

**FINAL REPORT FOR THE**

**SUMMER INTERNSHIP PROGRAMME,2025**

1. TITLE **:**  AI powered resume screening system
2. STUDENT NAME **: Khyanrick Borah** , **Syed Jeeshan Javed,**
3. **Nilim Kr. Gogoi**
4. SAMARTH ENROLMENT No. **:** 232010007027 , 232010007057, 242050007009
5. PROGRAM AND SEMESTER **:** B.Tech , 4th Sem
6. DEPARTMENT : Computer Science And Engineering
7. NAME OF MENTOR **: Dr. Purabi Sharma**
8. **Dr. Anjan Kumar Talukdar**
9. INTERNSHIP TYPE

(SKILL/RESEARCH) **:** Skill Based Internship

1. INTERNSHIP PROVIDING

ORGANISATION

/LAB/DEPARTMENT **:** DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING, GAUHATI UNIVERSITY

1. INTERNSHIP SUPERVISOR

(NAME AND DESIGNATION) **: Dr. Surajit Deka**

1. INTERNSHIP PERIOD

(GIVE DATES) **:** July 01, 2025 – July 31, 2025

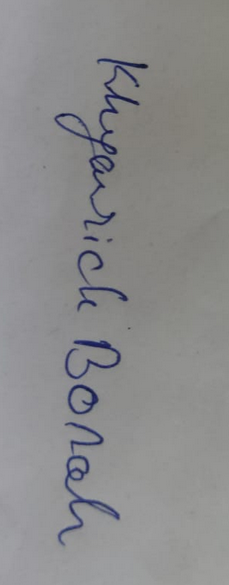
**Declaration**

I hereby declare that the project report titled "ML Based Student Grade Evaluation" submitted in partial fulfillment of the requirements for the award of the Summer Internship is a record of original work carried out by me under the guidance of Dr. Anjan Kumar Talukar and Mr. Angarag Das.

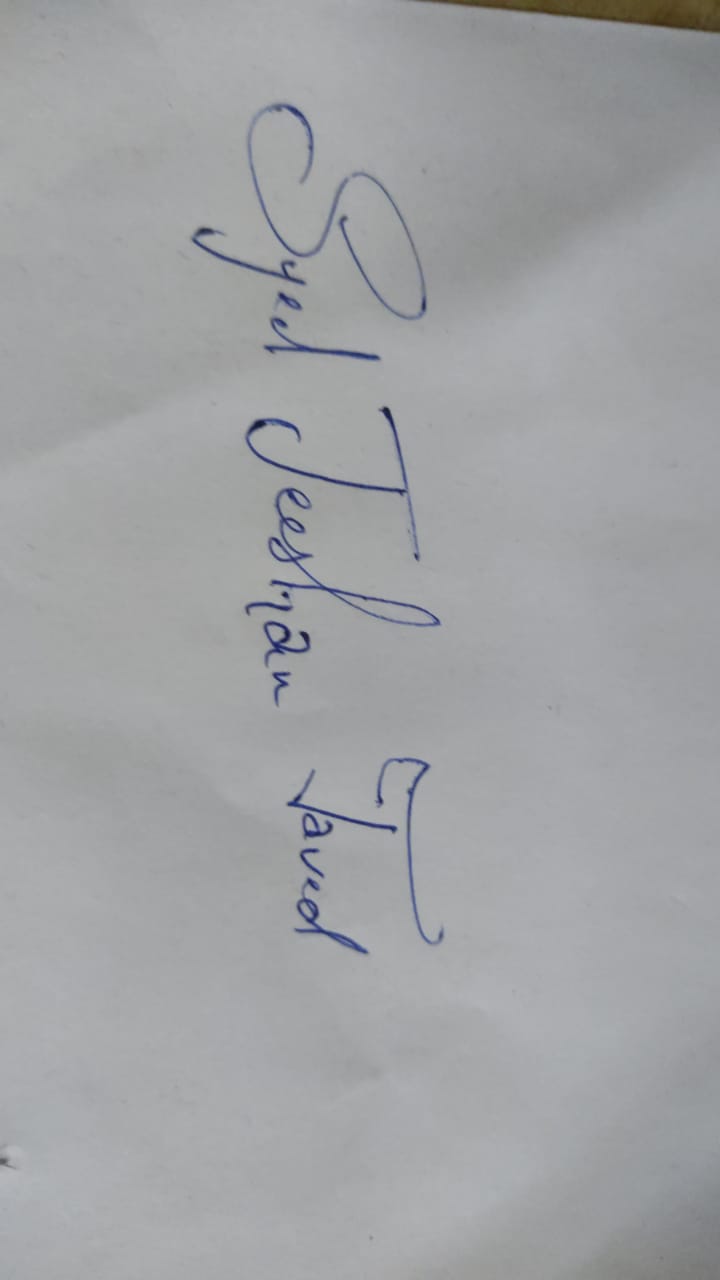
This report has not been submitted previously, either in part or full, to any other university or institution for the award of any degree, diploma, or certificate.

