Flight Fare Prediction

Using Machine Learning to Predict Flight Prices

CREATED BY:

VIJAY KUMAR SHAH

DATE :

04/09/2024

INTRODUCTION

INTRODUCTION TO FLIGHT FARE PREDICTION:

- FLIGHT PRICES ARE DYNAMIC AND FLUCTUATE BASED ON VARIOUS FACTORS SUCH AS DEMAND, TIME OF BOOKING, SEASONALITY, ETC.
- PREDICTING FLIGHT PRICES CAN HELP USERS SAVE MONEY AND PLAN BETTER.

PROBLEM STATEMENT:

- FLIGHT PRICE VOLATILITY CREATES UNCERTAINTY FOR TRAVELERS.
- THERE IS A NEED FOR A RELIABLE SYSTEM TO PREDICT FLIGHT PRICES ACCURATELY.

OBJECTIVE:

- TO DEVELOP A MACHINE LEARNING-BASED SYSTEM THAT PREDICTS FLIGHT FARES, ENABLING USERS TO MAKE INFORMED BOOKING DECISIONS.

VIJAY KUMAR SHAH

MOTIVATION

WHY THIS PROJECT?

- RISING FLIGHT COSTS MAKE PRICE PREDICTION ESSENTIAL FOR BUDGET-CONSCIOUS TRAVELERS.
- THE LACK OF ACCURATE PREDICTION TOOLS ON THE MARKET.

REAL-WORLD APPLICATIONS:

- HELPING TRAVELERS BOOK FLIGHTS AT OPTIMAL PRICES.
- ASSISTING TRAVEL AGENCIES IN OFFERING BETTER DEALS.

IMPACT:

- REDUCE TRAVEL COSTS FOR CONSUMERS.
- IMPROVE DECISION-MAKING FOR AIRLINES AND TRAVEL AGENCIES.

SCOPE AND OBJECTIVE

SCOPE:

- FOCUS ON DOMESTIC FLIGHT FARE PREDICTIONS WITHIN SPECIFIC ROUTES.
- UTILIZATION OF HISTORICAL DATA FOR TRAINING THE MODEL.

OBJECTIVES:

- DATA COLLECTION AND PREPROCESSING.
- MODEL SELECTION AND TRAINING.
- MODEL EVALUATION AND OPTIMIZATION.
- DEPLOYMENT OF THE PREDICTIVE MODEL ON A WEB INTERFACE.




LITERATURE WORK

RELATED WORK:

- OVERVIEW OF EXISTING FLIGHT FARE PREDICTION SYSTEMS (E.G., GOOGLE FLIGHTS, SKYSCANNER).
- COMPARISON OF DIFFERENT APPROACHES USED IN PREVIOUS RESEARCH, LIKE TIME-SERIES ANALYSIS AND MACHINE LEARNING.

TECHNOLOGY TRENDS:

- USE OF MACHINE LEARNING TECHNIQUES LIKE REGRESSION, TREE-BASED MODELS, AND DEEP LEARNING IN PRICE PREDICTION.

METHODOLOGY

DATA COLLECTION:

- SOURCES: KAGGLE, OPENSKY NETWORK, OR ANY OTHER AVIATION-RELATED DATASETS.
- DATA INCLUDES FLIGHT ROUTES, DATES, PRICES, AND AIRLINE INFORMATION.

DATA PREPROCESSING:

- CLEANING THE DATA BY HANDLING MISSING VALUES.
- FEATURE ENGINEERING: EXTRACTING DATE FEATURES (E.G., DAY OF THE WEEK, MONTH), ONE-HOT ENCODING CATEGORICAL VARIABLES (E.G., AIRLINES, SOURCE, AND DESTINATION).

