Visualization, Key Findings and Conclusion

Project Module 5

Areej Alnamlah – Nedaa Abuhussein – Safa Nasser

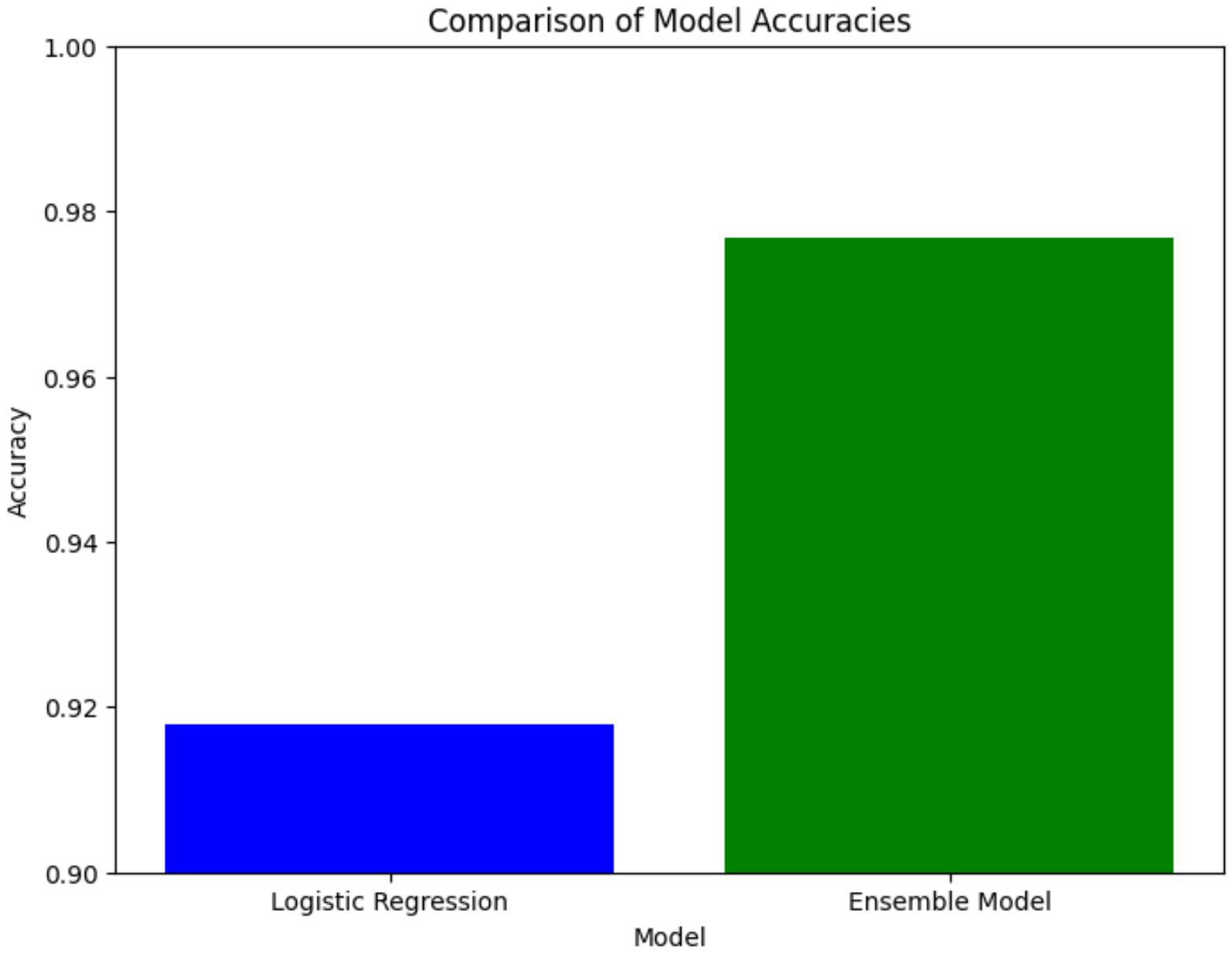
Process Overview

The project involved several key stages including data preprocessing, model implementation, hyperparameter tuning, and the use of various visualization techniques to enhance the understanding and presentation of the results. Each stage played a crucial role in ensuring the accuracy and efficiency of the machine learning models applied.

