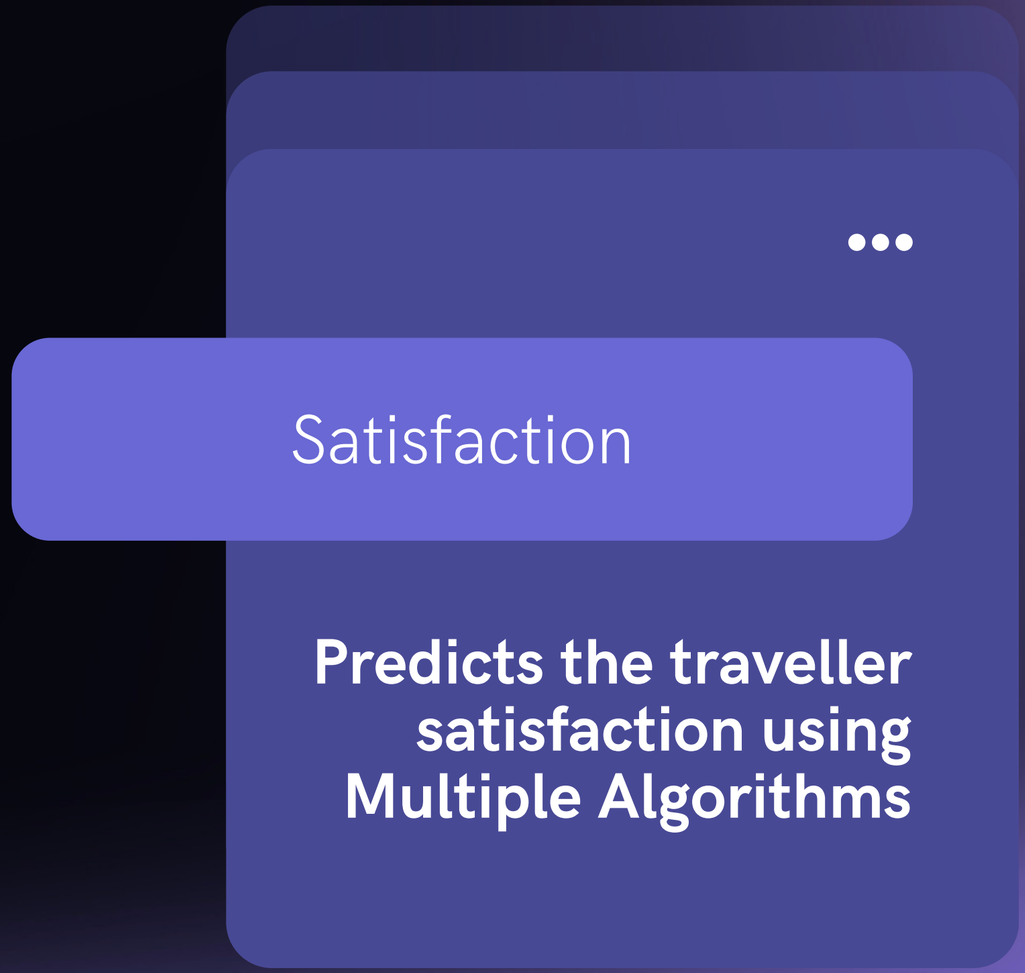
AI GENIUS FORGE

IDK

Traveller Satisfaction Prediction

A Project by Team IDK



Get to know us

M.M Navadeep - 99220041253

K.V Hitesh Kumar - 99220040586

T.V Sujith Gopi - 99220041389

B. Kushwanth Satya Sai - 99220041451

R.M.V.S Saketh - 9922008249

Problem Statement

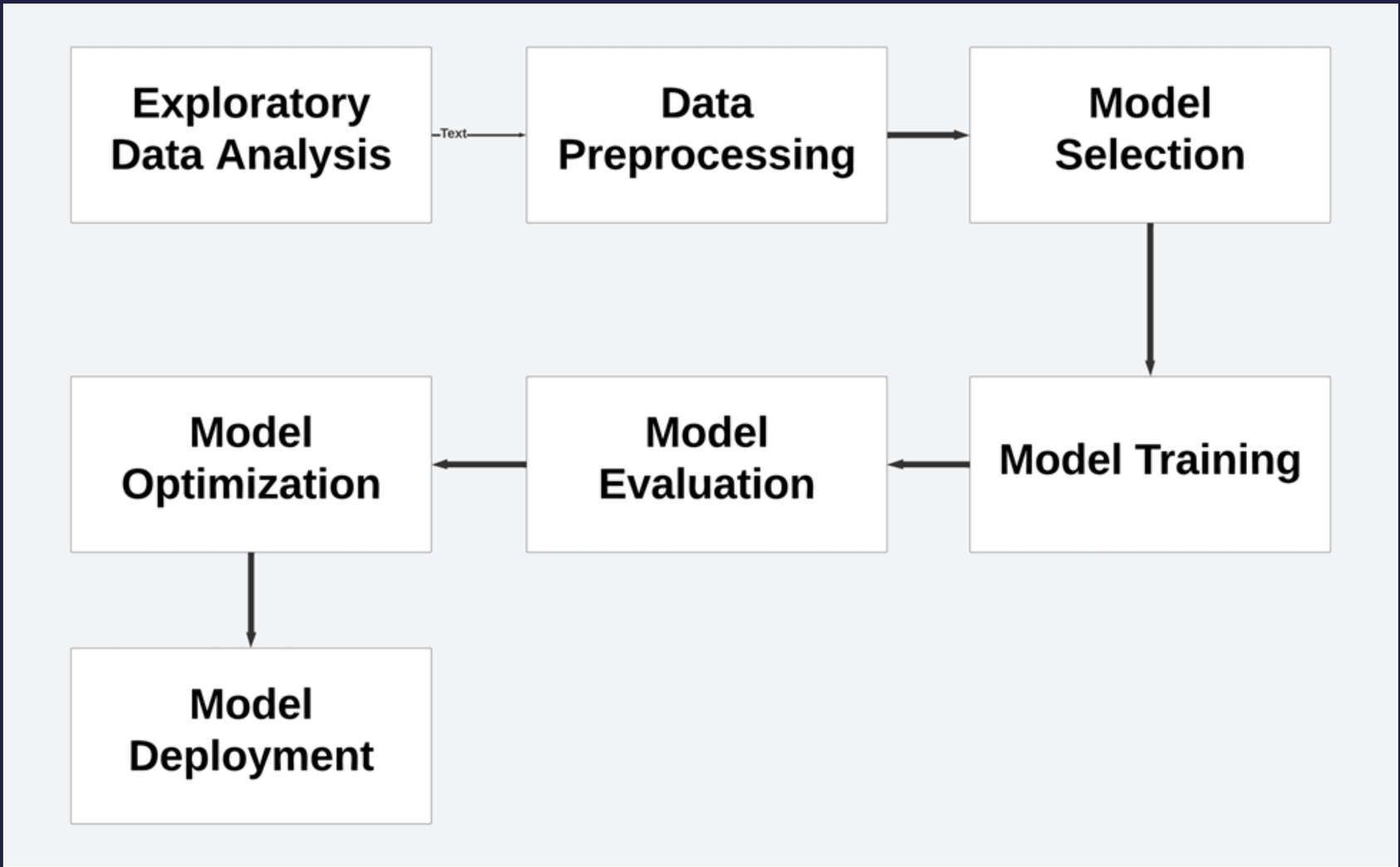
Travel Review Dataset Classification

This dataset contains passenger satisfaction surveys traveled in Flight.
Perform a granular level Exploratory Data Analysis on features.
Target : Predict passenger satisfaction.

Introduction

The objective of this study is to predict traveller satisfaction based on various factors captured in the dataset. The dataset includes features such as gender, customer type, age, type of travel, class, flight distance, and various aspects of the flight experience including inflight services, online booking ease, and departure delays. By analyzing this dataset, we aim to understand the key determinants of traveller satisfaction and develop predictive models to anticipate and improve overall satisfaction levels in the airline industry.

Flow Chart


Exploratory Data Analysis
Text
Data Preprocessing
Model Selection
Model Training
Model Evaluation
Model Optimization
Model Deployment

