PREDICTING PRICES OF FLIGHTS TO AND FROM ATL

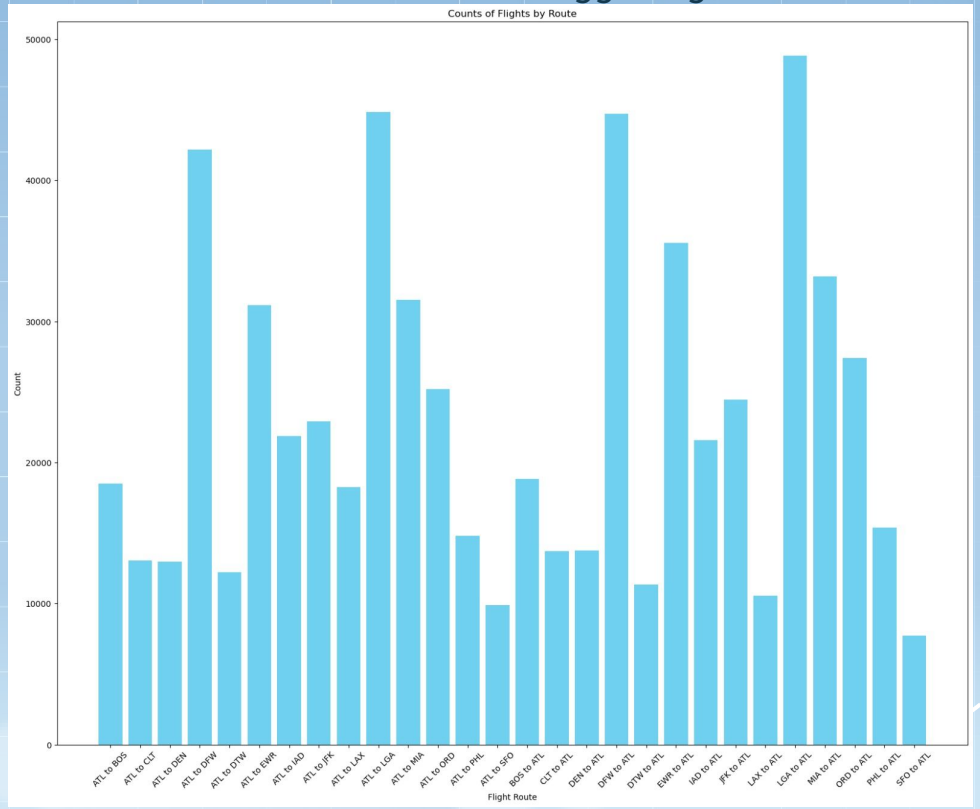
CS 4641 Group 31 Final Presentation

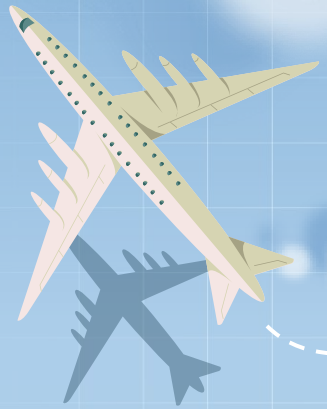
PROJECT DEFINITION

- Accurate flight cost prediction is vital in today's dynamic travel industry.
- Anticipating flight ticket prices significantly influences consumer choices and travel experiences.
- Our project aims to develop a machine learning model for effective flight cost prediction.
- Specifically geared towards Georgia Tech students, we focus on predicting costs for flights to and from ATL airport.
- To enhance specificity, our project excludes unrelated flights and is limited to economy-class seats.

