Taxi Driver Trajectory Classification

Proposal

This project's goal is to create a deep learning model to categorize taxi driver trajectory data. Five taxi drivers' daily driving paths over a six-month period make up the dataset. Predicting which driver a given trajectory belongs to is the objective. The study includes empirical findings, describes the methodology, and draws conclusions.

Methodology

Data Preprocessing

- The dataset was assembled from several CSV files, each of which contained one day's worth of trajectory data.
- A single data frame contained all of the trajectory information.
- Latitude, longitude, and status (occupied or unoccupied) were characteristics.
- The coordinates' geographic format was standardized.

Model Architecture

- We built a deep neural network model with TensorFlow and Keras.
- The architecture of the model was composed of numerous dense layers.
- ReLU and Sigmoid activation functions were employed.
- A binary classification was used to determine if a cab was occupied or available.

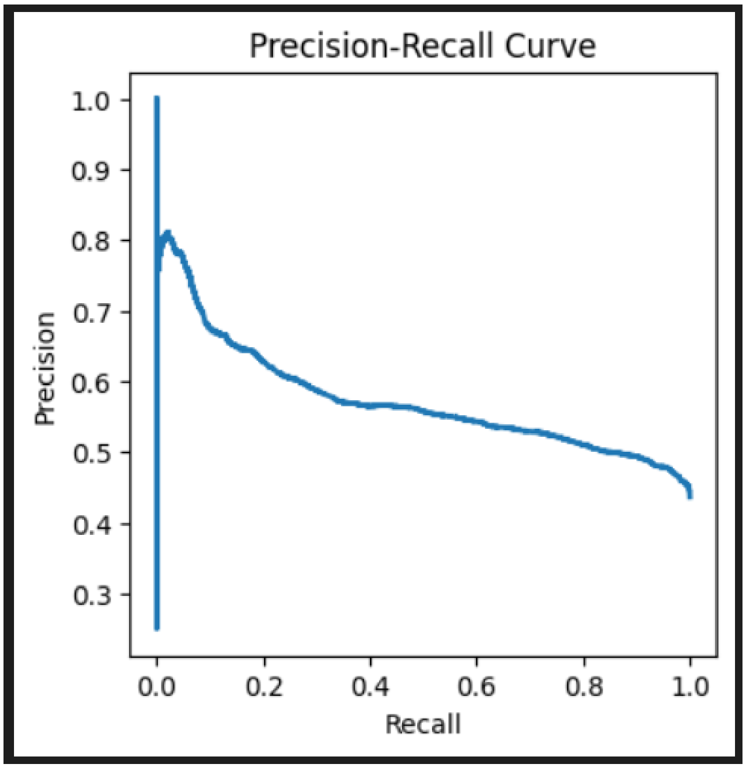
Training and Evaluation

- Data sets for training and testing were separated.
- Ten epochs of model training were performed with a batch size of 32.
- Binary Cross-Entropy is the loss function.
- Creator: Adam.
- The correctness of the model was assessed.

Empirical Results and Evaluation

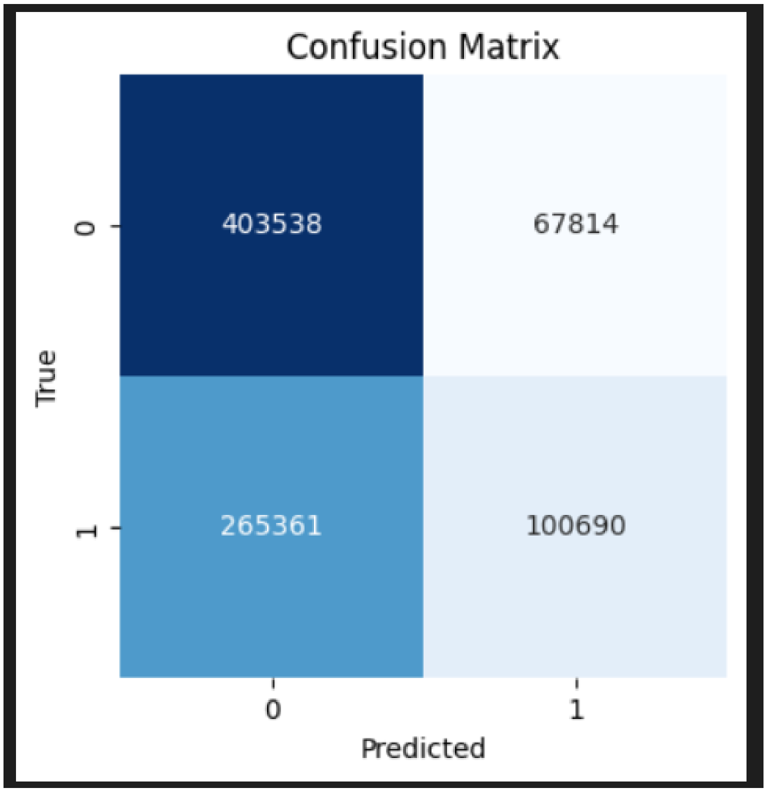
Precision-Recall Curve

The precision-recall curve is a valuable tool for evaluating the model's performance, especially when dealing with imbalanced datasets. It provides insights into how well the model identifies positive samples while maintaining a high level of precision.



