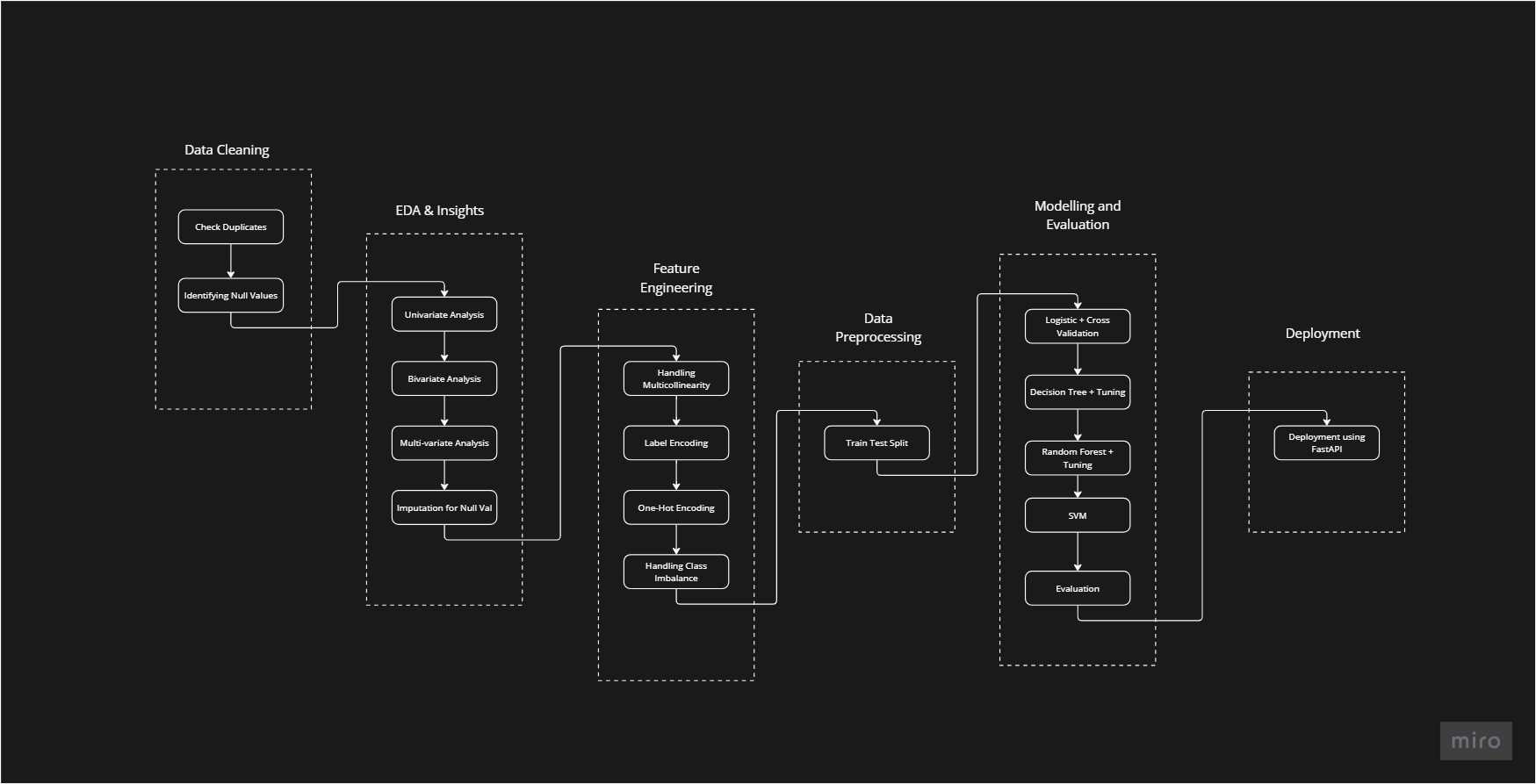
**Airline Referral Prediction Project**

****

**Introduction**

The project aims to predict whether a passenger referred by an existing customer will book a flight or not, based on various features such as seat comfort, cabin service, travel class, food & beverage, entertainment service, etc. The prediction model is developed using classification techniques in machine learning. The use of machine learning techniques allows for the development of a model that can learn from historical passenger and booking data and make accurate predictions on new data. The model can be used by airlines to target marketing campaigns towards potential passengers who are likely to book a flight based on a referral from an existing customer.

