AIRBNB PRICE PREDICTION

Domain - Machine Learning

Name of organization - Autotext.app

Name-Akash Patel
Gmail-akash1491625@gmail.com

High Level Design (HLD)

Document Version Control

Date Issued	Version	Description	Author
15-05-223		Air-bnb Price Prediction	Akash Patel

Contents Page No

Abstract

This study aims to predict Airbnb prices accurately by collecting a dataset with essential features that directly influence rental prices. The dataset is treated as a supervised learning problem, with known prices for each entry. The data will be divided into training, test, and cross-validation sets. Before proceeding, it is important to analyze the dataset's characteristics, including handling missing values, identifying data types, examining data size and shape, and conducting classical statistical analysis. These steps will inform subsequent decisions on data preprocessing and modeling techniques.

Introduction

Problem Statement-

To accurately predict Airbnb price, I aim to collect a dataset containing features which directly impact the rental price. No better place to start than by gathering a number of listings with fields directly from the site. Below you will find a list of the features that Ire taken from Airbnb and which turn out to be very important attributes in the price prediction. Since I know the price for each row, this can be classified as a supervised learning problem, and I will split my data into distinct training, test, and cross-validation sets. For now, I will examine the dataset as a whole, and come back to this division later. As a general rule, I like to examine a dataset's features for several characteristics before proceeding or deciding to gather additional data. These characteristics include:

- Number of missing values and how to deal with them (NaN or null)
- Type of data (categorical, boolean, image, numerical, text, etc)
- Shape and size of data (this impacts the type of model I will use)
- Classical statistical analysis (mean, median, range, variance, st. dev, etc)

Q- Why this High-Level Design Document?

Creating a project document or PDF is essential for my Airbnb pricing process. It is a complete list of the objectives, scope, approach and findings of my project. This information will enable effective communication with stakeholders, including managers, colleagues and customers. By sharing information, I can clearly understand the purpose, process and results of the project without direct intervention. Additionally, it helps to share information in my organization or work so that others can benefit from my understanding and thinking.In addition, project documentation documents key elements such as data, preliminary steps and modeling procedures, enabling repeatability. By incorporating this information into my PoIrPoint presentation, I improve collaboration, transparency, and my ability to expand and improve my work in the future.

Scope & Procedural Details -:

The scope of this project revolves around accurately predicting Airbnb prices by utilizing a comprehensive dataset comprising key features that directly impact rental prices. I aim to collect data on crucial factors such as location, property type, amenities, and listing attributes from Airbnb listings. By treating the problem as a supervised learning task, I will train a predictive model using the known prices for each listing. The dataset will be split into distinct sets for training, testing, and cross-validation to ensure reliable performance evaluation. Data preprocessing steps will be implemented to handle missing values, address categorical variables, and transform the data as required. Furthermore, I will conduct classical statistical analysis to gain insights into the dataset's distribution and characteristics. My objective is to develop a predictive model that accurately estimates Airbnb rental prices, which will benefit both hosts and guests in making informed decisions within the Airbnb marketplace.

- Data Collection: Gather a comprehensive dataset from Airbnb, including key features that directly impact rental prices, such as location, property type, amenities, and listing attributes.
- Supervised Learning: Approach the problem as a supervised learning task, utilizing the known prices for each listing to train a predictive model.
- Dataset Split: Divide the dataset into separate sets for training, testing, and cross-validation, ensuring accurate evaluation of the model's performance.
- Data Preprocessing: Perform necessary data preprocessing steps, such as handling missing values, dealing with categorical variables, and transforming the data as required by the selected modeling technique.
- Statistical Analysis: Conduct classical statistical analysis to gain insights into the dataset, including calculating mean, median, range, variance, and standard deviation, to understand the data's distribution and characteristics.
- Predictive Model Development: Utilize appropriate machine learning algorithms or regression techniques to develop a

predictive model capable of accurately estimating Airbnb rental prices based on the provided features.

- Model Evaluation: Assess the performance of the developed model using evaluation metrics like mean squared error (MSE), root mean squared error (RMSE), or R-squared (R^2) to measure the accuracy of price predictions.

- Application and Impact: Create a reliable and accurate pricing model that benefits both Airbnb hosts and guests. The model can empoIr hosts to set competitive prices for their listings, while helping guests make informed decisions when booking accommodations.

Product Perspective-:

This Airbnb price prediction project holds significant value from a product perspective. The development of an accurate pricing model for Airbnb listings offers several benefits for both hosts and guests. For hosts, the predictive model provides valuable insights into setting competitive prices for their accommodations, optimizing their revenue potential, and attracting potential guests. Hosts can make data-driven decisions based on factors such as location, property type, and amenities, leading to enhanced listing performance. On the other hand, guests benefit from the pricing model by gaining transparency and making informed choices when booking accommodations. The model helps them assess the fairness of prices and select suitable options within their budget. Overall, this project empoIrs both hosts and guests in the Airbnb marketplace, fostering better decision-making, increased satisfaction, and improved efficiency in the rental process.

Tools Used

Open-Smyce Ib Application - Jupyter Notebook.

Ib-based platform for version control and collaboration, primarily - GITHUB used for hosting and sharing code repository .

Web Engine for search - Google

The data preprocessing and analysis tasks Ire performed using popular Python libraries such as Pandas, NumPy, and Scikit-learn. I utilized Jupyter Notebook as my primary coding environment to execute data manipulation, exploratory data analysis (EDA), and model development. For the machine learning aspect, I employed regression techniques such as linear regression, decision trees, and random forests. I also utilized cross-validation techniques to evaluate and fine-tune my models. Visualizations Ire created using Matplotlib and Seaborn to gain insights into the data and present the finding effectively.

