# Computer Vision Intern Assignment: Resume Classifier

Approach Summary:

The Resume Classifier is a machine learning model designed for automated resume classification. The approach involves creating a custom dataset with a curated collection of resume and non-resume images, sourced from platforms like Kaggle and Google.

Dataset Details:

The dataset comprises around 2000 images, carefully selected to ensure diversity and relevance. Resumes and non-resume images are sourced from Kaggle and various Google repositories, contributing to a comprehensive training set. The dataset is balanced to address potential class imbalances, enhancing the model's ability to generalize across different resume types.

Model Architecture:

The model architecture is based on a CNN, featuring two convolutional layers with max-pooling, ReLU activation, and a fully connected layer for classification. The choice of architecture is motivated by the need to capture intricate features in resume images. The model is scalable and can handle images of varying sizes, assuming an input image size of 224x224.

Training Strategy:

Training involves fine-tuning the model with an emphasis on optimizing hyperparameters. Data augmentation techniques are employed to augment the dataset and improve the model's ability to generalize. The training process is carefully monitored to ensure efficient convergence, and regularization techniques like dropout are implemented to prevent overfitting.

Evaluation Metrics:

The model's performance is evaluated using a range of metrics, including accuracy, precision, recall, and F1 score. Confusion matrices are employed to provide a detailed breakdown of classification results, offering insights into the model's strengths and weaknesses. This comprehensive evaluation approach ensures a thorough understanding of the model's effectiveness in resume classification.

Confusion Matrix:

The confusion matrix is a critical evaluation tool in classification tasks, providing a detailed breakdown of the model's predictions.

- True Positives (TP) represent instances correctly classified as positive, indicating the model's ability to accurately identify relevant features.
- True Negatives (TN) are instances correctly classified as negative, showcasing the model's proficiency in discerning non-relevant elements.
- False Positives (FP) occur when the model incorrectly classifies a negative instance as positive, emphasizing potential areas of improvement in precision.
- False Negatives (FN) signify negative instances misclassified as positive, highlighting potential enhancements in recall

# **Highlights:**

The Resume Classifier Machine learning model provides a robust solution for automating resume classification:

*CNN Architecture:* Based on a Convolutional Neural Network for effective image-based classification.

*Custom Dataset:* Curated dataset of 2000 images from Kaggle and Google for diversity and relevance.

*Optimized Training*: Fine-tuned model with data augmentation and regularization for efficiency.

*Confusion Matrix Analysis*: Comprehensive evaluation using confusion matrices for detailed insights.

*Scalability*: Model designed to handle varying image sizes for real-world flexibility.

*Documentation*: Well-documented repository for easy setup, training, and integration into different environments.

*Deployment Ready*: Considerations for deployment environments, making it suitable for production systems.

*Community Collaboration:* Encourages contributions for future enhancements and improvements.