

Airbnb Sentiment Analysis

Ahmed Alghaith
Naif bin barakah
Fahad Alotaibi

Introduction

- Airbnb
- Why Guest satisfaction
- potential solution



Problem definition

The process of building an AI model :

- Choosing the dataset
- Pre-processing
- Selecting an appropriate machine learning algorithm
- Training and testing the AI model
- Evaluating and fine-tuning the model



Tools

dmlc
XGBoost



TensorFlow


pandas

 **NumPy**

 **scikit
learn**

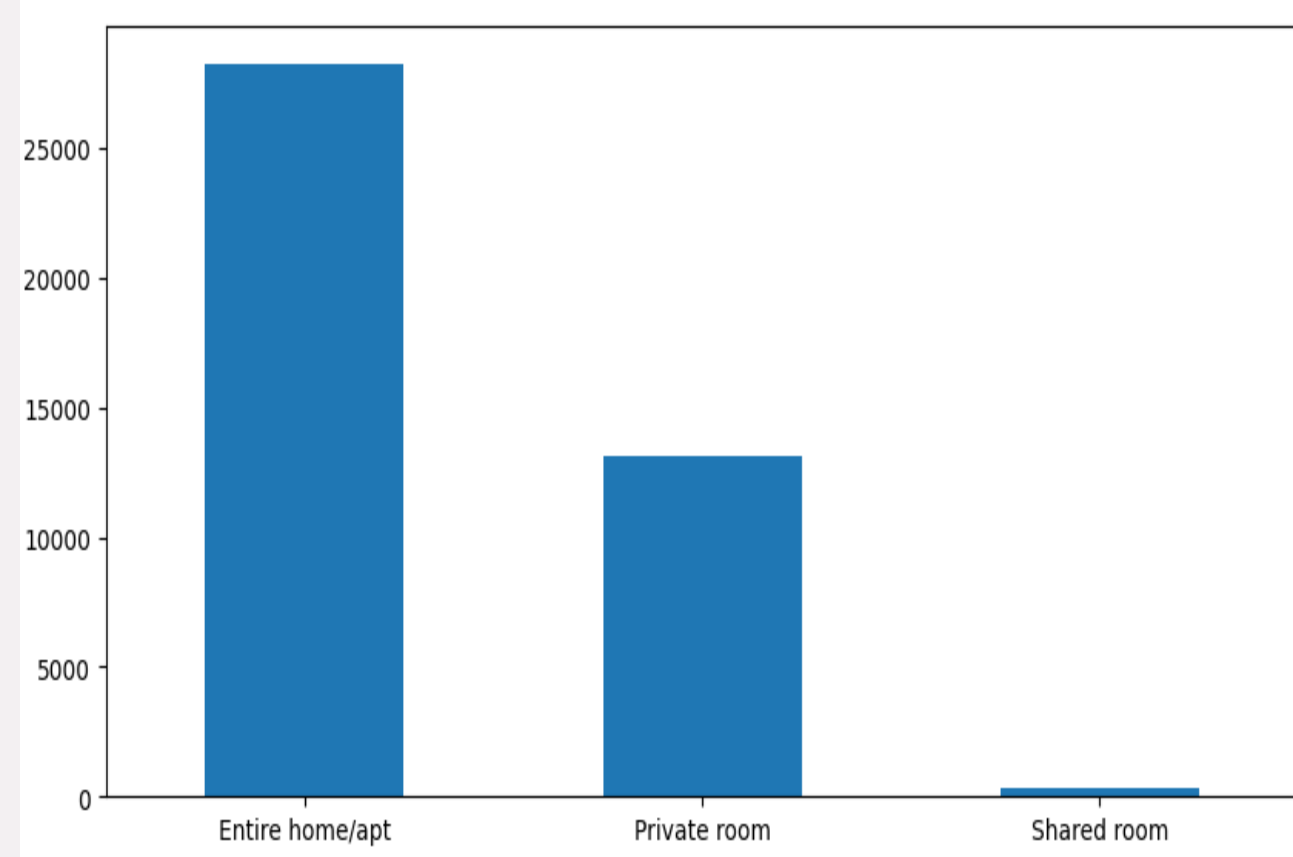
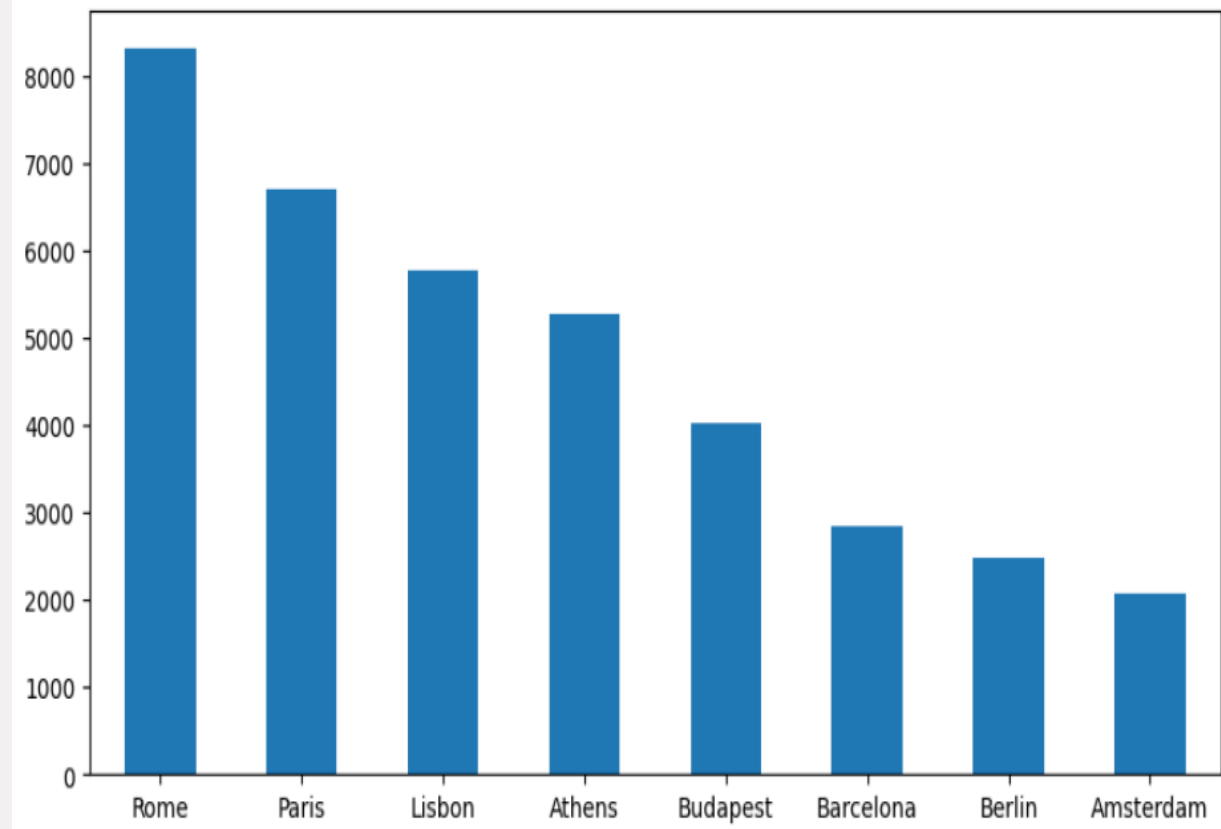
matplotlib

