED CLASS	
ACTUAL CLASS		Class -yes	Class-no
	Class-yes	a: IP	b: FN
	Class-no	c:FP	d: TN

Evaluation Metrics

Accuracy, root mean squared error and R-squared are typical metrics for regression models that help us evaluate the performance of our model.

Fine-Tuning Parameters

We fine-tuned the parameters of our models to optimize their performance and avoid overfitting through techniques like cross-validation, regularization and grid search.

Compare Models

We compared the performance of different models to understand which ones are best suited for predicting house prices.

Deploying the Model

Web Application

We created a web application to make our models more accessible to the public with the help of Flask, a Python web application fram

Hosting

We used GCP (Google Cloud Platform) and AWS (Amazon Web Services) to host our model and web application.

API

We also used REST API or GraphQL API to allow other developers to use our model in their applications.

Conclusion and Future Work

Conclusion

We successfully developed and deployed a model to predict home prices using machine learning. Our model could provide more accurate predictions than traditional methods.

Future Work

In the future, we could improve the model by adding more advanced machine learning algorithms, adding real-time data from MLS and expanding the market to other regions.

THANK YOU