METHODOLOGY

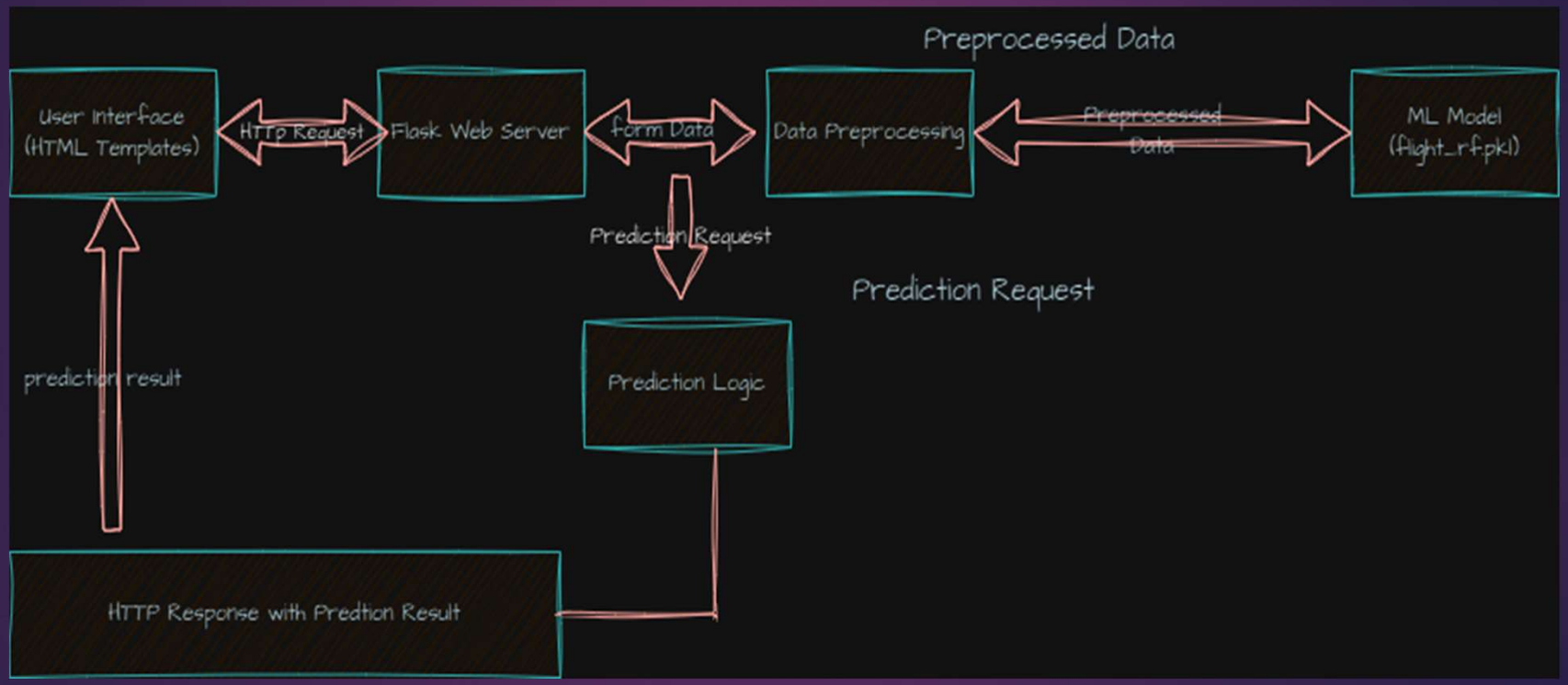
MACHINE LEARNING MODELS:

- INITIAL EXPLORATION WITH LINEAR REGRESSION.
- ADVANCED MODELS: RANDOM FOREST, XGBOOST.

MODEL TRAINING:

- SPLITTING DATA INTO TRAINING AND TEST SETS.
- TUNING HYPERPARAMETERS USING GRID SEARCH OR RANDOM SEARCH.

SYSTEM ARCHITECTURE



MODEL EVALUATION

PERFORMANCE METRICS:

- MEAN ABSOLUTE ERROR (MAE), ROOT MEAN SQUARED ERROR (RMSE), R-SQUARED (R^2).

COMPARISON OF MODELS:

- LINEAR REGRESSION: MAE = 500 INR, RMSE = 600 INR, R^2 = 0.70.
- RANDOM FOREST: MAE = 300 INR, RMSE = 400 INR, R^2 = 0.85.
- XGBOOST: MAE = 250 INR, RMSE = 350 INR, R^2 = 0.88.

BEST MODEL:

- XGBOOST PERFORMED THE BEST WITH THE LOWEST ERROR AND HIGHEST R-SQUARED SCORE.

