**PROJECT REPORT**

BUILDING NEURAL NETWORKS – IMDB DATASET

INTRODUCTION

This assignment gives an overview of IMDb Movie Reviews which is a binary sentiment analysis dataset made up of 50,000 reviews from IMDb that have been classified as positive or negative. The dataset includes an equal number of positive and negative reviews.

We've built a total of five models.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model Number | Layers | Activation | Nodes | Optimizer | Loss | Test  Accuracy |
| Model-1 | 3 | tanh | 16 | Rms prop | mse | 0.8824 |
| Model-2 | 2 | tanh | 16 | Rms prop | mse | 0.8821 |
| Model-3 | 3 | tanh | 16 | Rms prop | mse | 0.8799 |
| Model-4 | 2 | tanh | 32 | Rms prop | mse | 0.8632 |
| Model-5 | 2 | tanh | 64 | Rms prop | mse | 0.8684 |

Graphical user interface, application

Description automatically generated

* The "tanh" activation function and "mse" loss function have been used most frequently as instructed in the models discussed above.
* Model 1 has the highest accuracy, with a test accuracy of 88.24%, while the other models have slightly lower accuracy.
* As we increase the unit from 16, 32 to 64,  we find that accuracy varies very little while the loss function shrinks.
* Raising the number of hidden layers from one to three has no impact on the precision or deflection of the loss function.
* When using binary cross entropy as the benchmark for the IMDB dataset, the loss value using "mse" is negligible.
* The vanishing gradient issue affects the model's tanh activation function.
* When compared to the initial model, regularization produces much lower losses, less overfitting, and a somewhat more accurate model.
* Dropout does not influence the accuracy, however it does lessen the loss function.

**CONCLUSION**

The best model with the **highest accuracy** rate is **Model-1** i.e. 88.24%. Here the activation function used is “tanh”, loss function is “mse” and the optimizer is rms prop.