AUTOMATED JOB CANDIDATE SCREENING USING NLP

1) Our project is about making the hiring process faster and more efficient using Natural Language Processing (NLP). The program will read and analyze resumes, looking for important keywords and relevant details based on the job description. It will then rank candidates to help recruiters focus on the best matches. This saves time, reduces bias, and makes the process easier for everyone. Although it's just for learning, the project shows how AI can make hiring fairer and faster.

2) 1. Collector: Snehaan Bhawal
   2. Year: 2021
   3. Title of Dataset: Resume Dataset
   4. Version Number: Version 1
   5. Publisher: Kaggle
   6. URL: https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset
   7. Study/Paper/Reason: The dataset comprises resume examples sourced from livecareer.com, used for categorizing resumes into predefined lables.

3) Programming Language: Python 3.11
   Libraries:
   - Pandas, NumPy, Matplotlib, Scikit-learn, NLTK

4) I am implementing the entire classification model from scratch, without using any libraries for classification. This includes Text preprocessing, building the classifier from scratch and training the model. I am only using libraries for basic utilities like NumPy for matrix operations and Pandas for handling the dataset

5) For this project, I believe Logistic Regression**,** Support Vector Machines (SVM)**,** and Naive Bayes are strong choices for traditional machine learning, while BERT or other transformer-based models are ideal for advanced NLP.

6) The hyperparameters depend on the chosen model. For Logistic Regression, key hyperparameters include C(regularization strength) and solver, while for SVM, important ones are C, kernel, and gamma. I'll optimize these using Grid Search or Randomized Search with cross-validation to test different combinations and choose the best based on F1-score or accuracy.

7) I will evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score to assess its ability to classify resumes accurately and handle imbalanced classes. I'll also use a confusion matrix to visualize the model's true positives, false positives, false negatives, and true negatives.

8) https://github.com/Anusha2207Reddy/aerva-b00118232-spring-2025