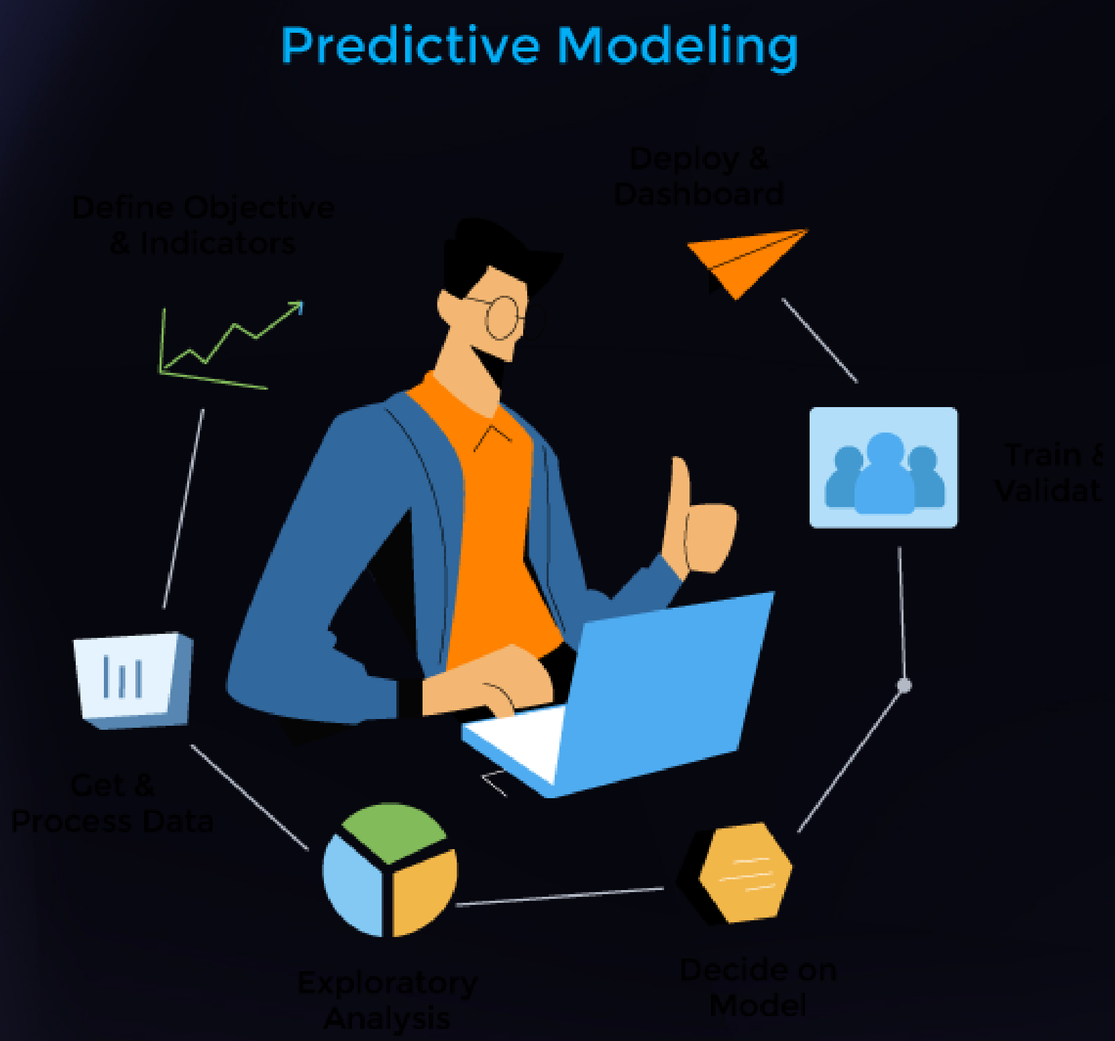
EDA

As a part of this, we've done:

- Data Collection and Loading:
  - Gather dataset from relevant sources (databases, files, APIs).
  - Load data into analysis environment (Python, spreadsheet software).
- Data Inspection:
  - Examine dataset structure, dimensions, and variables.
  - Identify issues like missing values, inconsistent data types, or outliers.
- Summary Statistics:
  - Calculate measures of central tendency (mean, median, mode).
  - Determine dispersion (standard deviation, range) and other relevant statistics.

EDA

- Data Visualization:
  - Create histograms, box plots, scatter plots, heatmaps.
  - Explore relationships between variables, detect patterns, and convey insights visually.
- Exploratory Analysis:
  - Investigate specific questions or hypotheses about the data.
  - Compare groups, examine correlations, identify trends over time.



Prediction Model's

Practical features or Prototypes to show the readers how your product or service makes money.

1 RandomForest Algorithm

2 SVC Algorithm

3 KNN Algorithm

RandomForest Classifier

- Random Forest Classifier is a robust and versatile,ensemble machine learning algorithm.
- Each tree is trained on a random subset of the data and features.
- It can be used for classification and regression tasks.
- Random sampling of data points (bootstrapping) is used.
- Random feature selection occurs at each split in the trees.
- The final prediction is determined by aggregating the outputs of the trees.
- For classification, this is done through majority voting.