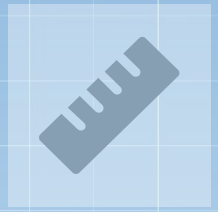
DATASET

from a Kaggle Flight Prices Dataset





CLEANING

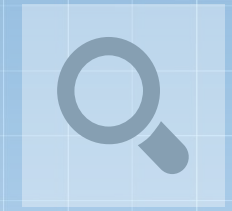


NORMALIZATION

- Numerical representations
- Feature extraction

PRUNING

- Removing features representing the same data
- Removing data irrelevant to our problem definition



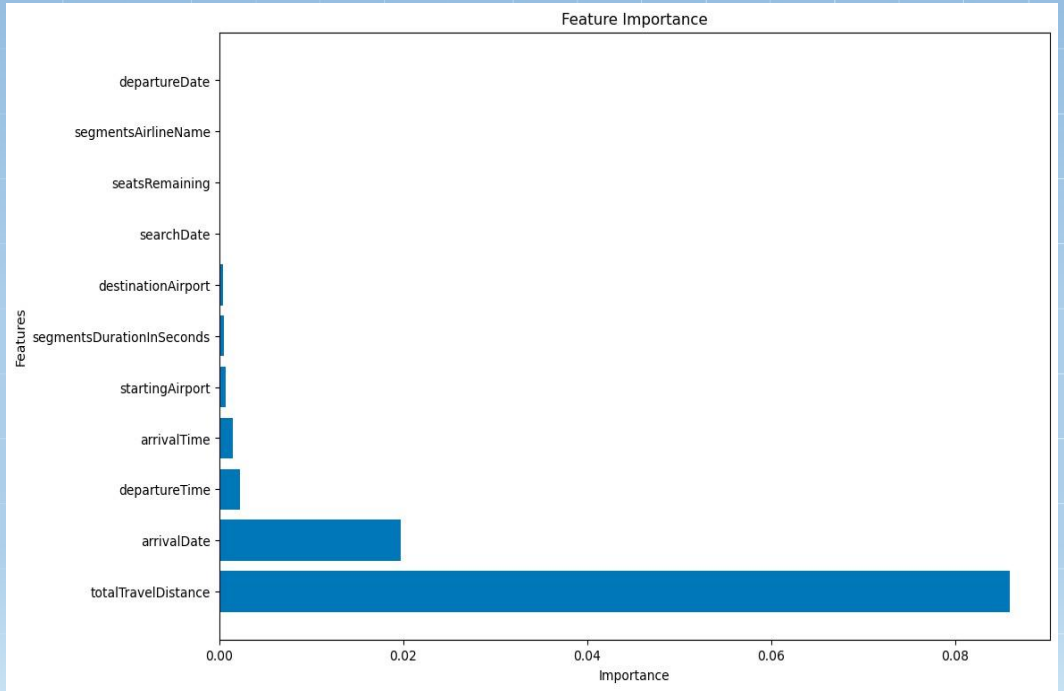
MISSING DATA

- Filling in missing values

PREPROCESSING

Lasso Feature Selection

- Determine feature importance
- Feature reduction



FINAL FEATURES

(in order of importance)

totalTravelDistance	Total travel distance in miles
arrivalDate	A number between 1-365 corresponding to the day of the year
departureTime	A number between 0-1439 corresponding to the time of the day in minutes
arrivalTime	A number between 0-1439 corresponding to the time of the day in minutes
startingAirport	Three-character IATA airport code for the initial location
segmentsDurationInSeconds	String containing the duration of the flight, in seconds, for the trip
destinationAirport	Three-character IATA airport code for the arrival location