All information provided in this report is authentic to the best of my knowledge, and any references or contributions from other authors or sources have been duly acknowledged.

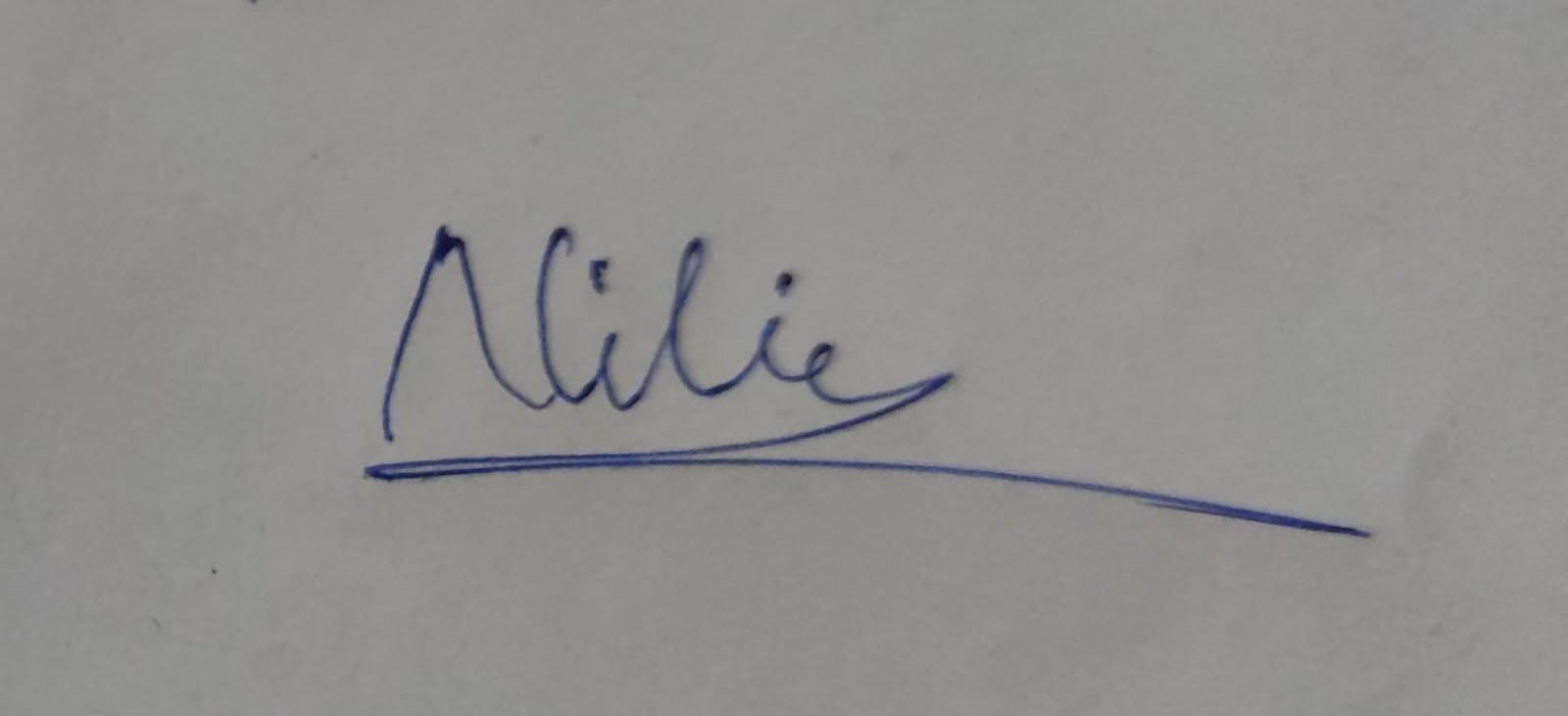
Signature:



Khyanrick Borah (232010007027)



Syed Jeeshan Javed (232010007057)

Nilim Kr. Gogoi (242050007009)

Date- 31/07/2025

## ****Acknowledgment****

On successful compilation and completion of this project titled “**AI powerd resume screening system”.**

First and foremost, we are deeply indebted to our supervisor, **Dr. Purabi Sharma**, for her unwavering support, insightful feedback, and constant encouragement. We extend our sincere thanks to Maam, as her expertise was instrumental in shaping the direction and depth of this project.

I am also profoundly grateful to the **Department of ECE, Gauhati University** for providing access to their extensive library resources and statistical databases, which were crucial for data collection and analysis.

I would like to express my sincere gratitude to all those who provided invaluable support and guidance throughout the preparation of this report. Their contributions were instrumental in its successful completion.

## ****Table of Contents****

1. Acknowledgment
2. Table of Contents
3. Introduction
4. Objective
5. Internship details and proposed work
6. Results & Observation
7. Skills Learned
8. Conclusion
9. References

**INTRODUCTION**

In today's competitive job market, recruiters and hiring managers are increasingly relying on data-driven tools to streamline the recruitment process and identify the most suitable candidates. One innovative application of this approach is resume ranking using artificial intelligence and machine learning, which allows organizations to efficiently match applicants to job requirements based on content relevance.

This project focuses on the development of an **AI-Powered Resume Ranking System** using natural language processing (NLP) and a **supervised learning algorithm**—the Decision Tree Classifier. The goal of the model is to analyze and score multiple resumes based on their textual similarity to a given job description. The core input features are extracted from uploaded resume PDFs and compared against the job description using the TF-IDF (Term Frequency–Inverse Document Frequency) method, enabling a more objective and scalable screening process.

The project adopts a step-by-step methodology that includes text extraction, preprocessing, vectorization, model training, and similarity-based scoring. To improve usability, a graphical user interface (GUI) is also integrated, allowing users to input job descriptions, adjust matching thresholds, upload resume files, and view ranked results with interactive visual feedback.

Through this project, I aimed to translate machine learning theory into a functional, real-world application. It provided hands-on experience in model development, user interface design, and performance evaluation. This mini project not only deepened my understanding of AI implementation in HR tech but also highlighted the potential of smart automation in solving modern recruitment challenges.

**OBJECTIVE**

The main goal of this project was to build an intelligent and interactive machine learning system that can automatically rank resumes based on their similarity to a specific job description. The aim was to reduce manual workload and bring efficiency and objectivity into the recruitment process using Python and basic NLP techniques.

### Objectives:

* **Resume Text Extraction**: Parsed multiple PDF resumes using the PyPDF2 library to extract clean, usable text data.
* **Feature Vectorization**: Used TF-IDF Vectorizer to convert both resumes and job descriptions into comparable numerical vectors.
* **Similarity Calculation**: Applied cosine similarity to score and rank each resume based on its relevance to the job description.
* **Threshold-Based Filtering**: Implemented logic to mark resumes as a match or not, depending on a user-defined similarity threshold.
* **Interactive GUI**: Developed a Tkinter-based graphical interface for easy uploading, job input, and live viewing of ranked results.
* **Automation**: Enabled automatic batch scoring of resumes, reducing human error and improving shortlisting speed.