 **seaborn**

 **Keras**



Data Distribution



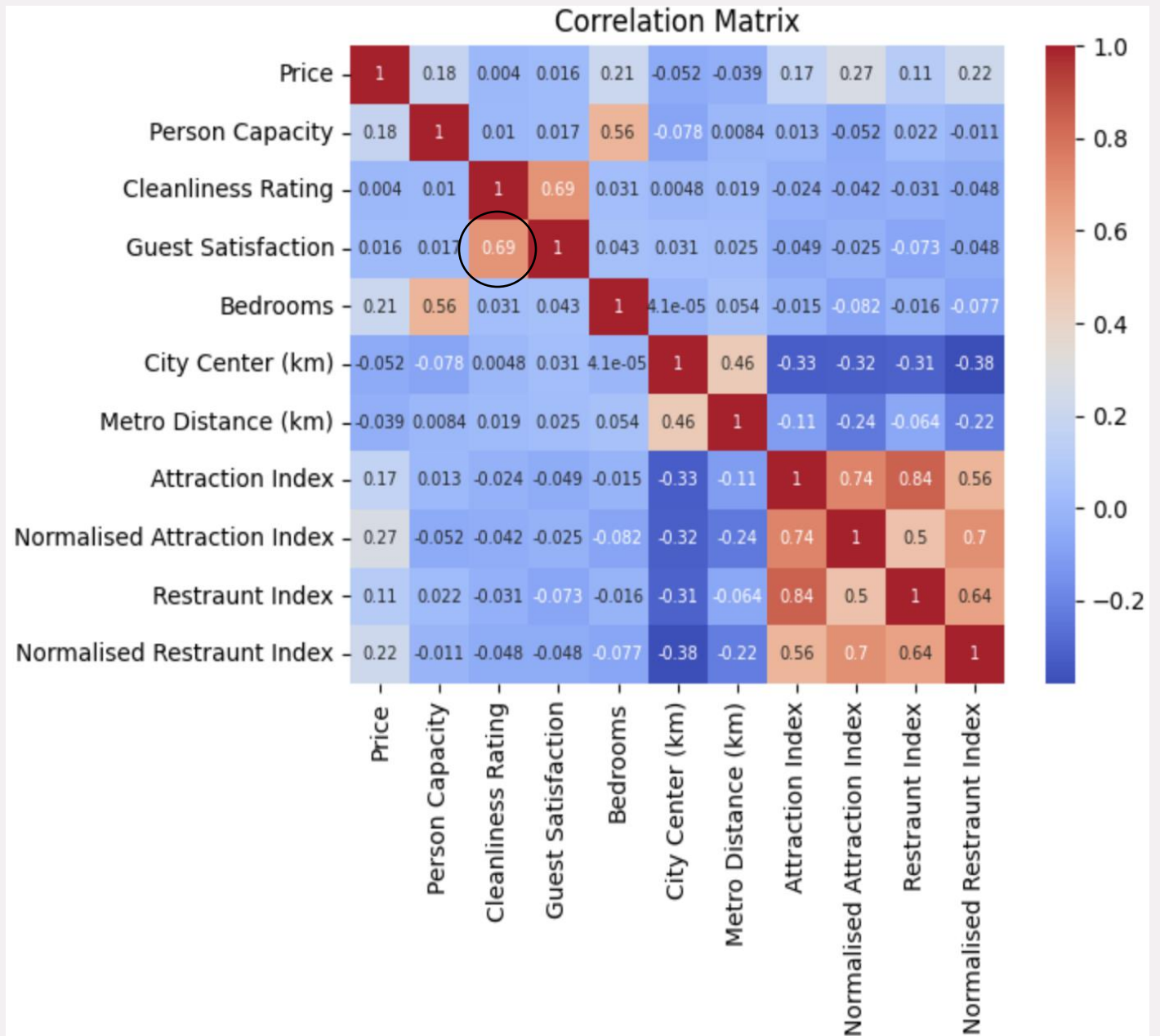
Data Pre-Processing

- **Handling Boolean data**
 - Convert true and false columns to 0 and 1.
- **Handling categorical data**
 - Convert categorical data (e.g., City) into numerical data using one-hot encoding.
- **Feature scaling**
 - Scale numerical features to have similar ranges using techniques such as standardization.
- **Feature engineering**
 - Create new features from existing ones that might be more useful for the machine learning algorithm.

Data Analysis

- High correlated features

- Cleanliness Rating
- Bedrooms
- City Center (km)



Neural Network architectures

	Hidden dense layers number	Neurons number with layers	Batch Normalization layers number	Validation accuracy	Training accuracy
Architecture1	3	All 16	0	92.85%	92.80%
Architecture2	6	All 16	0	92.81%	93.03%
Architecture 3	3	32,16,8	0	93.08%	92.92%
Architecture 4	3	All 16	3	92.99%	93.03%



Models cross validation scores

Gradient Boosting Classifier: 92.76%

Neural Network: 92.72%

Logistic regression: 92.45%

XGBClassifier: 94.07%

Support Vector machine: 92.54%

DecisionTreeClassifier: 92.81%

RandomForestClassifier: 94.15%





Model Hyperparameter tuning

Grid search with cross validation scores

RandomForestClassifier: 94.20%(+0.05%)

XGBClassifier: 95.34%(+1.27%)

DecisionTreeClassifier: 93.32%(+0.51%)