IMPLEMENTATION AND TOOLS

TOOLS AND TECHNOLOGIES USED:

- PROGRAMMING LANGUAGE: PYTHON
- LIBRARIES: PANDAS, NUMPY, SCIKIT-LEARN, XGBOOST, MATPLOTLIB
- FRAMEWORK: FLASK OR STREAMLIT FOR DEPLOYMENT
- DATABASE: SQLITE OR MYSQL FOR STORING PROCESSED DATA
- VERSION CONTROL: GITHUB FOR MANAGING CODEBASE

IMPLEMENTATION AND TOOLS

DEVELOPMENT PROCESS:

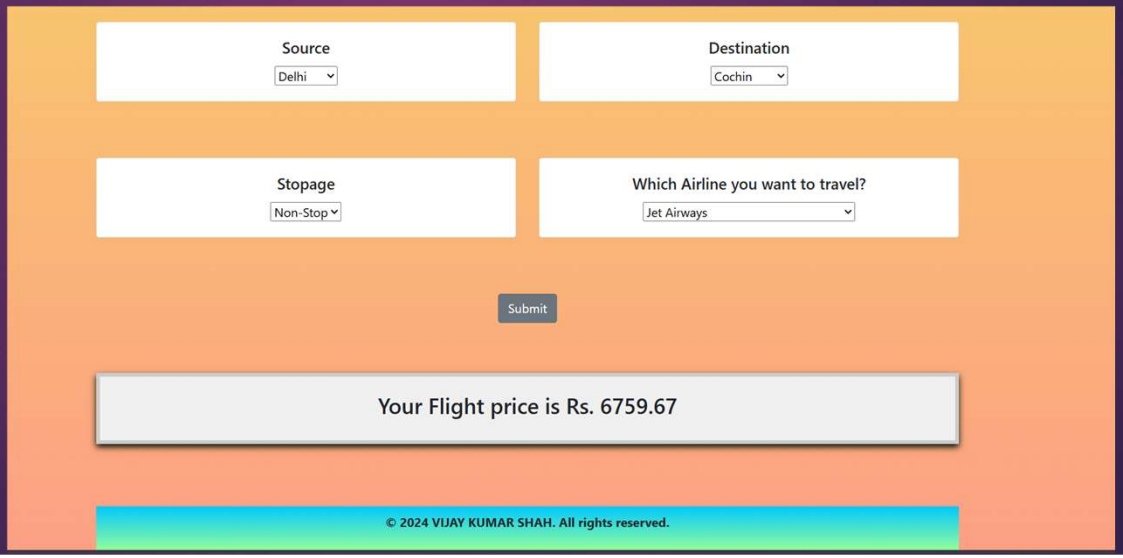
- DATA COLLECTION AND DATA PREPROCESSING.
- SPLITTING OF TRAIN DATA AND TEST DATA.
- MODEL TRAINING AND EVALUATION.
- HYPER PARAMETER TUNNING.
- DEPLOYMENT OF THE MODEL TO WEB APPLICATION.

WIREFRAMES AND USER INTERFACE

USER EXPERIENCE:

- USER-FRIENDLY INTERFACE WITH SIMPLE NAVIGATION.
- CLEAR PRESENTATION OF PREDICTED FARE WITH EASY-TO-UNDERSTAND VISUALS.

.



VIJAY KUMAR SHAH

CHALLENGES AND SOLUTION

CHALLENGES 1:

- HANDLING MISSING DATA AND OUTLIERS.

SOLUTION:

- USED DATA IMPUTATION TECHNIQUE TO HANDLE MISSING DATA.

CHALLENGES 2:

- DEPLOYING THE MODEL IN A USER-FRIENDLY INTERFACE.

SOLUTION:

- EMPLOYED FLASK/STREAMLIT FOR EASY AND EFFICIENT DEPLOYMENT

CONCLUSION

SUMMARY:

- SUCCESSFULLY DEVELOPED A MACHINE LEARNING MODEL TO PREDICT FLIGHT FARES.
- ACHIEVED A HIGH LEVEL OF ACCURACY WITH THE XGBOOST MODEL.

ACHIEVEMENTS:

- A FUNCTIONAL WEB-BASED APPLICATION THAT PREDICTS FLIGHT PRICES.
- POSITIVE IMPACT POTENTIAL FOR TRAVELERS AND THE TRAVEL INDUSTRY.

CONCLUSION

FUTURE WORK:

- EXPAND TO INTERNATIONAL FLIGHT ROUTES.
- INCORPORATE REAL-TIME DATA FOR EVEN MORE ACCURATE PREDICTIONS.
- DEVELOP A MOBILE APPLICATION FOR BROADER ACCESSIBILITY.

REFERENCE

CITATIONS AND RESOURCES:

- DATASETS USED (E.G., KAGGLE FLIGHT DATA).
- RESEARCH PAPERS OR ONLINE RESOURCES YOU REFERRED TO.
- DOCUMENTATION FOR LIBRARIES AND TOOLS (E.G., SCIKIT-LEARN, XGBOOST).

THANK YOU

feel free to ask queries

VIJAY KUMAR SHAH