Logo

Description automatically generated

**College of Professional Studies**

**Northeastern University San Jose**

**MPS Analytics**

**Course: ALY6020 – Predictive Analytics**

**Assignment:**

Module 5 Text Classification

**Submitted on:**

Feb 14th, 2024

**Submitted to:**  **Submitted by:**

Professor: BEHZAD AHMADI ARCHIT BARUA

**Introduction**

In the modern educational landscape, early identification of students who may require additional support in developing their motor skills is paramount. This paper explores the feasibility of utilizing handwriting tests as a diagnostic tool to assess motor skills in young students. By analyzing handwriting samples, specifically the way students write numbers, we aim to discern patterns that could indicate the need for intervention. Our approach employs two distinct machine learning models: a simplistic K-Nearest Neighbors (KNN) model and a more complex neural network. This comparative study seeks to determine the efficacy of these models in predicting handwriting characteristics, thereby offering insights into students' motor skill development.

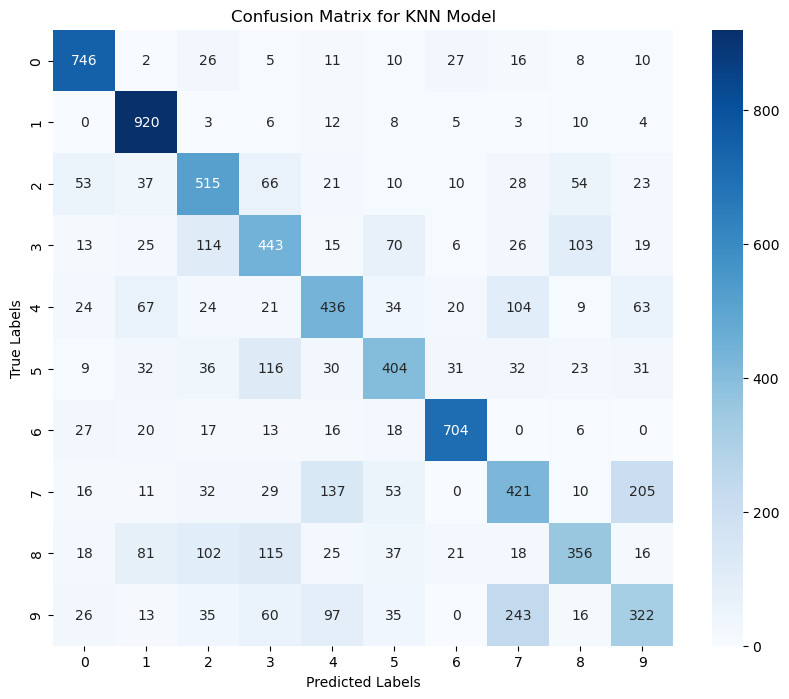
**Part 1: KNN Model for Handwriting Prediction**

The initial phase of our research involved the implementation of a KNN model to classify handwritten numbers. The KNN algorithm, known for its simplicity and effectiveness in pattern recognition tasks, was chosen as a starting point for its interpretability and ease of use.

Methodology: The model was trained using a dataset comprising handwritten digits, each represented by a set of pixel values. The dataset was divided into training and testing subsets to evaluate the model's performance.

Results: The KNN model achieved an overall accuracy of 62.70%. While certain classes (digits) were predicted with high precision and recall, others demonstrated lower scores, indicating variability in the model's ability to generalize across different handwriting styles.

Challenges: The primary challenges encountered with the KNN model included its sensitivity to the choice of 'k' value, the computational intensity due to the need to compute distances between all pairs of points, and the model's struggle with high-dimensional data, leading to the "curse of dimensionality."