VotingClassifier(Hard): 94.44%

VotingClassifier(Soft): 94.31%

StackingClassifier: 94.17%

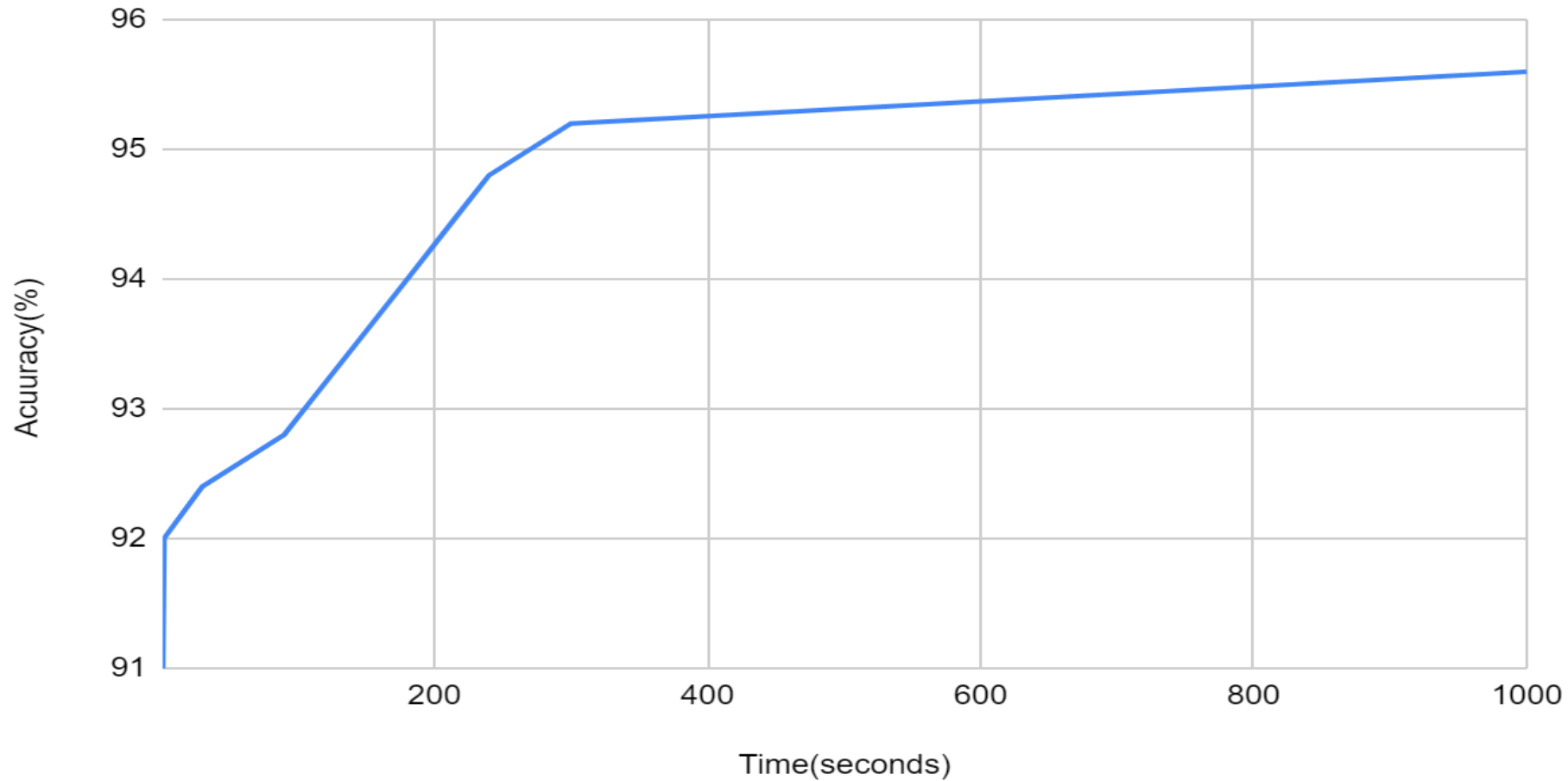
Report the results- XGBClassifier

Test accuracy: 96.27%

cross validation scores: 95.34%



Trade off Time Vs Accuracy





• Thank you!



Ahmed



Project repository



Fahad



Naif