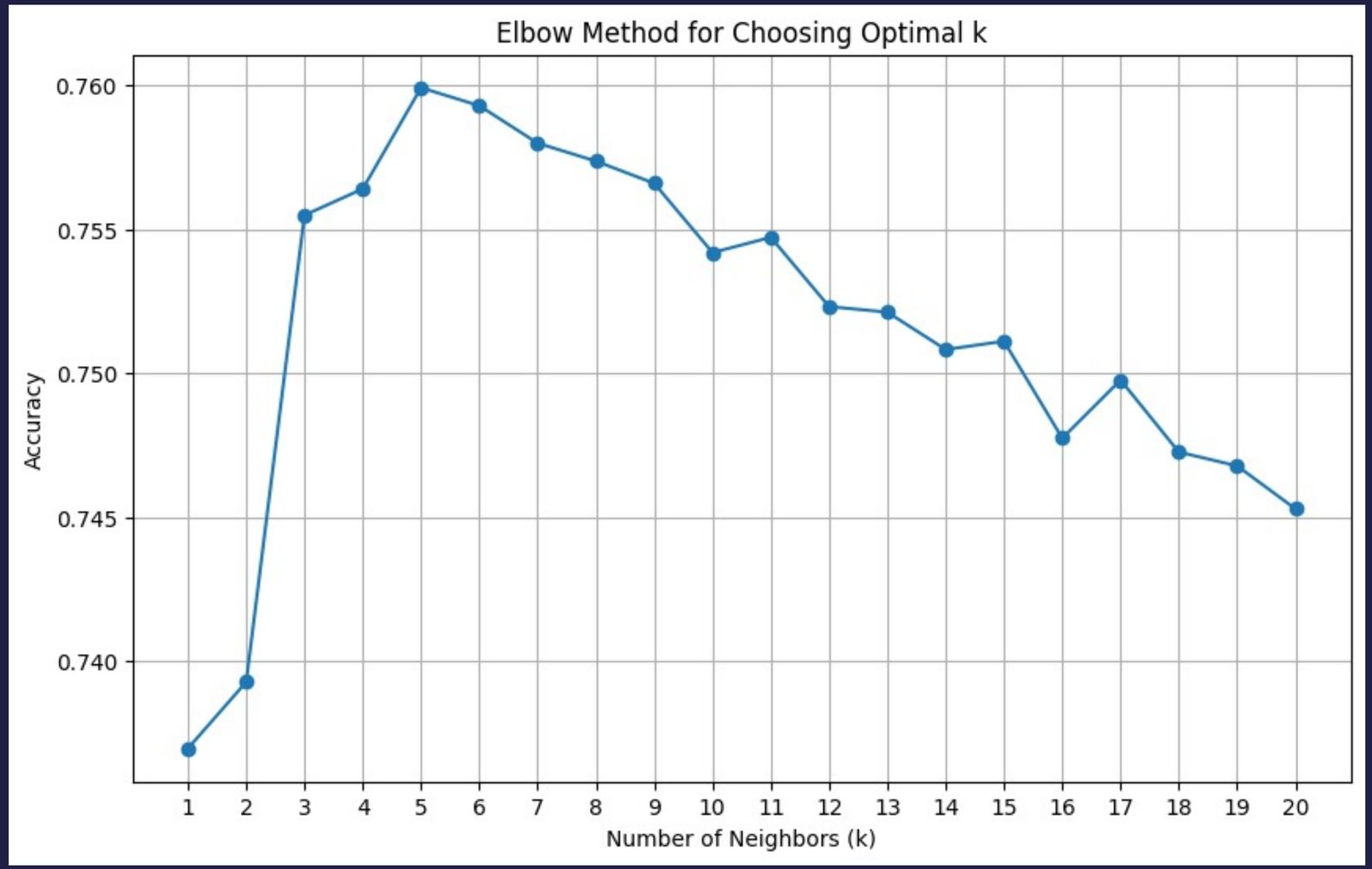
SVC Algorithm

- Linear SVC constructs a linear decision boundary to separate classes efficiently.
- It doesn't require kernel functions, making it computationally faster than non-linear SVMs.
- The algorithm aims to find the optimal hyperplane that maximizes the margin between classes.
- It uses a regularization parameter (C) to balance margin maximization and classification error minimization.
- While inherently a binary classifier, Linear SVC can be extended to multi-class classification.
- The algorithm is scalable and effective for various classification tasks.

KNN Algorithm

- K-Nearest Neighbors (KNN) is a straightforward and versatile algorithm used for classification and regression tasks.
- Elbow method in K-Nearest Neighbors (KNN) involves plotting the error rate against the number of neighbors (k) and identifying the point where the error rate starts to stabilize, helping to determine the optimal value of k.
- Predictions are made based on the majority class or average value of the k nearest neighbors.
- However, it can be computationally expensive for large datasets due to its reliance on calculating distances between data points.
- Despite this, KNN is robust to noisy data and works well with non-linear relationships.
- Its performance may degrade in high-dimensional spaces due to the curse of dimensionality.

Elbow Method for Choosing Optimal K



Conclusion

Through this project,

1. We analyzed the data
2. We preprocessed the data
3. We found correlation between features
4. Selected multiple models that best suit the current dataset
5. Trained and evaluated them
6. Generate a very user-friendly UI with Gradio and Streamlit
7. Implemented more features in the UI

THANK YOU!