**Project Title**

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning

with

TechSaksham – A joint CSR initiative of Microsoft & SAP

by

**Name of Student, Email id**

Under the Guidance of

**Name of Guide**

**ACKNOWLEDGEMENT**

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

Firstly, we would like to thank my supervisor, …………….., for being a great mentor and the best adviser I could ever have. His advice, encouragement and the critics are a source of innovative ideas, inspiration and causes behind the successful completion of this project. The confidence shown in me by him was the biggest source of inspiration for me. It has been a privilege working with him for the last one year. He always helped me during my project and many other aspects related to the program. His talks and lessons not only help in project work and other activities of the program but also make me a good and responsible professional.

……...

#### This Acknowledgement should be written by students in your own language (Do not copy and Paste)

#### …..

……

….

……

#### **ABSTRACT**

In today's job market, organizations receive an overwhelming number of resumes for each job opening, making manual screening a time-consuming and biased process. To address these challenges, this project presents an AI-powered Resume Screening & Ranking System that automates the process using Natural Language Processing (NLP) and Machine Learning (ML). The system consists of two main components: resume screening, which predicts the most relevant job category for each resume using TF-IDF vectorization and a K-Nearest Neighbors (KNN) classifier, and resume ranking, which compares resumes with a given job description and ranks them based on cosine similarity scores. To ensure accessibility and ease of use, a web-based application is developed using Streamlit, allowing recruiters to upload multiple resumes at once. The system extracts candidate details such as name and email, assigns similarity scores, predicts job categories, and presents a ranked list of resumes based on relevance. Unlike complex deep learning approaches, our method leverages efficient and interpretable ML techniques to achieve high accuracy while maintaining low computational costs. The implementation of this system significantly reduces hiring time and improves candidate selection efficiency, making it a valuable tool for modern recruitment.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**  **Introduction 1**

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 2

1.4. Scope of the Project 2

**Chapter 2.**  **Literature Survey 3**

**Chapter 3.**  **Proposed Methodology**

**Chapter 4.**  **Implementation and Results**

**Chapter 5. Discussion and Conclusion**

**References**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

The traditional resume screening process is a tedious, time-consuming, and often biased task for recruiters. With the increasing number of job applications, manually evaluating each resume to identify the most suitable candidates has become inefficient and impractical. Human-based screening is prone to inconsistencies, unconscious bias, and errors, leading to potential misjudgment in candidate selection. Additionally, recruiters often struggle to compare resumes effectively against job descriptions, making it difficult to shortlist the best candidates quickly. This inefficiency not only increases hiring time but also results in missed opportunities for highly qualified applicants. To address these challenges, this project introduces an AI-powered Resume Screening & Ranking System that automates the process using Machine Learning and Natural Language Processing. By extracting key details from resumes, predicting job categories, and ranking candidates based on their relevance to job descriptions, the system enhances recruitment efficiency, reduces bias, and improves candidate selection, making it a valuable tool for modern hiring processes.

* 1. **Motivation:**

Hiring the right candidate is one of the most crucial yet challenging aspects of recruitment. With companies receiving hundreds or even thousands of resumes for a single job opening, recruiters often find themselves overwhelmed by the sheer volume of applications. Manually screening resumes is not only time-consuming but also prone to human bias and errors, leading to missed opportunities for deserving candidates. This challenge sparked the motivation behind developing an AI-powered Resume Screening & Ranking System—one that can **automate, streamline, and optimize** the hiring process.

The potential impact of this system extends beyond just saving time. By leveraging **Machine Learning and Natural Language Processing**, this project ensures that every resume is evaluated **objectively and fairly**, reducing bias in hiring decisions. Organizations can make **data-driven recruitment choices**, ensuring that only the most relevant candidates move forward in the hiring pipeline. Additionally, by integrating a **user-friendly web application**, even non-technical recruiters can benefit from AI-driven insights without requiring expertise in complex algorithms.

Beyond corporate recruitment, this system has applications in **educational institutions, job portals, and talent acquisition platforms**, where efficient resume evaluation is crucial. Whether it's for **screening fresh graduates** or **hiring experienced professionals**, this solution can revolutionize the way resumes are processed, making hiring smarter, fairer, and more efficient.