METHODS

01

KNN

Supervised

02

LINEAR REGRESSION

Supervised

03

RANDOM FOREST

Supervised

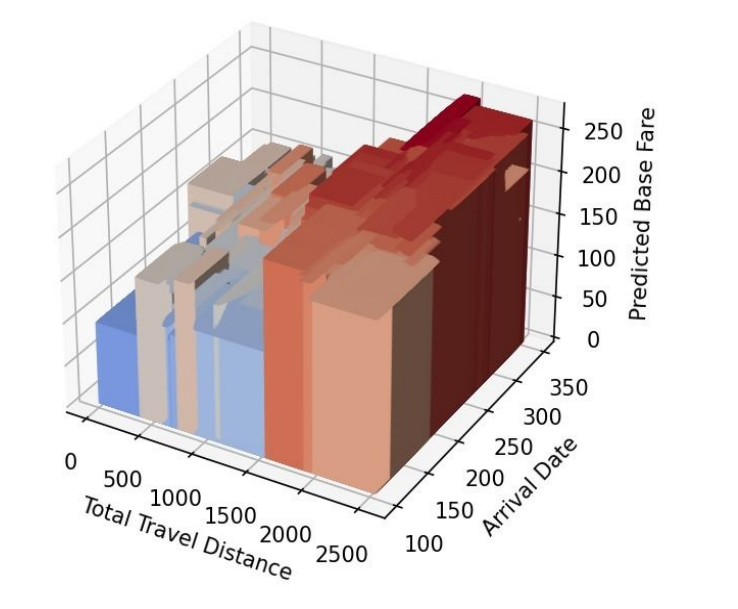
01

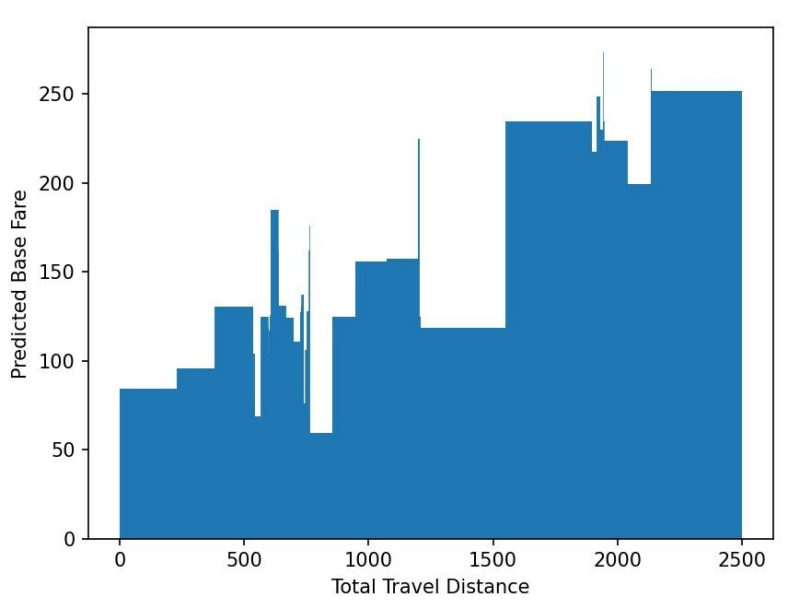
K NEAREST NEIGHBORS

Accuracy: ≈90%

KNN: RESULTS & ANALYSIS

- MSE: 208.86 - 211.54
