Net Rout Data Research

Analyzing the performance across all tested models allows us to pinpoint the best approach for predicting route optimality.

Based on the average cross-validation results for the classification task, here is the summary, performance ranking, and analysis of all the models tested.

Model Performance Summary (Classification)

The performance of the models is ranked based on two key metrics: the **F1-Score** (a balance of precision and recall) and the **Area Under the ROC Curve (ROC AUC)** (the model's ability to discriminate between classes).

Rank	Model Type	Average F1-Score (± Std Dev)	Average ROC AUC (± Std Dev)	Performance Tier
1	Random Forest	0.9644	0.9934	Elite
2	Gradient Boosting Machine (GBM)	0.9643	0.9823	Elite
3	Elastic Net (Logistic Regression)	0.9233	0.9485	Very Good
5	K-Nearest Neighbors (KNN)	0.8865	0.9143	Good
6	ANN (MLP) Classifier	0.6711	0.5923	Failed Baseline

Which Model is Better?

The models providing the best result and performance are the Random Forest Classifier and the Gradient Boosting Machine (GBM) Classifier.

Both models achieved a nearly perfect F1-Score of over 0.96 and ROC AUC scores extremely close to 1.0, indicating they can almost flawlessly predict whether a network route is optimal or not. The Random Forest edges out the GBM slightly in terms of ROC AUC, making it the most accurate discriminator.

Model Analysis and Interpretation

The distinct performance differences reveal crucial insights into your network route data:

1. **The Best: Tree-Based Ensemble Models (RF and GBM)**

Random Forest (RF)	**Gradient Boosting Machine (GBM)**
Performance: Excellent, highest ROC AUC.	**Performance:** Excellent, marginally lower ROC AUC than RF.

Analysis: The superior performance of these models suggests that the relationship between your features (Bandwidth, Delay, Trust Score, etc.) and the target Optimal is highly **non-linear**. These tree-based models excel at capturing complex, high-order feature interactions that simple linear models cannot. The data likely has many specific rules or thresholds that RF and GBM identify perfectly.

2. **The Contenders: Regularized Linear and Distance Models**

Elastic Net (Logistic Regression)	**K-Nearest Neighbors (KNN)**
Performance: Very Good (F1 =0.92).	**Performance:** Good (F1 = 0.89).

Analysis (EN): The strong performance of the Elastic Net (a linear model with regularization) suggests that, despite the underlying complexity, a core, approximate linear relationship *does* exist in the data. The regularization helped prevent overfitting, leading to highly robust performance for a linear classifier.

Analysis (KNN): KNN's good result indicates that **local similarity** is predictive. Routes that are close to each other in feature space (e.g., similar latency, similar bandwidth) tend to share the same optimal/non-optimal class. This is expected in highly structured data after proper feature scaling.

3. **The Failure: Artificial Neural Network (ANN/MLP)**

   ANN (MLP) Classifier
   Performance: Failed Baseline (F1 = 0.67, ROC AUC = 0.59).
   Analysis: The ANN model essentially **failed to learn** the difference between the two classes (predicting mostly Class 1). This is common for simple, untuned ANNs on tabular data. It suggests that the decision boundaries are either too subtle or too specific for the default network structure to capture effectively, and the optimization process (training) got stuck in a local minimum.

Finally

For practical application, the **Random Forest Classifier** is the best choice.

It offers the highest predictive power (ROC AUC of 0.9934), and because it is an **ensemble of decision trees**, it is generally easier to interpret than GBM, allowing you to identify which network metrics are most important for determining route optimality.