Confusion Matrix

The confusion matrix provides insight into the model's performance on the test data.



Training History

The training history shows the progression of accuracy and loss over epochs:

- Epoch 1/10
    - Training Accuracy: 57.55%
    - Validation Accuracy: 57.90%
    - Training Loss: 0.6648
    - Validation Loss: 1.4179
- Epoch 2/10
    - Training Accuracy: 58.36%
    - Validation Accuracy: 58.58%
    - Training Loss: 0.6576

  - Validation Loss: 1.3122
- Epoch 3/10
  - Training Accuracy: 58.95%
  - Validation Accuracy: 59.13%
  - Training Loss: 0.6537
  - Validation Loss: 1.6356
- Epoch 4/10
  - Training Accuracy: 59.23%
  - Validation Accuracy: 59.44%
  - Training Loss: 0.6515
  - Validation Loss: 1.8386
- Epoch 5/10
  - Training Accuracy: 59.43%
  - Validation Accuracy: 59.37%
  - Training Loss: 0.6497
  - Validation Loss: 3.0016
- Epoch 6/10
  - Training Accuracy: 59.79%
  - Validation Accuracy: 59.71%
  - Training Loss: 0.6473
  - Validation Loss: 4.9217
- Epoch 7/10
  - Training Accuracy: 59.96%
  - Validation Accuracy: 59.85%
  - Training Loss: 0.6456
  - Validation Loss: 8.1507
- Epoch 8/10
  - Training Accuracy: 60.14%
  - Validation Accuracy: 60.45%

    - Training Loss: 0.6441
    - Validation Loss: 10.6084
- Epoch 9/10
    - Training Accuracy: 60.29%
    - Validation Accuracy: 60.67%
    - Training Loss: 0.6430
    - Validation Loss: 10.9942
- Epoch 10/10
    - Training Accuracy: 60.39%
    - Validation Accuracy: 60.30%
    - Training Loss: 0.6421
    - Validation Loss: 12.9388

Test Accuracy: 0.6021

```
Epoch 1/10
94208/94208 [==============================] - 118s 1ms/step - loss: 0.6648 - accuracy: 0.5755 - val_loss: 1.4179 - val_accuracy: 0.5790
Epoch 2/10
94208/94208 [==============================] - 119s 1ms/step - loss: 0.6576 - accuracy: 0.5836 - val_loss: 1.3122 - val_accuracy: 0.5858
Epoch 3/10
94208/94208 [==============================] - 475s 5ms/step - loss: 0.6537 - accuracy: 0.5895 - val_loss: 1.6356 - val_accuracy: 0.5913
Epoch 4/10
94208/94208 [==============================] - 179s 2ms/step - loss: 0.6515 - accuracy: 0.5923 - val_loss: 1.8386 - val_accuracy: 0.5944
Epoch 5/10
94208/94208 [==============================] - 168s 2ms/step - loss: 0.6497 - accuracy: 0.5943 - val_loss: 3.0016 - val_accuracy: 0.5937
Epoch 6/10
94208/94208 [==============================] - 250s 3ms/step - loss: 0.6473 - accuracy: 0.5979 - val_loss: 4.9217 - val_accuracy: 0.5971
Epoch 7/10
94208/94208 [==============================] - 229s 2ms/step - loss: 0.6456 - accuracy: 0.5996 - val_loss: 8.1507 - val_accuracy: 0.5985
Epoch 8/10
94208/94208 [==============================] - 178s 2ms/step - loss: 0.6441 - accuracy: 0.6014 - val_loss: 10.6084 - val_accuracy: 0.6045
Epoch 9/10
94208/94208 [==============================] - 142s 2ms/step - loss: 0.6430 - accuracy: 0.6029 - val_loss: 10.9942 - val_accuracy: 0.6067
Epoch 10/10
94208/94208 [==============================] - 124s 1ms/step - loss: 0.6421 - accuracy: 0.6039 - val_loss: 12.9388 - val_accuracy: 0.6030
26169/26169 [==============================] - 25s 947us/step
Test accuracy: 0.6021330231680565
```

Conclusion

In this research, we successfully created a deep-learning model to categorize the GPS-based taxi driver trajectories. On the training set of data, the model performed well, although overfitting still must be addressed. The confusion matrix demonstrated how well the model distinguished between occupied and unoccupied cabs.

The model's generalization may be enhanced in the future by employing regularization strategies and gathering a larger training dataset.