**Assignment 7 (Final group project)**

**Assignment Problem:**

The ultimate objective of this final project is to develop a model capable of predicting flight delays and estimating the likelihood of delays based on departure and arrival times and airports. The comprehensive solution involves a multi-step process encompassing data collection, analysis, feature engineering, model training, algorithm selection, and performance evaluation.

**Solution Steps:**

1. Data Collection: Gather the necessary data for analysis.
2. Data Analysis: Examine the dataset, identifying correlations between columns.
3. Feature Engineering: Select relevant columns, handle missing values, and address outliers. Represent airports as binary values.
4. Data Splitting and Model Training: Divide processed data into training and test sets and proceed with model training.
5. Algorithm Evaluation: Explore various algorithms using methods like confusion matrices and AUC-ROC curves.
6. Model Refinement: Iteratively fine-tune the model for enhanced performance.
7. Predictive Testing: Utilize the trained model to make predictions about flight delays.

**Algorithms Used:**

1. Random Forest: Introduces randomness, preventing overfitting, with fast training speed and the ability to handle high-dimensional data. It accommodates both discrete and continuous data without requiring normalization.
2. Gradient Boosting Machine (GBM): Demonstrates high prediction accuracy, captures nonlinear relationships well, and is robust to noise and outliers. Combining multiple models reduces the risk of overfitting.
3. Neural Network: Possesses strong nonlinear fitting ability, excelling in capturing complex relationships. Through dynamic tracking and deep learning, it reflects prediction results realistically. Continuous iteration and structure adjustment ensure stability, eliminating human subjective factors for authentic and objective results.

**Performance:**

1. Random Forest: Swift with relatively high accuracy.
2. GBM: Faster with higher accuracy than Random Forest. Iterative refinement enhances comprehensiveness and accuracy.
3. Neural Network: Slower but exhibits relatively high accuracy. Adjustments further enhance performance.

**Reflection on Learning Outcome:**

This project, initiated from scratch and independently completed, amalgamates knowledge acquired throughout the learning journey. From data analysis to feature engineering, model training, and iterative refinement, the experience uncovered challenges, led to the discovery of intriguing insights, and enriched understanding. The process not only showcased the application of acquired skills but also facilitated continuous learning in problem-solving contexts.