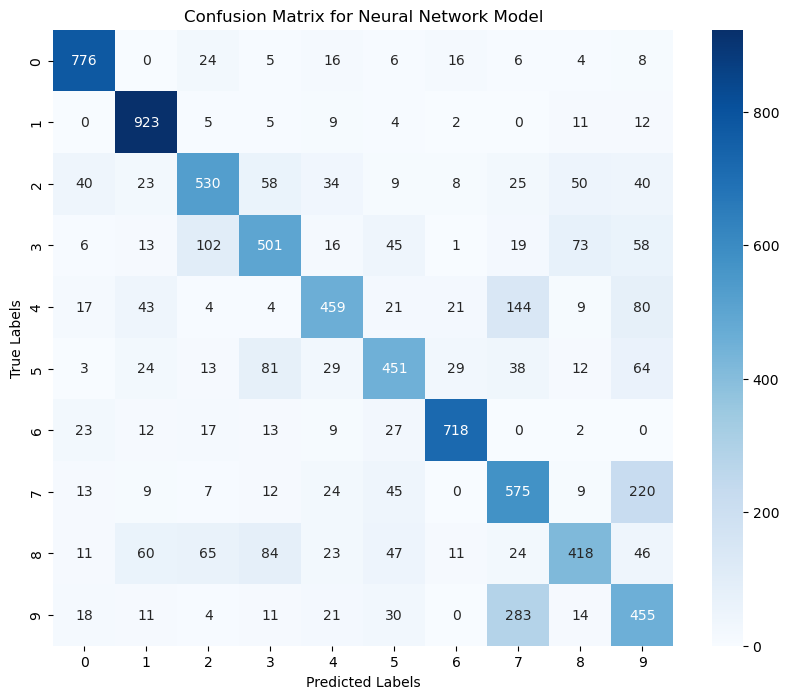
**Part 2: Neural Network Model for Handwriting Prediction**

To explore the capabilities of a more sophisticated model, we implemented a neural network, known for its prowess in handling complex pattern recognition tasks, especially in image and handwriting analysis.

Methodology: The neural network consisted of multiple layers, including input, hidden, and output layers, designed to learn the intricate patterns in handwriting data. The data was preprocessed, scaled, and converted into a format suitable for neural network training.

Results: The neural network model demonstrated a notable improvement, with an overall accuracy of 69.12%. The model showed enhanced precision and recall across most classes, indicating a better grasp of the diverse handwriting styles present in the dataset.

Challenges: The complexities of designing and training a neural network were evident. These included determining the optimal architecture, preventing overfitting through regularization techniques, and the increased computational resources required for training and inference.



**Part 3: Comparative Analysis and Model Recommendation**

In this section, we perform a comprehensive comparison between the K-Nearest Neighbors (KNN) model and the Neural Network model using detailed performance metrics. This analysis aims to identify the most suitable model for the school's handwriting recognition task, considering the ability to accurately predict handwritten digits as an indicator of students' motor skill development.

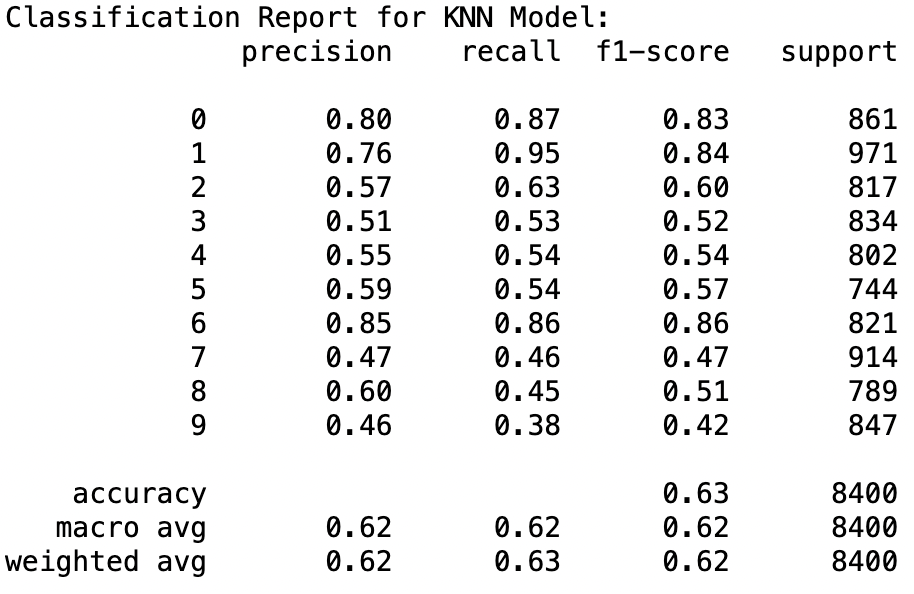
KNN Model Performance:

The KNN model exhibited varying degrees of success across different digit classes. Notable observations include:

High Precision and Recall for digits '0' and '1', with F1-scores of 0.83 and 0.84, respectively, indicating strong performance in these categories.

Moderate to Low performance for digits '8' and '9', with F1-scores of 0.51 and 0.42, suggesting difficulties in accurately classifying these digits.

