Machine Learning and Pattern Recognition

CA Two Individual Report

Student Number as per your student card:

Suprith Sreekanta Swamy – 20016392

Course Title: Data Analytics

Lecturer Name: Satya Prakash

Module/Subject Title: Machine Learning and Pattern Recognition (B9DA109)

Assignment Title: Task 2, Task 3

In Task 2 of our group assignment, my assigned task was to delve into data augmentation. Commencing with the development of a model for fine-tuning, I did experimentation with various optimizers. After doing multiple layers into the model, my initial inclination was towards the Stochastic Gradient Descent (SGD) optimizer. Upon partitioning the data into training sets, I observed good accuracy as the epochs progressed. However, after the visualized data left me dissatisfied, prompting me to explore an alternative optimizer 'Adam.'

Upon applying the 'Adam' optimizer and undergoing further training, the outcomes revealed a higher accuracy of 96.15% accompanied by a reduced validation loss. In contrast, SGD exhibited a similar pattern but with a higher validation loss. Consequently, the decision was made to proceed with the 'Adam' optimizer. In my conclusive analysis, I posited that the test accuracy stood at 98.15%, while the validation accuracy reached 92.31%, suggesting a difference. The lowest validation loss, recorded at epoch 4, was 0.1983. Crucially, the graphical representations illustrated minimal discrepancies between training and validation, indicative of an absence of overfitting.

To address potential overfitting, I implemented data augmentation, introducing further enhancements to the model. Incorporating 'Adam' optimizer in fine-tuning, I achieved great results. Notably, I introduced dropouts and extended the number of epochs. The integration of early

stopping proved to be an effective strategy, ensuring optimal accuracy post-dropouts and 'Adam' optimization.

In task 3, I split data into dependent and independent variables. The task involved coding to extract the top 15 words from each genre and creating word clouds. Aggregating all poems into a unified string, I undertook the removal of stop words and leveraged frequency distribution analysis. Employing common words, I identified and compiled the top 15 words for each genre, culminating in the creation of distinct word clouds for each Genre.

In essence, the journey encompassed model refinement, optimizer exploration, meticulous analysis of accuracy metrics, and the creative exploration of textual data through word cloud generation. The collaborative effort resulted in a comprehensive understanding of the interplay between model architecture, optimization techniques, and insightful data representation.