Muxin Liu Internet Prediction Project

Objective

The project aimed to showcase basic pre-processing techniques and train a Gradient Boosted Trees (GBT) model using TensorFlow Decision Forests on a dataset.

Pre-processing

- Ticket names were divided into sections, and passenger names were tokenized.
- Transformations were applied to the dataset, resulting in additional columns like "Ticket_number" and "Ticket_item".

Model Training

- Initially, a GBT model was trained using default parameters.
- Subsequently, the model was trained again using specific parameters.

Model Summary

- The model used is named "gradient_boosted_trees_model_2".
- The model type is "GRADIENT_BOOSTED_TREES" and it's designed for classification tasks.
- The model's architecture, including the number of parameters, layers, and other details, is given in the `model.summary()` output.

Key Insights

The document emphasizes the importance of variables in the model, suggesting that some variables significantly influence predictions. However, specific details on variable importance were not extracted from the provided content.

Recommendations

For a deeper understanding and evaluation of the model's performance, consider:

- Reviewing variable importance details to identify key predictors.
- Evaluating the model's accuracy, precision, recall, and other metrics.
- Conducting a feature importance analysis to understand the relative importance of each variable in the model's predictions.