**continue building the public transport optimization**

**model**

1. The aim of feature engineering is to prepare an input data set that best fits the machine learning algorithm as well as to enhance the performance of machine learning models
2. In contrast, during model optimization, you either increase or decrease depth and width depending on your goals. If your model quality is adequate, then try reducing overfitting and training time by decreasing depth and width. Specifically, try halving the width at each successive layer
3. Optimization modeling is a mathematical approach used to find the best solution to a problem from a set of possible choices, considering specific constraints and objectives
4. In contrast, during model optimization, you either increase or decrease depth and width depending on your goals. If your model quality is adequate, then try reducing overfitting and training time by decreasing depth and width. Specifically, try halving the width at each successive layer.
5. The raw data is subjected to cleaning, normalization, and type conversion in the pre-processing stage to ensure data consistency and usability.
6. The feature engineering process involves further processing the pre-processed dataset using the mRMR algorithm. This algorithm selects highly relevant features, reduces redundant information, and enhances the representation of the classification model.
7. The next step involves training and validating the classification model using the feature-engineered dataset. The dataset is divided into a training set and a test set, with a ratio of 7:3.
8. The Catboost model is optimized using the Optuna framework, which selects the best hyperparameter configuration to enhance the model generalization ability and performance.
9. Due to the presence of missing data, duplicate data, and character data in the dataset, it is essential to address these factors to mitigate their adverse impact and improve the overall data quality
10. Finally, the results are evaluated and analyzed. Performance metrics such as accuracy, precision, recall, and F-value are calculated to assess the classification performance of the model on the test set. These metrics provide insights into the effectiveness of the proposed approach.