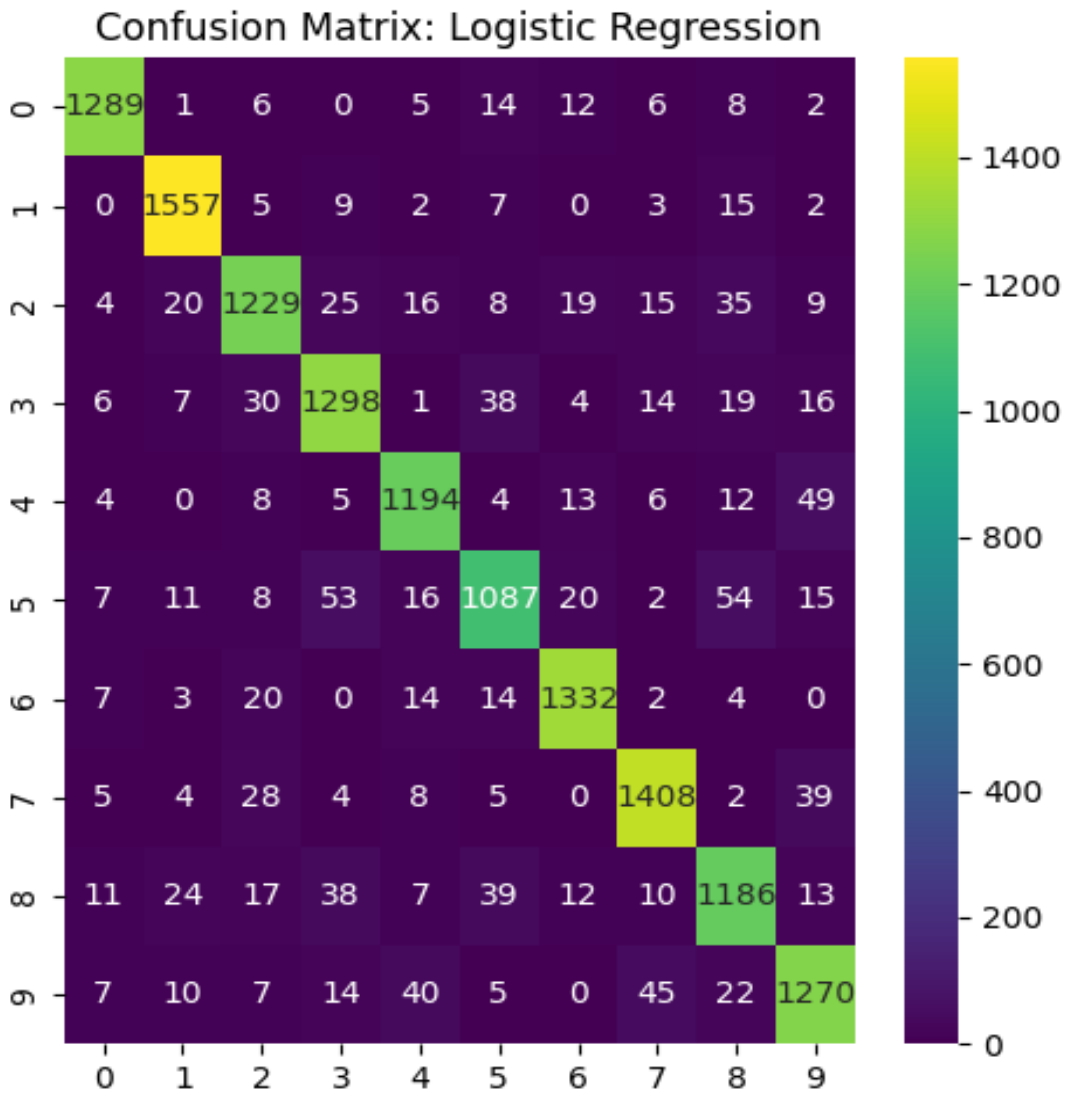
Process Overview

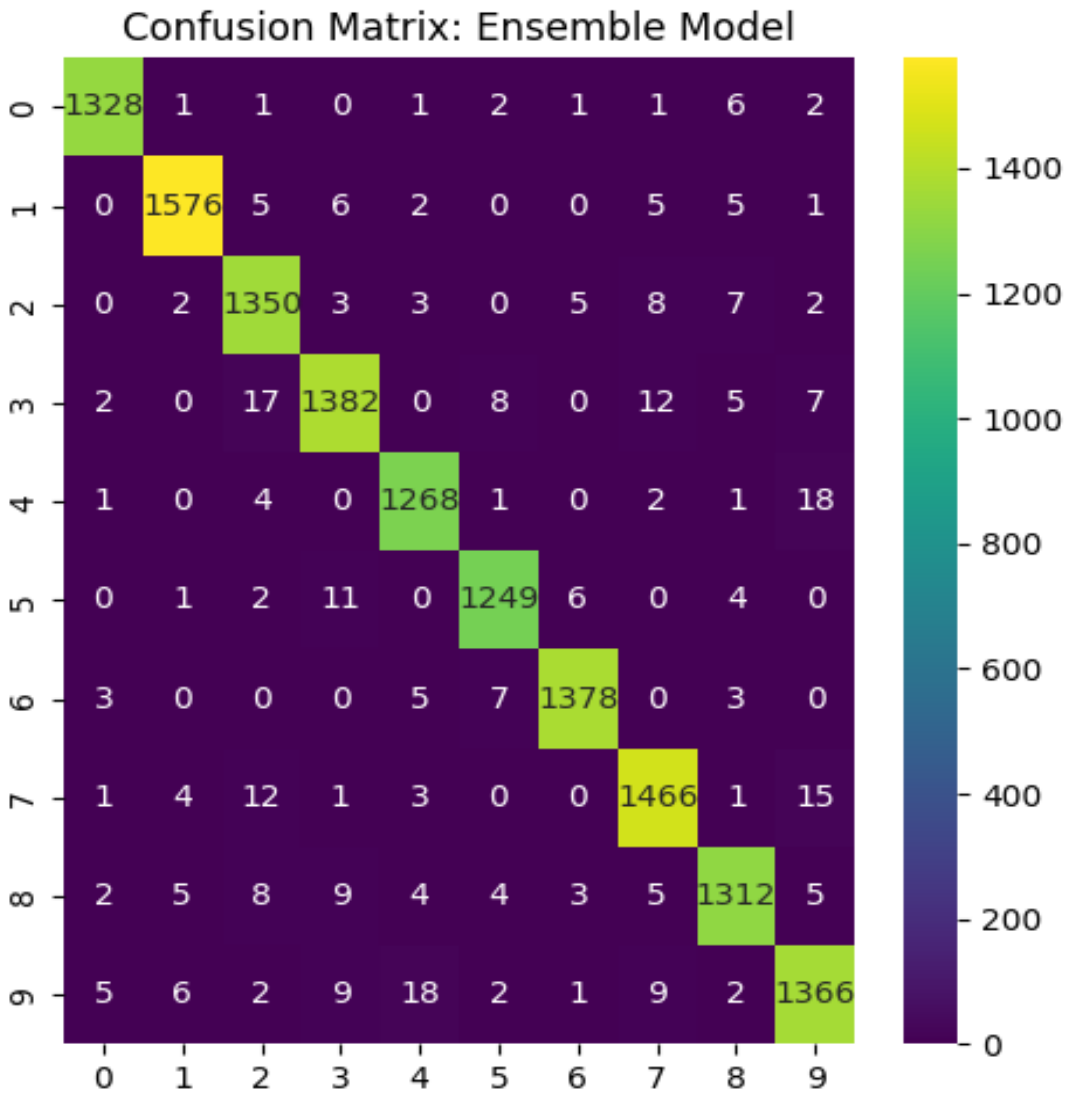
- **Data Preprocessing:** The initial stage involved cleaning the data, dealing with missing values, and normalizing the data to ensure that the models would perform optimally.
- **Model Implementation:** Multiple machine learning models were implemented, including both traditional algorithms and neural network approaches. This allowed for a comprehensive analysis of the dataset from different perspectives.
- **Hyperparameter Tuning:** Extensive hyperparameter tuning was conducted, particularly for the neural network models, to optimize performance and accuracy. Techniques such as grid search and random search were utilized to find the best parameters.
- **Visualization:** Various visualization tools were employed to illustrate the data distribution, model performance, and comparison between different models. These visualizations were critical for interpreting the complex model outputs and making the data accessible to stakeholders.