**Objective**

The given data includes airline reviews from 2016 to 2019 for popular airlines around the world with multiple-choice and free-text questions. Data was scraped in Spring 2019. The main objective is to predict whether passengers will refer the airline to their friends.

**Data Description**

The dataset consists of 131,895 entries with the following columns:

1. **Airline**: Name of the airline.
2. **Overall**: Overall rating given by the passenger.
3. **Author**: Name of the reviewer.
4. **Review Date**: Date of the review.
5. **Customer Review**: Text of the customer review.
6. **Aircraft**: Type of aircraft.
7. **Traveller Type**: Type of traveler (solo, couple, family).
8. **Cabin**: Class of travel (economy, business, etc.).
9. **Route**: Route of the flight.
10. **Date Flown**: Date of the flight.
11. **Seat Comfort**: Rating for seat comfort.
12. **Cabin Service**: Rating for cabin service.
13. **Food & Beverage**: Rating for food and beverage service.
14. **Entertainment**: Rating for entertainment service.
15. **Ground Service**: Rating for ground service.
16. **Value for Money**: Rating for value for money.
17. **Recommended**: Whether the passenger recommended the airline or not.

**Data Analysis and Preprocessing**

* Missing values were handled appropriately.
* Data types were converted as necessary for model compatibility.
* Features were selected based on their relevance to the prediction task.
* Exploratory data analysis was conducted to understand the distribution and relationships of the features.

**Key Insights**

1. **Traveler Type**: 37.22% of passengers travel solo, followed by couples and families.
2. **Travel Class**: 78.44% of passengers choose economy class.
3. **Cabin Service**: 50% of passengers gave a rating of 4.0 and 5.0 for cabin service, which positively impacts the airline business.
4. **Entertainment Service**: 30% of passengers are not satisfied with the entertainment service, giving a rating of 1.0.
5. **Seat Comfort**: 38% of passengers gave a rating of 4.0 and 5.0 for seat comfort, indicating that poor seat comfort negatively impacts overall satisfaction.
6. **Overall Rating**: 42% of passengers gave an overall rating below 3.0, suggesting significant room for improvement in airline services.

**Model Development**

Various classification models were developed and evaluated:

1. **Logistic Regression**
2. **Decision Tree**
3. **Random Forest**
4. **K-Nearest Neighbour**
5. **Support Vector Machine**
6. **Naive Bayes**

**Model Evaluation**

* Hyperparameter tuning was performed using Grid Search CV to optimize model performance and avoid overfitting.
* Evaluation metrics indicated that the Logistic Regression model provided the best performance.
* Support Vector Machine showed the highest accuracy rate by a very small margin.

**Feature Importance**

The most important features contributing to the model's prediction of whether a passenger will recommend an airline to friends were:

1. **Overall Rating**
2. **Value for Money**

**Recommendations**

To improve their business and customer satisfaction, airlines should focus on enhancing:

1. **Cabin Service**
2. **Ground Service**
3. **Food & Beverage**
4. **Entertainment**
5. **Seat Comfort**

**Conclusion**

The project successfully developed a predictive model with over 90% accuracy. By focusing on key service areas, airlines can increase customer satisfaction and referral rates, thereby growing their business.

**Future Work**

* Further refinement of models using advanced techniques such as ensemble methods.
* Incorporation of additional features such as demographic data for more personalized predictions.
* Continuous model evaluation and updating with new data to maintain accuracy and relevance.