This project illustrates how basic AI tools and automation can modernize hiring workflows, making candidate evaluation faster, consistent, and scalable.

**INTERNSHIP DETAILS AND PROPOSED WORK**

The internship was centered on developing a simple AI-powered application to rank resumes based on their textual similarity to a given job description. The project aimed to build a Resume Ranking System using Python, with the help of key libraries such as pypdf for reading resumes in PDF format, scikit-learn for implementing the machine learning logic, and tkinter for creating an interactive desktop interface.

At the core of this system is a basic content-matching algorithm powered by **TF-IDF vectorization** and **cosine similarity**, along with a user-defined **threshold feature** that filters out resumes falling below a certain match score. This makes the system both customizable and practical for real-world use in academic or professional hiring scenarios.

### **Steps Followed**:

**1. Resume Extraction:**

Resumes were loaded in bulk from a folder. The pypdf library was used to extract text content from PDF files. This raw text data was stored for further processing.

**2. Text Vectorization and Similarity Scoring:**

The input job description and all extracted resumes were converted into numerical feature vectors using **TF-IDF (Term Frequency-Inverse Document Frequency)** via TfidfVectorizer from sklearn. Then, **cosine similarity** was calculated between the job description vector and each resume vector to quantify how closely each resume matched the job description.

**3. Resume Ranking and Threshold Filtering:**

Each resume was scored based on similarity. The user could define a **threshold value** (e.g., 0.2), and only resumes with a similarity score **equal to or above this threshold** were considered relevant. The results were then ranked from highest to lowest similarity.

**4. GUI Development:**

A simple **graphical user interface (GUI)** was developed using tkinter. This interface allowed users to:

* Input a job description.
* Set a similarity threshold.
* Load resume files from a directory.
* View ranked results along with match scores and acceptance/rejection labels.

**5. Output and Presentation:**

Final results were displayed in a scrollable text area within the GUI and optionally saved to a CSV file. Resumes that met the threshold were labeled as a "Match ", while others were flagged as "Not Match ".

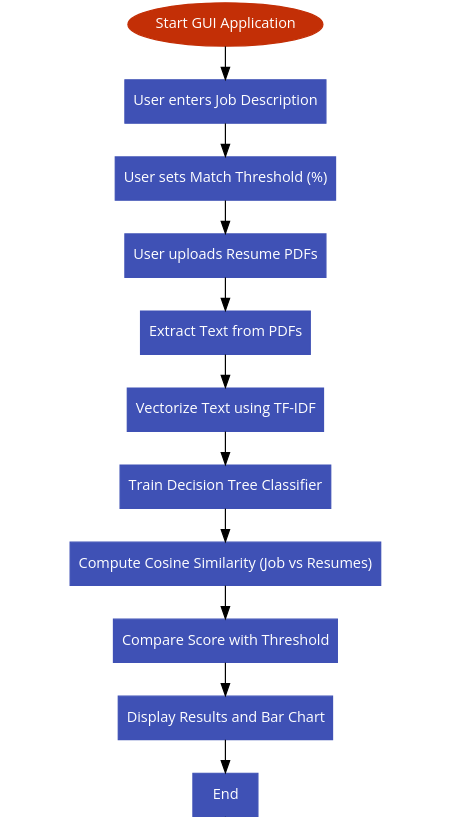


Fig 1 : workflow of AI powered Resume screening System

**SKILLS LEARNED**

This internship project provided valuable hands-on experience in applying machine learning and automation to solve real-world problems in the domain of recruitment. Throughout the development of the Resume Ranking System, I gained a variety of technical and practical skills:

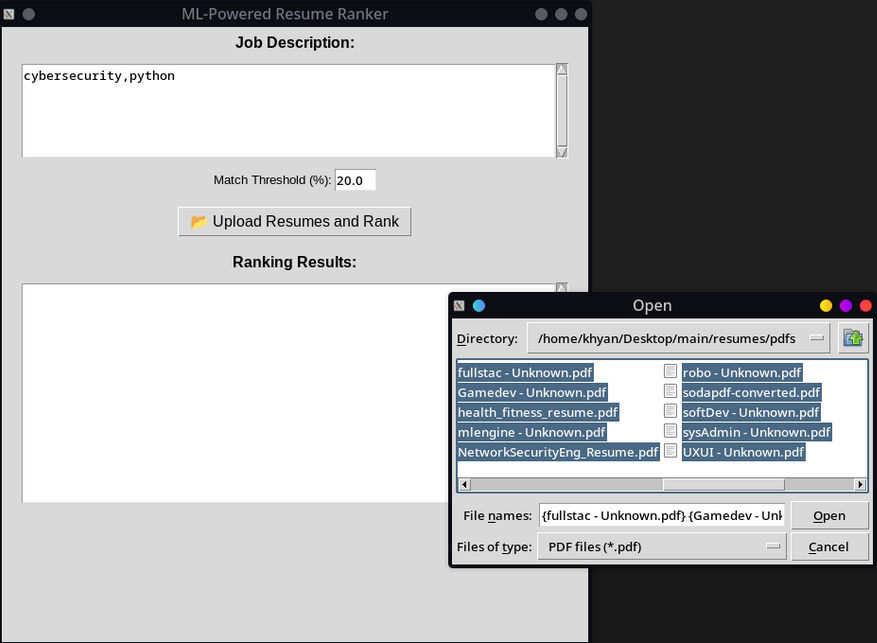
* **Python Programming**: Strengthened my command over Python for tasks such as file handling, data parsing, logic implementation, and building end-to-end applications.
* **Text Processing & NLP**: Learned how to extract meaningful text from PDF resumes using the PyPDF2 library and apply TfidfVectorizer to convert unstructured text into numerical vectors suitable for comparison.
* **Similarity Scoring**: Understood how cosine similarity works and implemented it to measure the relevance of each resume against a job description.
* **Threshold Logic**: Added logic to filter resumes based on a predefined similarity threshold to distinguish between relevant and irrelevant candidates.
* **GUI Development**: Designed and built an intuitive graphical interface using tkinter to allow users to upload resumes, input job descriptions, and view results easily.
* **Automation**: Automated the reading, scoring, and ranking of multiple resumes, reducing manual effort and improving consistency.
* **Problem Solving**: Improved debugging skills while handling challenges such as missing data, poorly formatted files, and text extraction failures.

