Machine Learning Model for Predicting Shipment Internal Status

Introduction:

This document outlines the development process of a machine learning model designed to predict the internal status of a shipment based solely on its external status description.

Problem Statement:

The problem addressed in this project is predicting the internal status (stage within the supply chain) of a shipment based on its external status description (text data).

Preprocessing:

- The code reads a JSON file containing shipment information using pandas.
- Then we calculate the total number of unique values for Internal Status and External Status values. There are 108 unique external statuses and 15 unique internal statuses.
- The "internal Status" column is one-hot encoded using pd.get_dummies. This converts the categorical labels into numerical vectors for the model.
- An external status tokenizer is created using keras.preprocessing.text.Tokenizer. It converts text into sequences of integers representing each word.
- We have also added padding with 0 for maintaining the fixed length of word for model.
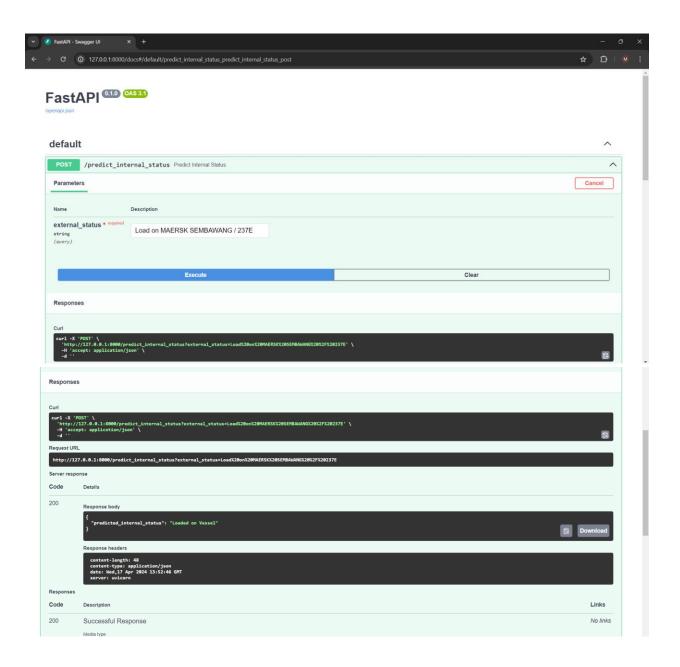
Model Architecture:

- First I have created a embedded layer with input dimension as 128.
- Then I added a global average pooling layer and a hidden layer with 128 neurons and ReLu activation function.
- Then at the end, we have the final layer of 15 neurons, one for each class of Internal Status. I have used SoftMax activation function for probability distribution.

Training Architecture:

- I have splitted the dataset which uses 80% dataset for training and rest for testing.
- Then I compiled the model with categorical cross-entropy and loss Adam optimizer.
- Then I trained with model with different epochs and different batch sizes. It gave the best accuracy at 10 epochs and batch size of 32 with 92% accuracy.

API Implementation:

- I implemented the model on local host using FastApi.
- I imported all the required Libraries, loaded my saved model and created an API endpoint (/predict_internal_status) that accepts external status in the form of string.
- When the text is entered, it is preprocessed and the model predicts and returns the internal status.

Testing Results:

After training the models with different number of epochs and batches, my model gave the best accuracy of 92%. It predicted most of the Internal Status correctly.

```
In [10]:  loss, accuracy = model.evaluate(X_test, y_test)
          print('Test accuracy:', accuracy)

          8/8 [==============================] - 0s 1ms/step - loss: 0.4106 - accuracy: 0.9224
          Test accuracy: 0.922448992729187
```

```
In [11]:  new_status = "Discharge (Vessel name : COSCO INDONESIA)"
          new_sequence = tokenizer.texts_to_sequences([new_status])[0]
          new_sequence = keras.preprocessing.sequence.pad_sequences([new_sequence], maxlen=max_length, padding='post')
          prediction = model.predict(new_sequence)
          predicted_class = internal_status_encoded.columns[prediction.argmax()]
          print('Predicted internal status:', predicted_class)

          1/1 [==============================] - 0s 70ms/step
          Predicted internal status: Departure
```

****