Project 02: Name Classification for Regional Identification

Initial Model Architecture and Parameters Analysis

The initial neural network model for this classification task was created using TensorFlow's Keras API and includes 3 layers. This consists one input layer, two hidden layers and one output layer. Input layer is designed to accept 13,003 features which corresponds to the preprocessed input data. The first hidden layer consist of 32 neurons and second hidden layer consist of 16 neurons and both employs the ReLU activation function. The output layer containe 18 neurons and that match the total number of unique class identified in the target variables SR_Name and used the Softmax activation function which outputs a possibility distribution over the 18 classes making it well-suited for multi-class classification. The total number of trainable parameter in this initial model is 416,962 . The model was trained for 50 epochs with a batch size of 32 and training time is 281.18 seconds.

```
[18] import pandas as pd

     # To see the number of input features and output classes (for context)
     print("Number of input features (N):", X_train.shape[1])
     print("Number of output classes:", len(df['SR_Name'].unique()))

     Number of input features (N): 13003
     Number of output classes: 18
```

```
[19] # To see the model summary (for total parameters and layers)
     # Re-create the initial model definition temporarily to get its summary
     model_initial_cls = create_classification_model(input_shape=[X_train.shape[1]], num_classes=len(df['SR_Name'].unique()))
     model_initial_cls.summary()

     Model: "sequential_1"
```

Layer (type)	Output Shape	Param #
dense_3 (Dense)	(None, 32)	416,128
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 18)	306

```
     Total params: 416,962 (1.59 MB)
     Trainable params: 416,962 (1.59 MB)
     Non-trainable params: 0 (0.00 B)
```

```
Initial classification model training time: 281.18 seconds
```

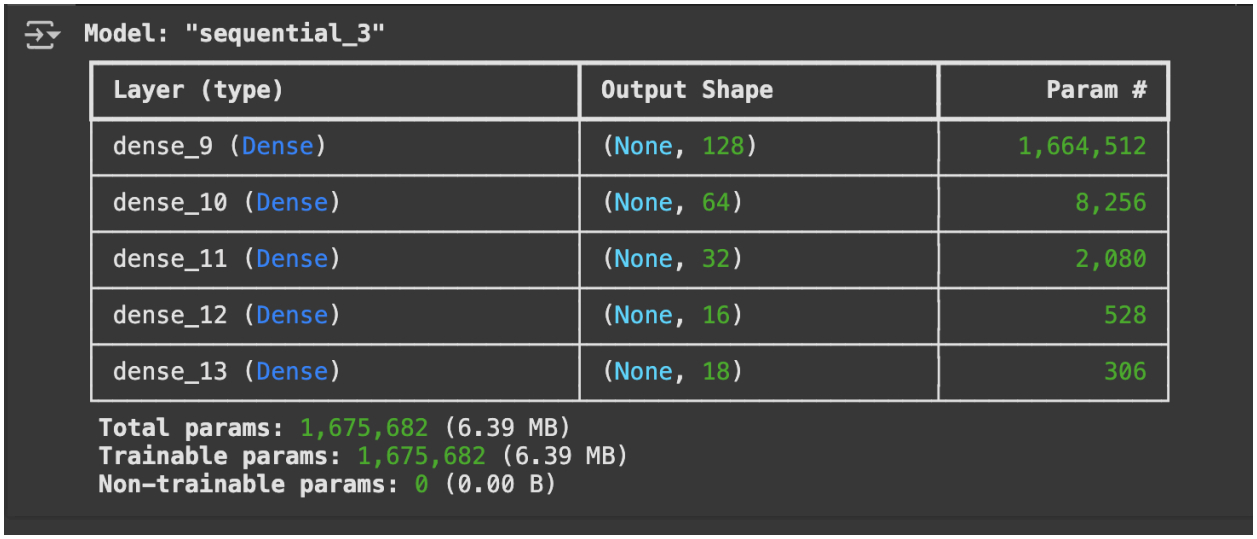
The performance metics for this initial model on both the training and test datasets are as follows:

Initial Model	Accuracy	Macro Avg F1-score	Weighted Avg F1-score
Training Data	0.87	0.87	0.88
Testing Data	0.27	0.27	0.29

There is a significant difference between the training and testing dataset of the model performance metics. While the model achieve 87% of high accuray rate on training data, its accuray rate extremely dropped to 27% on the unseen test data. This significant performance gap is a strong indicator of model overfitting.