- R-squared: 0.9041 - 0.9065





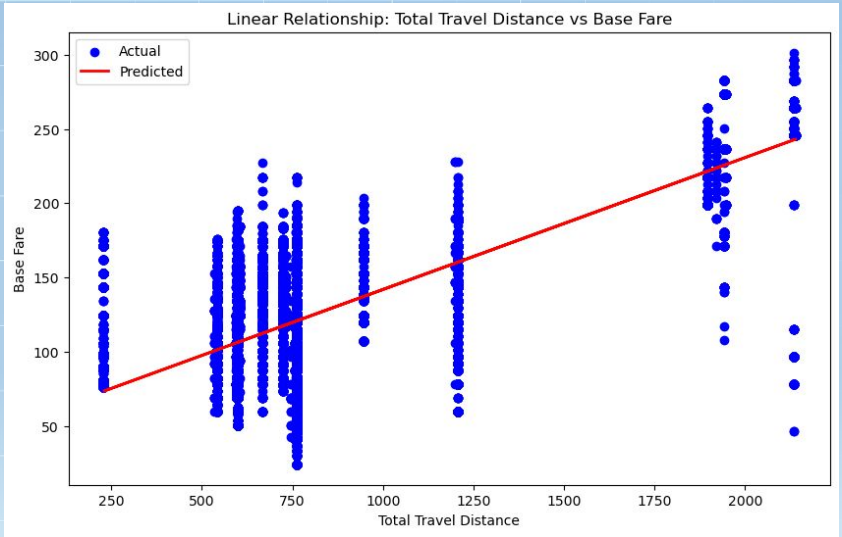
02

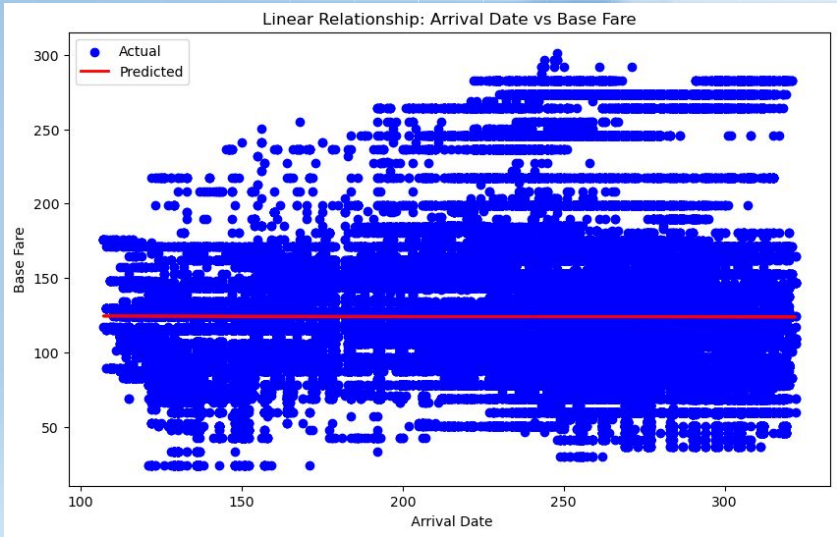
LINEAR REGRESSION

Accuracy: ≈50%

LR: RESULTS & ANALYSIS

- MSE: 1076
- R-squared: 0.51
- Reason for inaccuracy: non-linear relationships for some features