Functional Architecture

- Data Processing: In this phase, I implemented functions to handle missing values, encode categorical variables, and perform data scaling. Additionally, I incorporated techniques for data transformation to ensure the suitability of the data for modeling.
- Feature Engineering: Explain the functions or techniques used to extract and create meaningful features from the raw data. This may include methods like deriving new variables, aggregating information, or incorporating external datasets.
- Model Development: My project utilized various machine learning algorithms and regression techniques. I implemented functions to train and evaluate models, including linear regression, decision trees, and random forests. These functions Ire responsible for fitting the models to the training data and assessing their performance.
- Model Deployment and Prediction: Once the model was trained and validated, I deployed it for real-time predictions. I created functions that loaded the saved model, performed necessary data preprocessing steps, and generated price predictions based on new input data.
- Additional Functionality: My project also included additional functionality to enhance user experience and performance monitoring. This included functions for visualizing data, building interactive interfaces, and monitoring the model's performance.
- Flow and Dependencies: The functions within my architecture Ire designed to work together seamlessly, with data flowing from the initial preprocessing stage to the final prediction phase. I relied on external libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib to support my implementation.

Optimization

To enhance the performance of my Airbnb price prediction model, I employed several optimization strategies:

- Feature Selection: Through careful analysis and domain expertise, I selected the most influential features for my model. By considering feature importance, correlation analysis, and domain knowledge, I identified the key attributes that have a significant impact on rental prices, thereby improving the predictive power of my model.
- Hyperparameter Tuning: I systematically optimized the hyperparameters of my model to maximize its performance. Employing techniques such as grid search and random search, I explored various combinations of hyperparameter values to identify the optimal configuration. This process allowed me to fine-tune my model and achieve better prediction accuracy.
- Cross-Validation: To prevent overfitting and assess the model's generalization ability, I employed cross-validation techniques. Specifically, I used k-fold cross-validation to train and evaluate the model on different subsets of the data. This approach helped me estimate the model's performance and select the best-performing configuration.
- Regularization: To improve the model's robustness and prevent overfitting, I applied regularization techniques. By introducing penalties, such as L1 or L2 regularization, I controlled the complexity of the model and encouraged simpler, more generalizable solutions.
- Ensemble Methods: I leveraged ensemble methods to further enhance the predictive accuracy of my model. By combining multiple models, such as bagging or boosting, I achieved better prediction performance through aggregation or weighted voting.
- Evaluation Metrics: To assess the performance of my optimized model, I utilized appropriate evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or R-squared (R^2). These metrics allowed me to quantitatively measure the model's prediction accuracy and compare different configurations.
- By employing these optimization strategies, I significantly improved .

KPI(Key Performance Indicators)

Key Performance Indicators (KPIs) for my Airbnb price prediction project:

- Prediction Accuracy: I will measure the accuracy of my price predictions using evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or R-squared (R^2). A lower error value indicates higher prediction accuracy, ensuring reliable price estimates for hosts and guests.
- Feature Importance: I will determine the importance of each feature in my model. By using techniques like feature importance scores or permutation importance, I can identify the key factors that significantly impact rental prices, providing insights into the pricing dynamics.
- Model Training Time: I will measure the time required to train my predictive model. This metric will help me assess the efficiency of my model development process and ensure timely model updates as new data becomes available.
- Model Size: I will evaluate the size of my model in terms of memory usage or storage requirements. A smaller model size will facilitate easier deployment and scalability, allowing for efficient integration into production systems.
- Scalability: I will assess the scalability of my model by measuring its performance on different dataset sizes. This KPI will enable me to determine if my model can handle larger datasets and accommodate future growth.
- User Satisfaction: I will gather feedback from users, such as Airbnb hosts and guests, to gauge their satisfaction with the price predictions provided by my model. This feedback will help me understand the practical value and effectiveness of my model in meeting their needs.
- Return on Investment (ROI): I will assess the financial impact of my project by quantifying the cost savings or revenue increase achieved through more accurate price predictions. This KPI will demonstrate the value and potential return on investment of my project.
- Model Stability: I will monitor the stability of my model by evaluating its performance over time. This will involve tracking the consistency of predictions and detecting any significant changes or fluctuations in model performance.

- By tracking these KPIs, I can effectively evaluate the performance, efficiency, and impact of my Airbnb price prediction project, ensuring its success and identifying opportunities for further improvement.

Deployment

As a developer working on Airbnb's price forecasting project, I will focus on implementing the forecasting model and making it available to users. Integration of the model into a production environment will be done by placing it in a user-friendly web application. The app will provide users with a seamless interface to interact with models and get quotes for listings. I will use cloud services and container technologies like Docker to be flexible and flexible. This will enable the model to perform different tasks and accommodate future growth.

A good record will be created to collect, prioritize and replace new information, store the model with new information, to avoid conflicting information. Regular monitoring and auditing will be done to keep track of performance metrics such as forecast accuracy and response time to ensure the model is stable and accurate. In addition, detailed information will be provided to guide users on how to interpret the forecasts and solve any problems they may encounter. By following this distribution strategy, my goal is to provide hosts and guests with a user-friendly and friendly Airbnb price estimation solution so they can make a decision without knowing it.