Peer Pair Group 17 Report

Water Quality Prediction Using Custom Models

Introduction

This report summarizes the work done by our team of three members in building and evaluating custom models to predict water potability using the Water Quality Dataset. Each member developed a unique model using different architectures, optimizers, and regularization techniques. Below, we provide a detailed summary of each model, the results obtained, and the challenges faced during the process.

Model 1: Regularized Deep Neural Network (DNN) with Adam Optimizer

Engineer Name: Geu Aguto Garang
Regularizer: L2 (0.01)
Optimizer: Adam
Early Stopping: Yes (patience=10)
Dropout Rate: 0.5

Model Architecture

- Input Layer: 128 neurons, ReLU activation, L2 regularization
- Hidden Layers: 64, 32, and 16 neurons with ReLU activation and L2 regularization
- Dropout: 0.3 after each hidden layer
- Output Layer: 1 neuron, Sigmoid activation

Training Summary

- **Training Accuracy:** 0.602
- **Testing Accuracy:** 0.624
- **F1 Score:** 0.3041
- **Recall:** 0.623984
- **Precision:** 0.389356

Results

- The model achieved moderate accuracy but struggled with precision and recall, particularly for the "Potable" class.
- The F1 score was low, indicating poor balance between precision and recall.

Challenges

- The model faced challenges in learning meaningful patterns due to the imbalanced nature of the dataset.
- Overfitting was observed despite the use of L2 regularization and dropout.

Model 2: Lightweight ANN with RMSprop Optimizer

Engineer Name: Abubakar Ahmed
Regularizer: None
Optimizer: RMSprop
Early Stopping: Yes (patience=5)
Dropout Rate: 0.5

Model Architecture

- Input Layer: 128 neurons, ReLU activation
- Hidden Layers: 64, 32, and 16 neurons with ReLU activation
- Dropout: 0.3 after each hidden layer
- Output Layer: 1 neuron, Sigmoid activation

Training Summary

- **Training Accuracy:** 0.708
- **Testing Accuracy:** 0.689
- **F1 Score:** 0.668695
- **Recall:** 0.695122
- **Precision:** 0.690427

Results

- This model performed better than the first, achieving higher accuracy and F1 score.
- The recall and precision were more balanced, indicating better generalization.

Challenges

- The model still struggled with the imbalanced dataset, but the use of RMSprop and dropout helped mitigate overfitting.
- Early stopping with a patience of 5 helped prevent overfitting but may have stopped training prematurely.

Model 3: CNN with Stochastic Gradient Descent (SGD) Optimization

Engineer Name: Peris Wangui
Regularizer: L2 (0.01)
Optimizer: SGD with Momentum
Early Stopping: Yes (patience=10)
Dropout Rate: 0.5

Model Architecture

- Conv1D Layers: 64 and 32 filters, kernel size 3, ReLU activation
- Batch Normalization after each Conv1D layer
- MaxPooling1D: Pool size 2
- Fully Connected Layers: 64, 32, and 16 neurons with ReLU activation
- Output Layer: 1 neuron, Sigmoid activation

Training Summary

- **Training Accuracy:** 0.6762
- **Testing Accuracy:** 0.473333
- **F1 Score:** 0.642534
- **Recall:** 1.0
- **Precision:** 0.473333

Results

- The model achieved high recall but very low precision, indicating a high number of false positives.
- The F1 score was moderate, but the overall accuracy was low.

Challenges

- The CNN architecture may not have been suitable for this tabular dataset, as it is typically used for sequential or image data.

- The model struggled with generalization, likely due to the dataset's characteristics.

Summary Table

Train Instance	Engineer Name	Regularizer	Optimizer	Early Stopping	Dropout Rate	Accuracy	F1 Score	Recall	Precision
Model 1	Geu Aguto Garang	L2 (0.01)	Adam	Yes (patience=10)	0.5	0.624	0.3041	0.623984	0.389356
Model 2	Abubakar Ahmed	None	RMSprop	Yes (patience=5)	0.5	0.689	0.668695	0.695122	0.690427
Model 3	Peris Wangui	L2 (0.01)	SGD with Momentum	Yes (patience=10)	0.5	0.473333	0.642534	1.0	0.473333

Insights and Challenges

1. **Dataset Imbalance:** All models struggled with the imbalanced nature of the dataset, particularly in predicting the minority class ("Potable").
2. **Overfitting:** Despite using regularization and dropout, overfitting was a common issue, especially in the DNN and CNN models.
3. **Model Suitability:** The CNN model, while innovative, was not well-suited for this tabular dataset, leading to poor performance.
4. **Optimizer Choice:** RMSprop performed better than Adam and SGD, likely due to its adaptive learning rate.
5. **Early Stopping:** Early stopping was effective in preventing overfitting but may have stopped training too early in some cases.

Model Comparsion and Best Model Identification

The Lightweight ANN by Abubakar Ahmed is the best-performing model based on the following metrics:

- **Accuracy:** 0.695 (highest among the three models).
- **F1 Score:** 0.669 (balanced precision and recall).

- **Precision and Recall:** Both metrics are relatively balanced, indicating better generalization.

Why it is the best:

- It achieved the highest accuracy and F1 score.
- The use of RMSprop and dropout provided a good balance between bias and variance.
- It demonstrated consistent improvement during training, suggesting effective learning.

Conclusion

- **Best Performing Model:** The lightweight ANN with RMSprop optimizer (Model 2) achieved the highest accuracy and F1 score, making it the most effective model for this dataset.
- **Recommendations:** Future work could focus on addressing dataset imbalance using techniques like SMOTE or class weighting. Additionally, experimenting with other architectures like Gradient Boosting Machines (GBMs) or ensemble methods could yield better results.
- **Team Collaboration:** Each team member contributed unique insights and approaches, highlighting the importance of experimentation and collaboration in machine learning projects.

Each model had its strengths and weaknesses, but the Lightweight ANN by Abubakar Ahmed emerged as the best performer. Future work could focus on addressing class imbalance, experimenting with more advanced architectures (e.g., Gradient Boosting Machines or ensemble methods), and performing hyperparameter tuning to further improve performance.