The overall accuracy of the KNN model stood at 62.70%, reflecting a decent baseline capability in handwriting recognition.



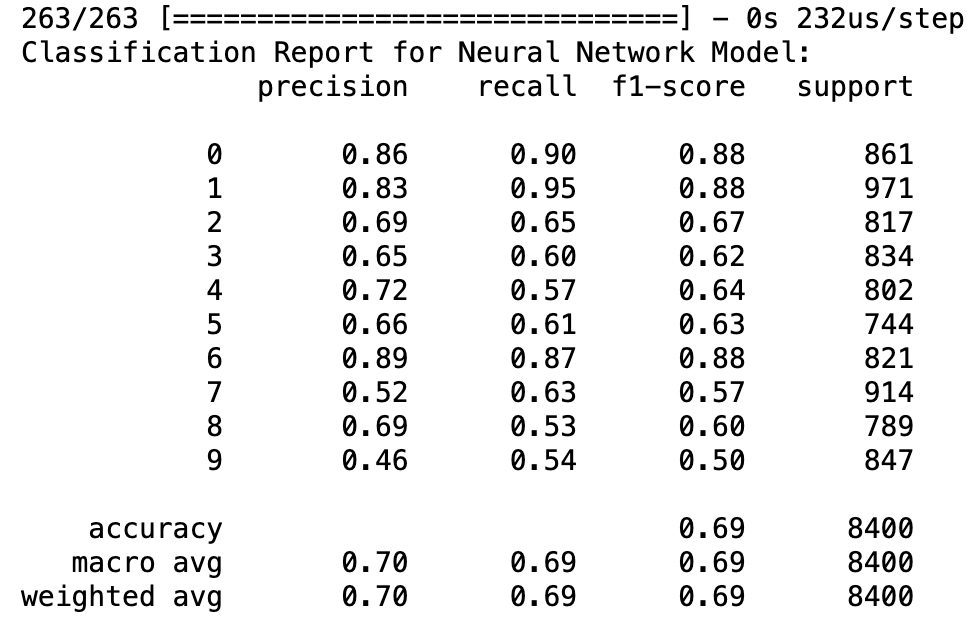
Neural Network Model Performance:

The Neural Network model demonstrated improved and more consistent performance across all classes:

Enhanced Precision and Recall for digits '0' and '1', with F1-scores reaching 0.88 for both, showcasing the model's effectiveness in recognizing these digits.

Significant improvements were observed in traditionally challenging classes for KNN, such as '8' and '9', with the Neural Network achieving F1-scores of 0.60 and 0.50, respectively.

The overall accuracy of the Neural Network model was 69.12%, indicating a substantial enhancement in predictive capability over the KNN model.



Comparative Analysis:

Accuracy: The Neural Network model outperforms the KNN model with a higher overall accuracy (69.12% vs. 62.70%), indicating better generalizability and effectiveness in digit recognition.

Consistency Across Classes: The Neural Network shows more balanced performance across different digit classes, evidenced by higher and more uniform F1-scores, suggesting a better ability to handle the diversity in handwriting styles.

Precision and Recall: The Neural Network model generally exhibits higher precision and recall values, indicating a lower rate of false positives and a higher success rate in identifying true positives across digit classes.

Challenges and Considerations:

The KNN model, while simpler and less resource-intensive, shows limitations in handling complex patterns and variability in handwriting, leading to lower accuracy and imbalanced performance across classes.

The Neural Network model, despite its higher computational demand and complexity, offers superior performance, making it a more effective tool for this application.

Recommendation:

Considering the detailed analysis and the superior performance of the Neural Network model across key metrics, we recommend its adoption for the school's handwriting recognition task. The Neural Network's ability to accurately classify digits, even in challenging categories, makes it a valuable asset for identifying students who may require additional support in developing their motor skills. The investment in computational resources and expertise required to implement and maintain the Neural Network model is justified by its potential to enhance early educational interventions and support tailored learning experiences for students.

This recommendation is made with the understanding that early and accurate identification of developmental needs can significantly impact students' educational outcomes, and the Neural Network model's advanced capabilities align well with this objective.

**References**

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: with applications in R. Springer.

Plamondon, R., & Srihari, S. N. (2000). Online and off-line handwriting recognition: a comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(1), 63–84. <https://doi.org/10.1109/34.824821>