Visualization ML

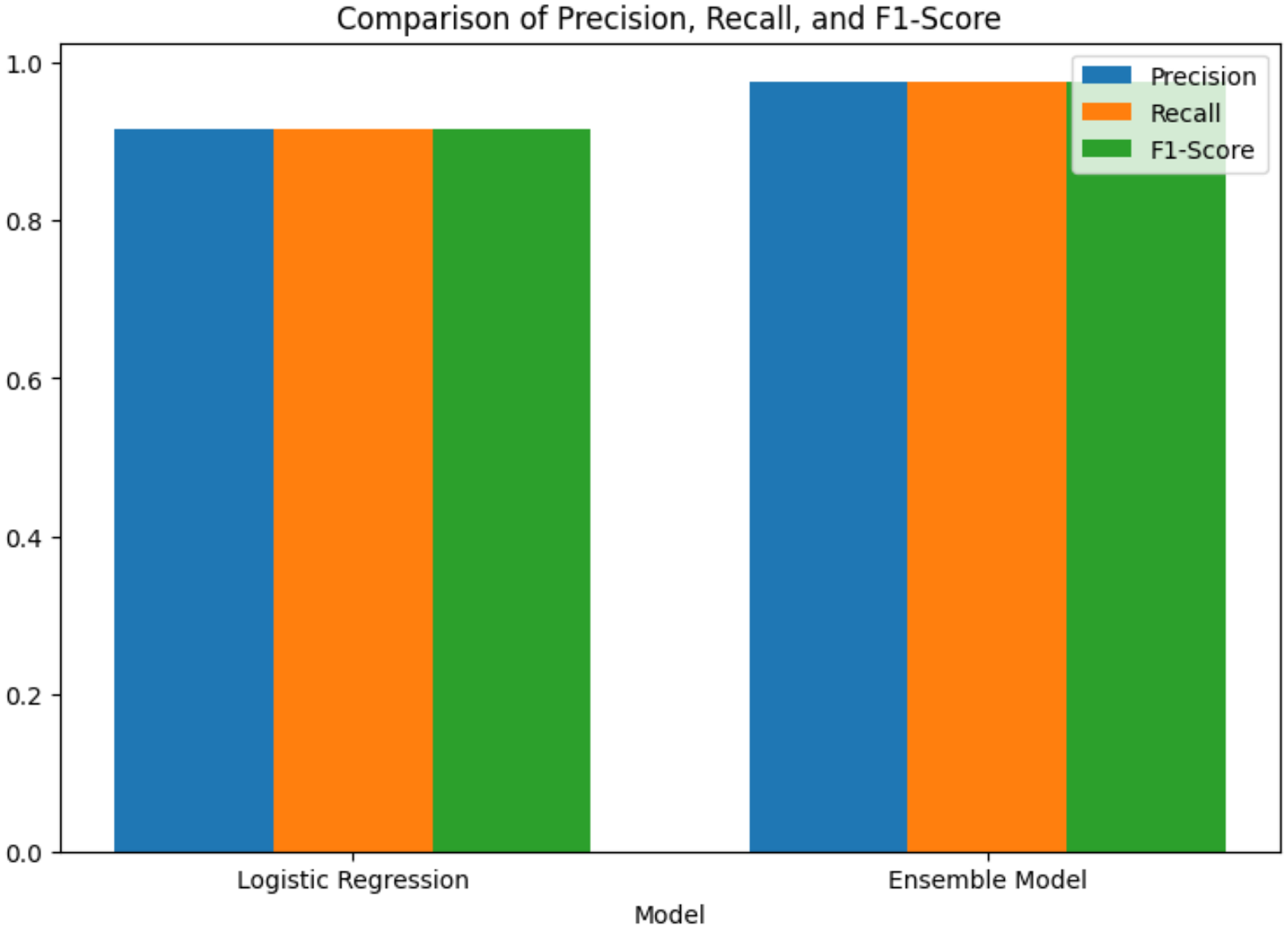


Visualization ML

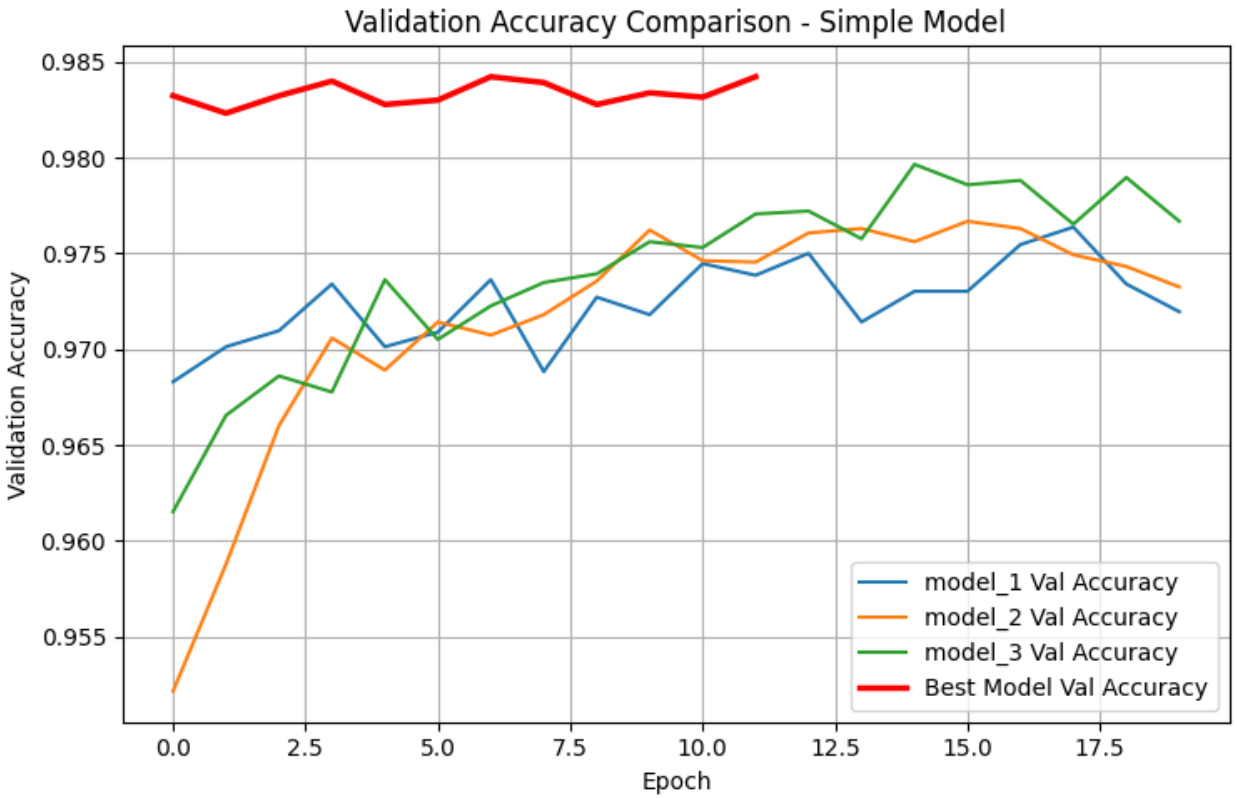


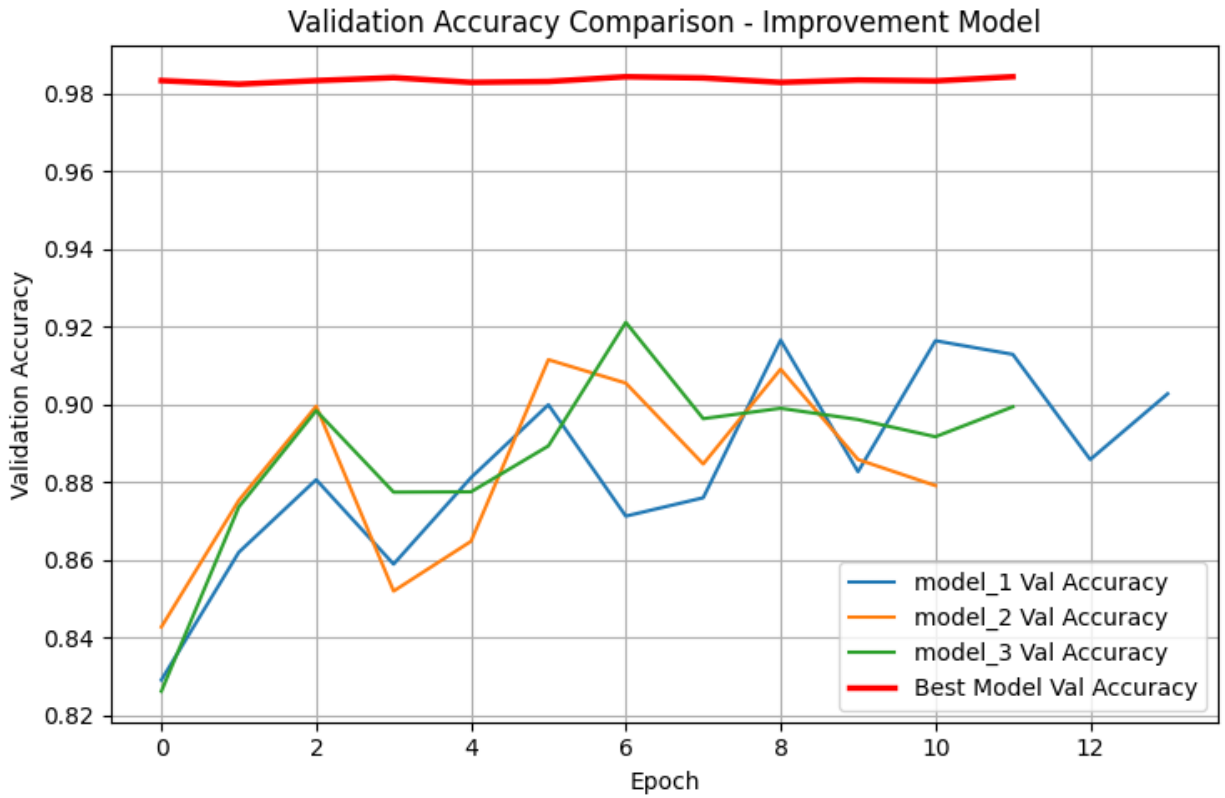


Visualization ML

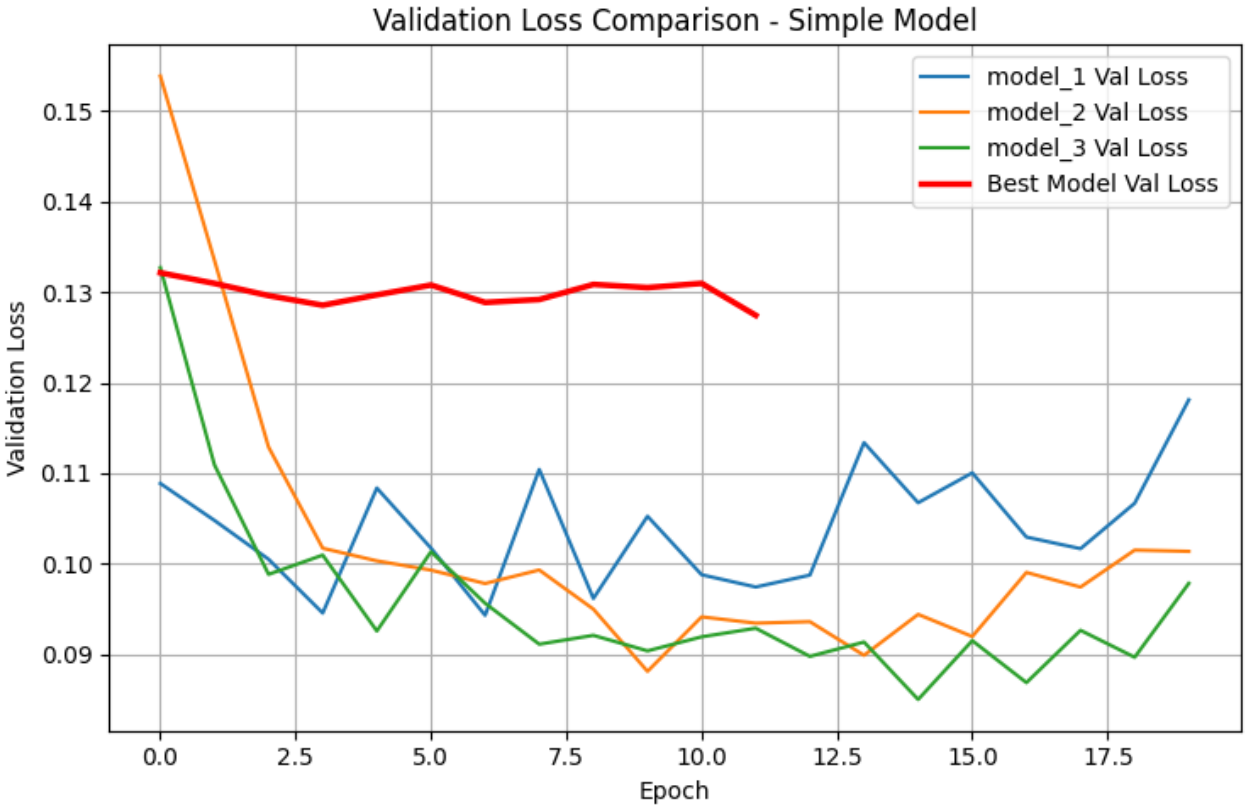


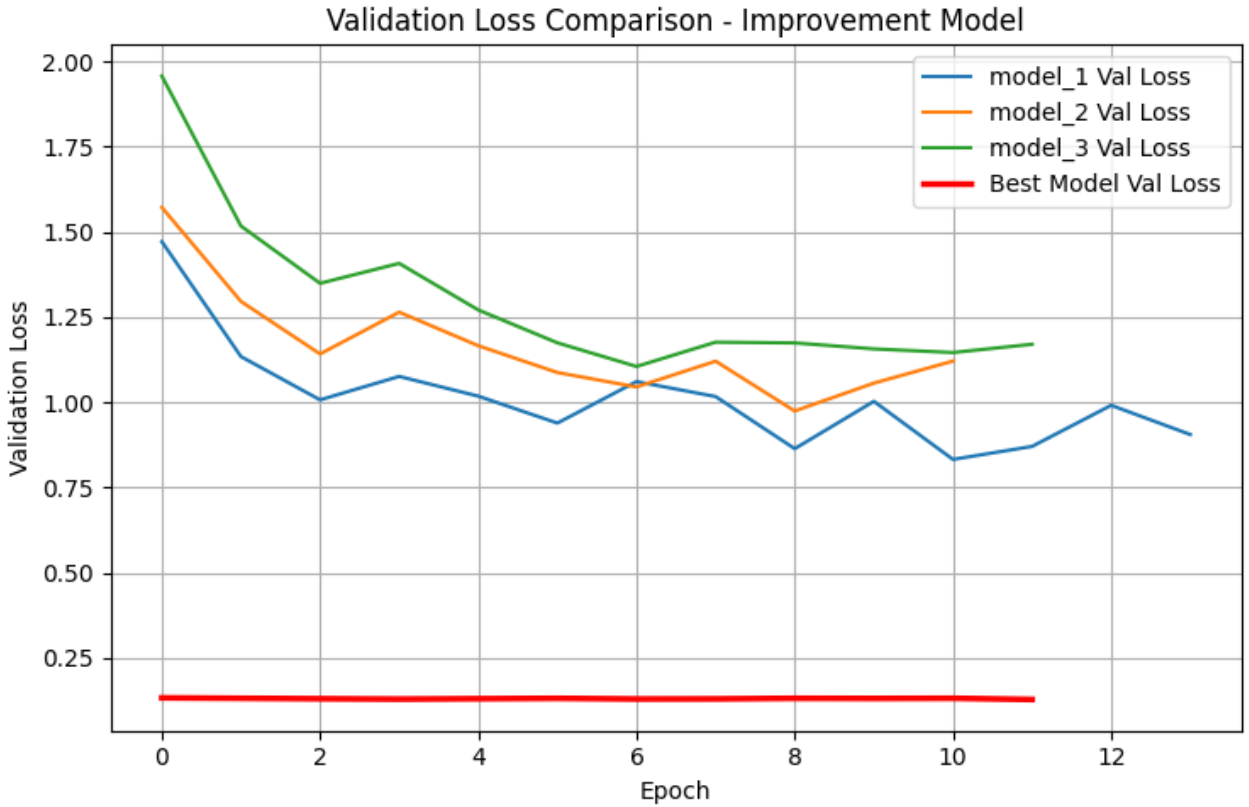
Visualization NN – acc





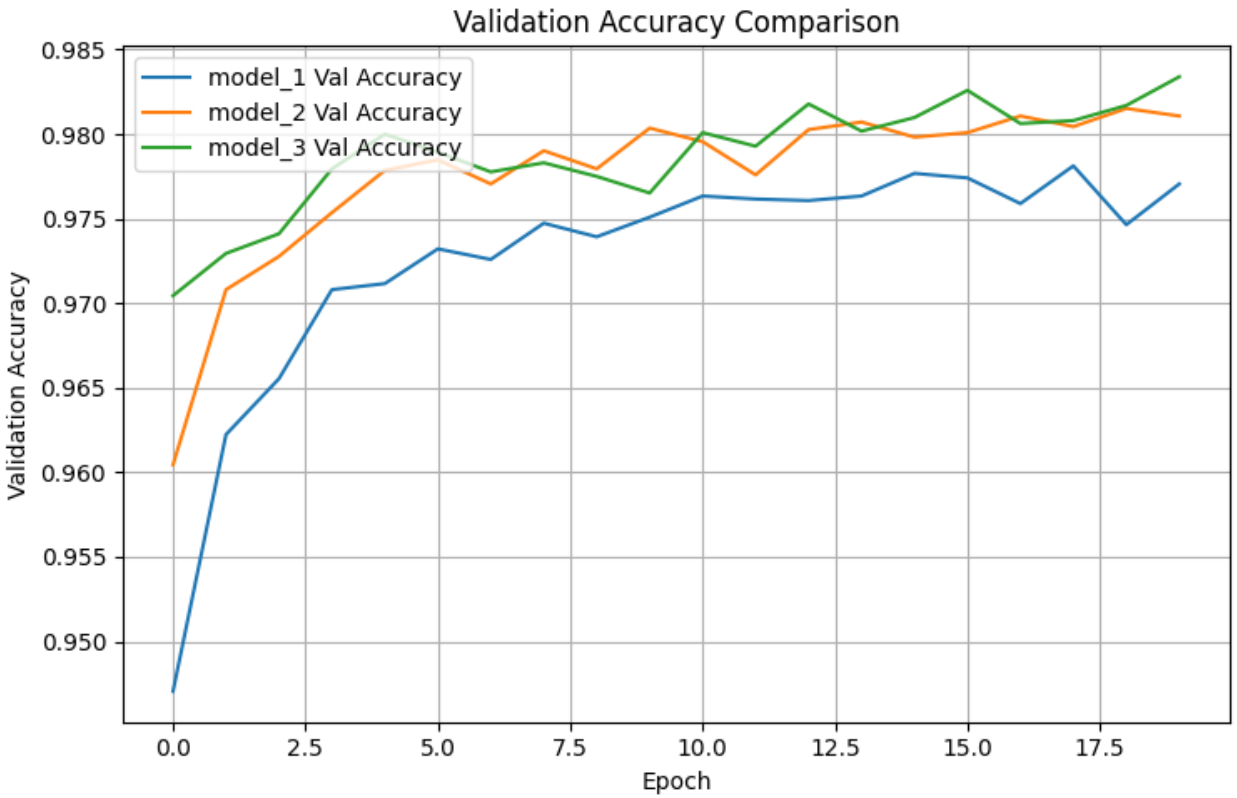
Visualization NN – loss