**Results and Observation**

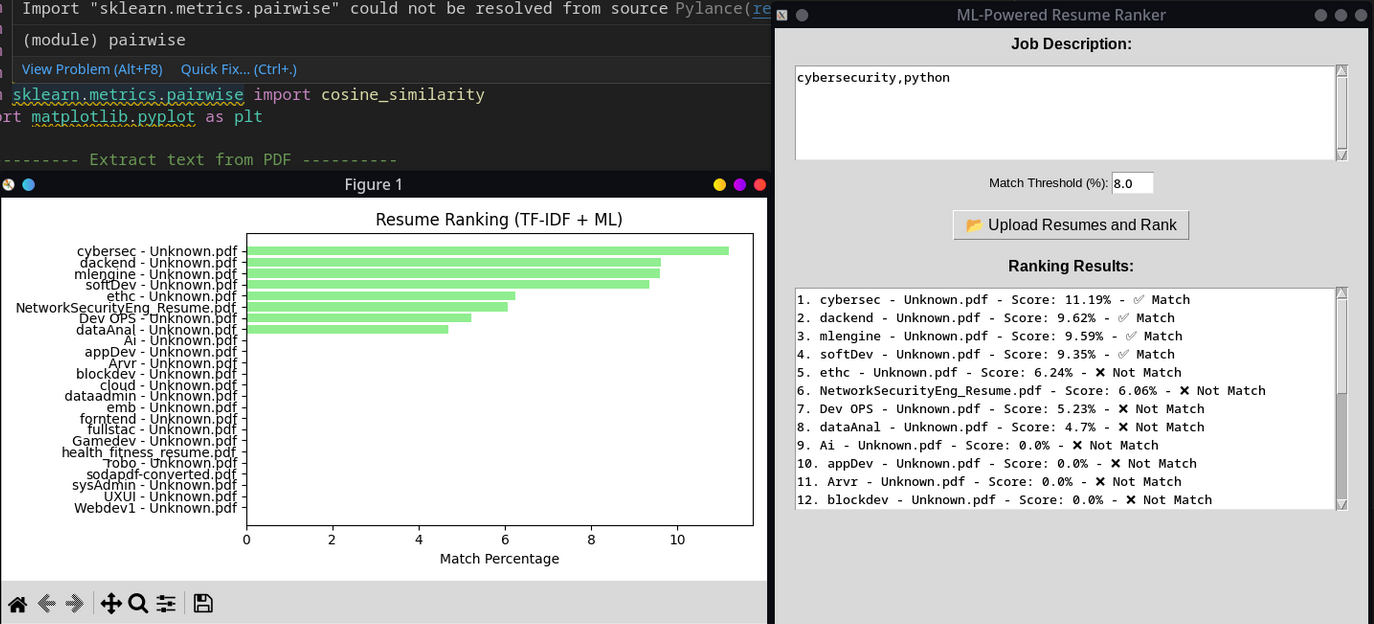
After entering the Job description and Match Threshold, we upload the resumes into the program by clicking on the “Upload Resumes and Rank”

Button

### ****Program GUI****



### ****Result****



****Output Visualization:****

**◦ The system presents the resume similarity scores using a bar chart, providing a clear and intuitive comparison of each candidate’s relevance score.  
◦ Each resume’s score is visually compared against the user-defined threshold, with appropriate labels (“Match” or “Not Match”) assigned to indicate its suitability. This improves interpretability and helps users quickly identify top candidates.**

## ****Conclusion****

This internship project offered valuable insights into applying artificial intelligence for resume screening—a critical task in modern recruitment. By building an AI-powered Resume Ranking System, I gained experience in combining natural language processing, machine learning, and automation to solve practical problems efficiently.

* **Python Proficiency**: Improved my fluency in Python by integrating multiple libraries for text extraction, data processing, and system logic.
* **Document Handling**: Utilized PyPDF2 to parse and process large volumes of PDF resumes, handling various formats and ensuring robust data input.
* **Text Vectorization & NLP**: Applied TfidfVectorizer to convert job descriptions and resume texts into comparable vector formats, enabling meaningful analysis.
* **Cosine Similarity**: Implemented cosine similarity to evaluate textual relevance between resumes and job descriptions, forming the basis for scoring and ranking.
* **Threshold-Based Filtering**: Developed logic to classify candidates as relevant or irrelevant based on a similarity threshold, mimicking a basic AI hiring filter.
* **GUI Design**: Built a user-friendly desktop interface using tkinter, allowing non-technical users to operate the system seamlessly.
* **Automation & Scalability**: Automated resume scanning and ranking, paving the way for scaling this system in real-world HR pipelines.

This project enhanced my understanding of AI applications in HR tech and helped me bridge the gap between theoretical knowledge and deployable solutions.

## ****References****

1. **Reddit – r/learnpython: How to Extract Text from PDFs**  
   <https://www.reddit.com/r/learnpython/comments/12x9xgi/>
2. **Feature Extraction**

https://en.wikipedia.org/wiki/Feature\_extraction

1. **Tkinter GUI Applications**[**https://github.com/ParthJadhav/Tkinter-Designer**](https://github.com/ParthJadhav/Tkinter-Designer)
2. **Cosine Similarity**

https://en.wikipedia.org/wiki/Cosine\_similarity