Project Summary

By Vaatsalya Babbar

Dataset Details:

- **Dataset**: Collection of image data for resume and non-resume categories
- **Size**: Total of 124 images(62 resumes and 62 non-resumes)
- **Features**: Image pixels, categorised into classes (resume/non-resume).
- **Preprocessing**: Image normalisation, resizing to (256x256), and augmentation.

Model Architecture:

- Model Type: Convolutional Neural Network (CNN).
- **Architecture:** Custom CNN with 3 convolutional layers, 2 max-pooling layers, and 2 fully connected layers.
- **Activation:** ReLU activation for hidden layers, Sigmoid for output.
- **Loss Function:** Categorical Cross-Entropy.
- **Optimiser:** Adam optimiser.

Training Strategy:

- **Train-Validation Split:** 80-20 split for training and validation.
- **Batch Size:** Batch size of 16 because of a small Dataset.
- **Data Augmentation:** Applied augmentation techniques (rotation, horizontal flip) to increase dataset diversity.
- **Callbacks:** Early stopping, model checkpointing.

Evaluation Metrics

- **Accuracy:** Overall model accuracy on validation set.
- **Confusion Matrix:** Visualised model performance using a confusion matrix to analyse true positive, true negative, false positive, and false negative rates.

Conclusion

- Achieved an accuracy of 85% on the validation set.
- Confusion matrix revealed higher accuracy in identifying non-resumes but slightly lower in resumes due to dataset imbalance.
- The model demonstrates potential in classifying resume and non-resume images, with room for improvement through dataset expansion and fine-tuning.