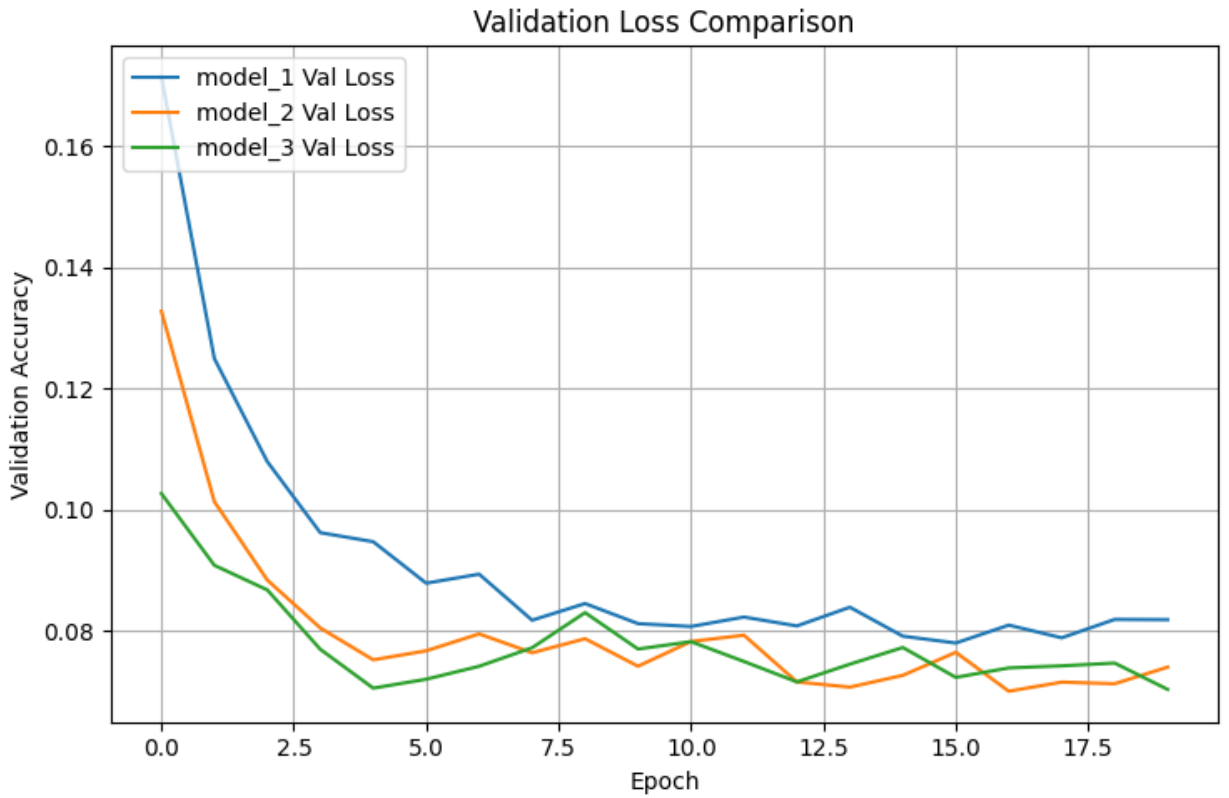




Visualization NN – train vs valid

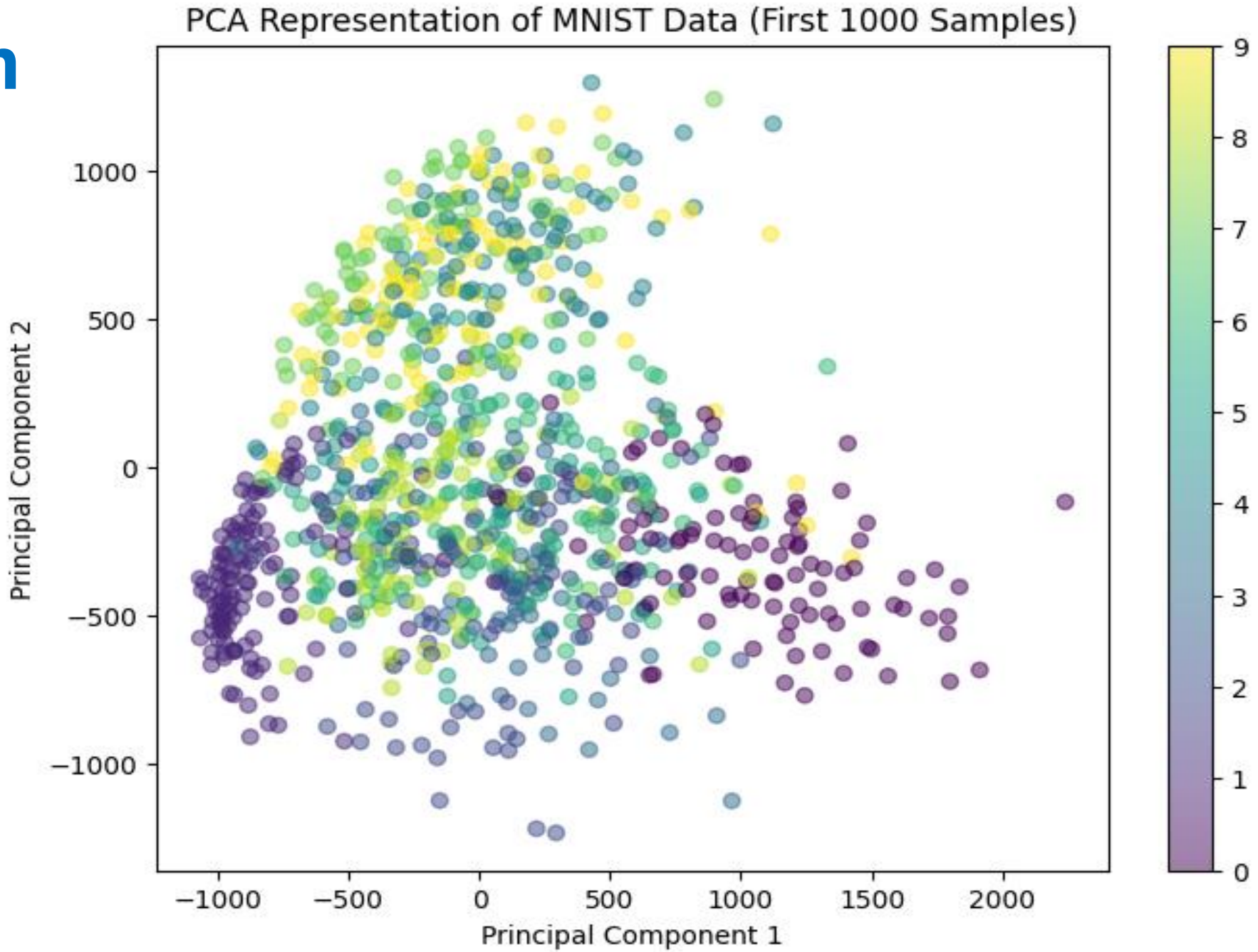
Visualization PCA with NN



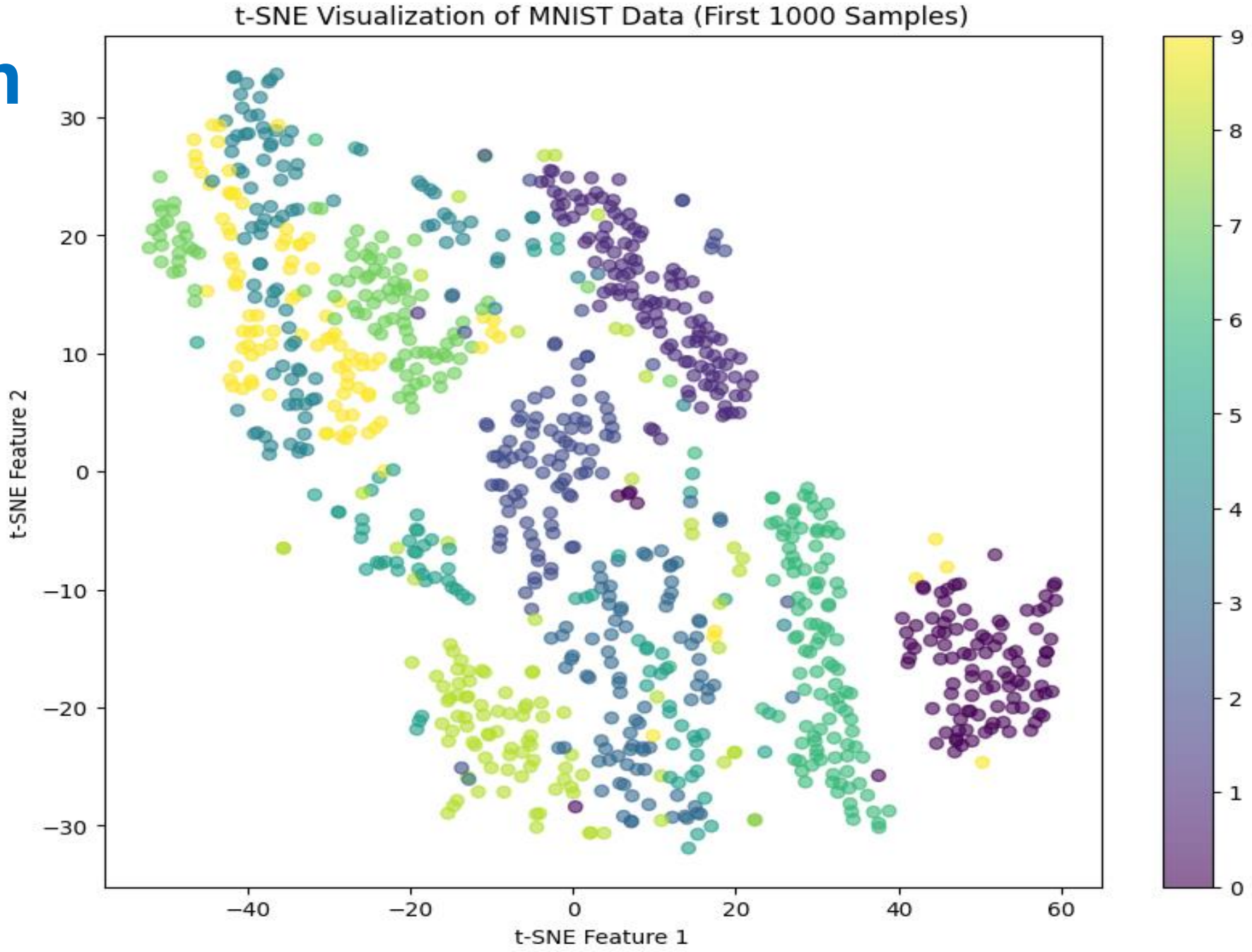



```
Best validation accuracy of 0.9834
In Model # 3
```

Visualization
PCA



Visualization
t-SEN



Key Findings and Insights

1. The **neural network model**, specifically *(best model)* witch configuration detailed in the second notebook *in building model section*, provided the best performance in terms of accuracy and efficiency. This model benefitted significantly from deep layers and a carefully tuned learning rate.

Key Findings and Insights

2. **Principal Component Analysis (PCA)** was effective in reducing dimensionality and enhancing model performance by focusing on the most informative features.

Key Findings and Insights

3. Comparative analysis across models revealed that while traditional algorithms were faster to implement and provided a good baseline, **deep learning models excelled** in handling large and complex datasets with higher accuracy.

Lessons Learned

1. The **importance of** comprehensive **data preprocessing** was reaffirmed, as it directly influences the success of the subsequent modeling stages.

Lessons Learned

2. **Hyperparameter tuning**, although time-consuming, is vital for achieving the best model performance.

Lessons Learned

3. **Visualization** not only aids in interpreting results but also in communicating findings to non-technical stakeholders, making it an indispensable part of any data science project.

Conclusion

The project demonstrated the effectiveness of a structured machine learning workflow, incorporating sophisticated techniques and tools to tackle the challenges presented by complex datasets. The insights gained from this project are expected to guide future projects and strategies in similar contexts.