New Model Architecture and Parameters Analysis

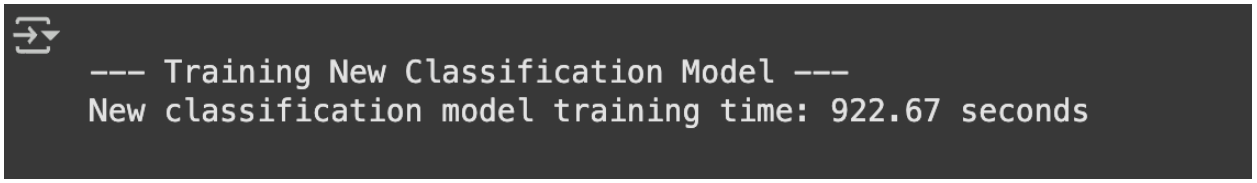
In the phase of new neural network model for this classification task include 5 layers that includes 4 hidden layers and 1 output layer. Four hidden layers with 128, 64, 32 and 16 neurons respectively both using the ReLU activation function. The total number of trainable parameter in this new model is 1,864,578. The model was trained for 50 epochs with a batch size of 32 and training time is 922.67 seconds. The training duration is significantly longer than the initial model's training time (281.18 seconds) which is likely increased number of layers and parameters.



Model: "sequential_3"

Layer (type)	Output Shape	Param #
dense_9 (Dense)	(None, 128)	1,664,512
dense_10 (Dense)	(None, 64)	8,256
dense_11 (Dense)	(None, 32)	2,080
dense_12 (Dense)	(None, 16)	528
dense_13 (Dense)	(None, 18)	306

Total params: 1,675,682 (6.39 MB)
Trainable params: 1,675,682 (6.39 MB)
Non-trainable params: 0 (0.00 B)



```
--- Training New Classification Model ---
New classification model training time: 922.67 seconds
```

The performance metics for this new model on both the training and test datasets are as follows:

New Model	Accuracy	Macro Avg F1-score	Weighted Avg F1-score
Training Data	0.86	0.87	0.86
Testing Data	0.27	0.27	0.29

Model Improvement Analysis

The following table summarizes the key performance metrics for both the initial and new classification models.

Performance Metics	Initial Model - Train	Inital Model - Test	New Model - Train	New Model- Test
Trainable Parameters	416,962	-	1,864,578	-
Training Time	281.18 seconds	-	922.67 seconds	-
Accuracy	0.87	0.27	0.86	0.27
Macro Avg F1-score	0.87	0.27	0.87	0.27
Weighted Avg F1-score	0.88	0.29	0.86	0.29

Adding the initial models with 3 layers to new model with 5 layers significantly increased the number of trainable parameters from 416,962 to 1,864,578. This increased complexity led to a major rise in training time from 281.18 seconds for initial model to 922.67 seconds for new model. Both of the initial and new models show persistent overfitting. The initial model achieved an 87% accuracy on training dataset but declined to 27% on the test dataset showing a poor generalization. The new model also indicated very high accuray rate 86% but continued a smililary low accuray rate 27% on test data. Despite of adding more layers and parameters, the new model did not show any significant improvement in test performance metrics compared to the initial model. Beside, its test accuracy and F1-scores are reamined identical to the initial model test data performance.

In conclusion, while the new model increased complexity and computational cost, it failed to reduce the overfitting issues present in the initial model. In future improvement should be focus on stronger regulaization techniques such as dropout, L1 (Lasso) /L2 (Ridge) regularization or early stopping to prevent the model from becoming too dependent on specific training examples and lead it to learn more robust features and reduction strategies to improve the model's generalization capabilities.