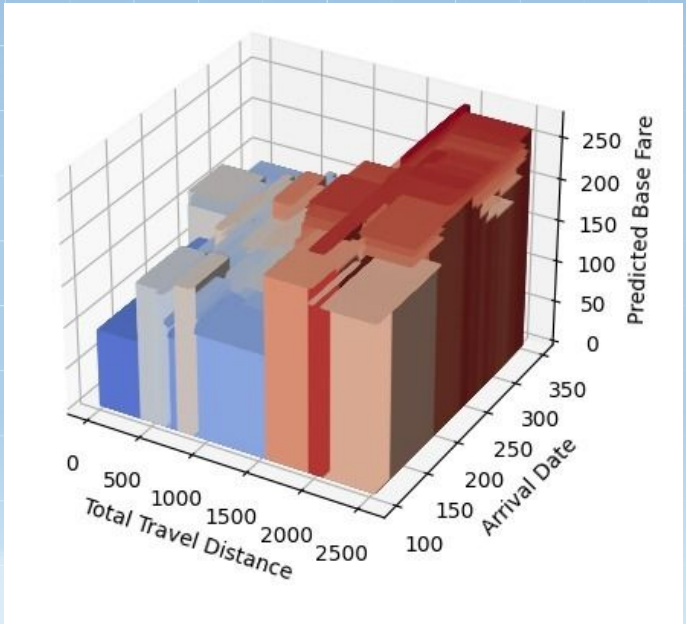
03

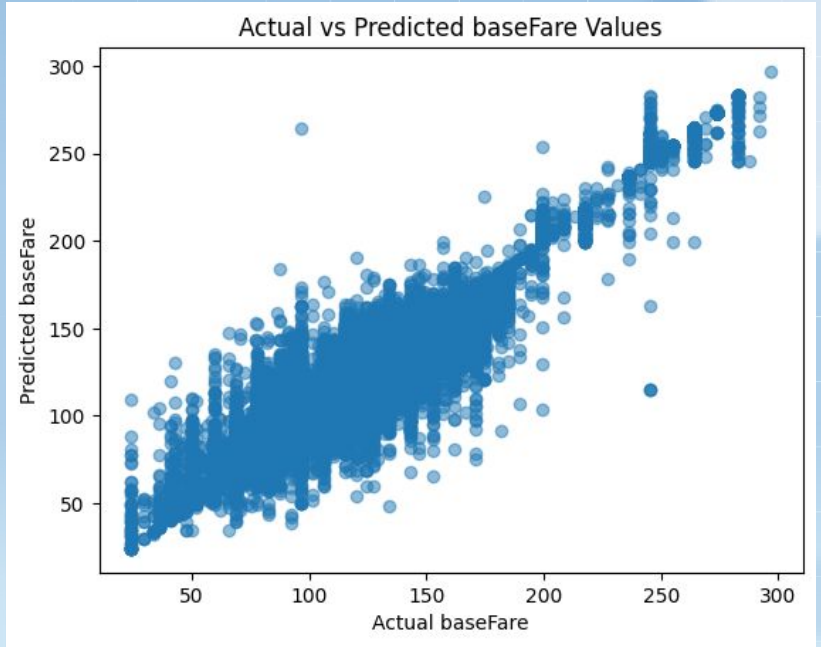
RANDOM FOREST

Accuracy: ≈96%

RF: RESULTS & ANALYSIS

- MSE: 88.46
- R-squared: 0.9599





MODEL COMPARISON

Aspect	KNN	Linear Regression	Random Forest
Non-linearity Handling	Effective in capturing complex, non-linear relationships in flight features.	Assumes a linear relationship between variables, may struggle with non-linear correlations.	Capable of capturing non-linear relationships, robust performance observed.
Prediction Accuracy	Achieved over 90% accuracy with optimized k-value and feature selection.	Achieved approximately 50% accuracy, reliable predictor despite non-linear correlations.	Achieved outstanding accuracy up to 95.99%, with consideration for computational costs.
Interpretability	Limited interpretability due to the complex nature of the algorithm.	May struggle with interpretability, especially in the presence of non-linear correlations.	Limited interpretability, challenging to explain the decision-making process.
Overfitting Concerns	Prone to overfitting, especially with small k-values. Requires careful tuning.	Potential for overfitting, especially with correlated features.	May exhibit overfitting, but observed performance on testing data alleviated concerns.

DISCUSSION & CONCLUSION

DISCUSSION

- KNN: showed high accuracy, but lacked interpretability
- Linear Regression: not reliable w/ less accuracy
- Random Forest: high accuracy, but at the expense of some interpretability

FUTURE WORK / NEXT STEPS

Test on future flight prices and explore prices for flight routes outside Atlanta

REFERENCES

K. Tziridis, T. Kalampokas, G. A. Papakostas and K. I. Diamantaras, "Airfare prices prediction using machine learning techniques," *2017 25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, 2017, pp. 1036-1039, doi: 10.23919/EUSIPCO.2017.8081365.

M. Mulkalla, Deepika and A. Joshi, "Predicting the Fare of a Flight Ticket with Machine Learning Algorithms," 2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022, pp. 1-5, doi: 10.1109/ICCR56254.2022.9995886.

S. N. Prasath, S. Kumar M and S. Eliyas, "A Prediction of Flight Fare Using K-Nearest Neighbors," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2022, pp. 1347-1351, doi: 10.1109/ICACITE53722.2022.9823876.