**Comprehensive Documentation for Airline Passenger Referral Prediction Project**

**1. Introduction**

**Project Overview:**

Customer referrals play a significant role in airline businesses. Satisfied customers are more likely to recommend an airline to others, which can drive revenue growth and enhance brand loyalty. This project aims to build a predictive model to determine whether an airline passenger is likely to refer the airline based on various service-related factors.

**Business Objectives:**

* Understand key factors influencing passenger referrals.
* Identify patterns in customer feedback.
* Improve service quality and customer retention.
* Provide actionable insights for enhancing customer experience.

**Dataset Description:**

The dataset consists of airline passenger reviews, including:

* **Numerical Ratings**: Seat Comfort, Food & Beverage, In-flight Entertainment, Staff Service, Value for Money, etc.
* **Categorical Variables**: Gender, Travel Type (Business/Leisure), Airline Class (Economy/Premium/Business), etc.
* **Textual Reviews**: Customer-written feedback (if available).

**2. Data Preprocessing**

**2.1 Handling Missing Values:**

* Imputed missing numerical values using **mean/median strategies**.
* Categorical variables were handled using **mode imputation** or **category encoding**.
* Removed highly missing data fields to prevent model bias.

**2.2 Feature Engineering:**

* Created new features such as **average rating score**, **service quality index**, and **loyalty factors**.
* Performed **sentiment analysis** on textual feedback to classify customer sentiment as Positive, Neutral, or Negative.

**2.3 Encoding Categorical Variables:**

* Applied **One-Hot Encoding** for nominal categorical variables.
* Used **Label Encoding** for ordinal variables (e.g., Seat Class: Economy < Premium < Business).

**2.4 Scaling and Normalization:**

* Standardized numerical variables using **Min-Max Scaling** for better model convergence.

**2.5 Data Splitting:**

* Split dataset into **Training (80%)** and **Testing (20%)** to evaluate model performance.

**3. Machine Learning Models Implemented**

**3.1 Logistic Regression**

* A simple and interpretable model that predicts customer referrals based on probability scores.
* Used as a baseline model for comparison.

**3.2 Random Forest Classifier**

* An ensemble learning technique using multiple decision trees to reduce variance and improve accuracy.
* Works well with both categorical and numerical data.

**3.3 Random Forest with Hyperparameter Tuning**

* Optimized model performance using **GridSearchCV** and **RandomizedSearchCV** to fine-tune parameters such as:
  + n\_estimators: Number of decision trees.
  + max\_depth: Maximum depth of each tree.
  + min\_samples\_split: Minimum number of samples required to split a node.
  + min\_samples\_leaf: Minimum number of samples in a leaf node.

**3.4 Support Vector Machine (SVM) Classifier**

* A powerful model that identifies the optimal decision boundary between referring and non-referring passengers.
* Used **RBF Kernel** for better separation of complex data.

**3.5 SVM with Hyperparameter Tuning**

* Optimized model using:
  + C: Regularization parameter to balance complexity and accuracy.
  + Kernel: Linear, Polynomial, RBF for best feature transformation.
  + Gamma: Controls the influence of each data point.

**4. Model Evaluation Metrics**

**4.1 Confusion Matrix**

* Analyzed True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) to measure classification performance.

**4.2 Accuracy, Precision, Recall, and F1-Score**

* **Accuracy**: Overall correctness of predictions.
* **Precision**: Measures how many predicted referrals were actual referrals.
* **Recall (Sensitivity)**: Measures how many actual referrals were correctly identified.
* **F1-Score**: Harmonic mean of Precision and Recall.

**4.3 ROC Curve & AUC Score**

* **ROC Curve**: Plots True Positive Rate vs. False Positive Rate for different threshold values.
* **AUC Score**: Measures model’s ability to distinguish between classes.
  + **AUC = 1.0**: Perfect model.
  + **AUC > 0.9**: Excellent model.
  + **AUC ~ 0.5**: Random guessing.

**5. Insights & Business Recommendations**

**5.1 Key Factors Influencing Passenger Referrals**

* **Seat Comfort**: Highly correlated with customer referrals.
* **Staff Friendliness**: Significant impact on satisfaction levels.
* **Food & Beverage Quality**: Complaints about food influenced negative reviews.
* **Frequent Flyer Programs**: More loyal customers were likely to refer the airline.

**5.2 Observations from Recent vs. Older Reviews**

* Recent reviews were more **critical** of airline services compared to past years.
* Declining service quality in certain areas was a **major cause of customer dissatisfaction**.

**5.3 Actionable Recommendations**

* Improve **seat comfort and in-flight entertainment** to increase positive referrals.
* Invest in **customer service training** for staff to improve friendliness ratings.
* Enhance **food and beverage options** to cater to customer preferences.
* Personalize marketing offers for **loyal passengers** to retain long-term customers.

**6. Conclusion**

This project successfully implemented multiple machine learning models to predict airline passenger referrals. By analyzing customer feedback and service-related factors, we provided actionable insights for improving customer satisfaction and loyalty.

The evaluation metrics (such as accuracy, precision, recall, F1-score, and AUC) confirmed the effectiveness of the models used. Further enhancements, including advanced NLP for sentiment analysis and deep learning approaches, can be considered for future research.

**7. Future Enhancements**

* **Incorporate Deep Learning Models** such as LSTMs and Transformers for sentiment analysis on textual reviews.
* **Analyze Social Media Data** to capture real-time customer sentiments.
* **Expand the Dataset** by integrating more airline data from different regions.
* **Automate Hyperparameter Tuning** using advanced techniques like **Bayesian Optimization**.

By implementing these strategies, the airline industry can make more data-driven decisions and enhance customer experiences.