* 1. **Objective:**

The primary objective of this project is to develop an AI-powered Resume Screening & Ranking System that automates and optimizes the recruitment process by leveraging **Machine Learning (ML) and Natural Language Processing (NLP)** techniques. The system aims to **extract key details** such as name and email from resumes, **classify resumes into relevant job categories**, and **rank candidates based on their similarity to a given job description**. By implementing an **efficient and unbiased resume evaluation process**, this project seeks to reduce the time and effort required for manual screening while improving the accuracy and fairness of candidate selection. Additionally, the system is integrated into a **user-friendly Streamlit-based web application**, allowing recruiters to **upload multiple resumes, view ranked results, and download a structured report** for further decision-making. Ultimately, this project aims to enhance recruitment efficiency, minimize human bias, and provide a scalable, automated solution for modern hiring challenges.

* 1. **Scope of the Project:**

The **AI-powered Resume Screening & Ranking System** is designed to streamline the recruitment process by automating resume evaluation and ranking. The system focuses on **extracting key details from resumes, predicting suitable job categories, and ranking candidates based on their relevance to a given job description**. This project is primarily intended for **HR professionals, recruiters, and hiring managers** to assist in **efficiently shortlisting candidates** from a large pool of applicants.

The system is implemented as a **web-based application using Streamlit**, allowing recruiters to **upload multiple resumes at once, analyze them, and download ranked results in a structured format**. It leverages **TF-IDF vectorization and Cosine Similarity** for ranking, and a **K-Nearest Neighbors (KNN) classifier** for job category prediction.

**Limitations**

* The system currently supports **only PDF resumes** and does not process other formats like DOCX or TXT.
* **Predicted job categories** are based on pre-trained data, meaning accuracy may depend on the diversity and quality of training data.
* **Similarity ranking is based on textual content** and does not evaluate soft skills, work ethics, or cultural fit.
* **Limited contextual understanding** compared to deep learning-based models like BERT, though it ensures efficiency and interpretability.
* The system relies on **well-structured resumes**; highly unstructured resumes may affect accuracy in entity extraction (Name, Email).

Despite these limitations, the project provides a **fast, scalable, and efficient** solution to improve **resume evaluation in recruitment processes**. Future enhancements may include **support for multiple file formats, deep-learning-based models for better ranking, and an improved entity extraction mechanism**.

**CHAPTER 2**

**Literature Survey**

* 1. **Review relevant literature or previous work in this domain.**

The increasing adoption of Artificial Intelligence (AI) and Machine Learning (ML) in recruitment has led to significant advancements in automated resume screening and ranking systems. Several studies and implementations have explored the use of Natural Language Processing (NLP), classification algorithms, and similarity-based ranking techniques to enhance the efficiency and accuracy of resume evaluation. Traditional recruitment processes rely heavily on manual screening, which is time-consuming, prone to bias, and inefficient in handling large applicant pools. Research has shown that text classification models, entity recognition techniques, and vectorization-based similarity measures can significantly improve candidate shortlisting.

* 1. **Mention any existing models, techniques, or methodologies related to the problem**

Several methodologies have been proposed in literature for resume screening and ranking. Support Vector Machines (SVM), Naïve Bayes, Decision Trees, and Random Forests have been commonly used for resume classification tasks, with SVM often outperforming others in structured text classification. More recent approaches leverage Deep Learning models such as BERT (Bidirectional Encoder Representations from Transformers) for semantic understanding and contextual resume-job matching. In terms of ranking resumes, studies have explored the use of TF-IDF (Term Frequency-Inverse Document Frequency) and Cosine Similarity to measure textual relevance between resumes and job descriptions. The Sentence-BERT (SBERT) model has been introduced as an improved version of BERT for sentence-level semantic similarity, achieving state-of-the-art performance in ranking-based applications.

* 1. **Highlight the gaps or limitations in existing solutions and how your project will address them.**

While existing solutions provide promising approaches, they often come with several **limitations**. **Deep Learning models such as BERT require extensive computational resources**, making them **inaccessible for small-scale applications**. Many traditional methods, including **SVM and Naïve Bayes, lack interpretability** and struggle with handling large-scale resume datasets efficiently. Additionally, existing solutions **primarily focus on classification or ranking separately** but do not integrate both aspects into a **single, user-friendly system**. Another significant gap is the **lack of interactive, easy-to-use interfaces** for recruiters, who may not have technical expertise.

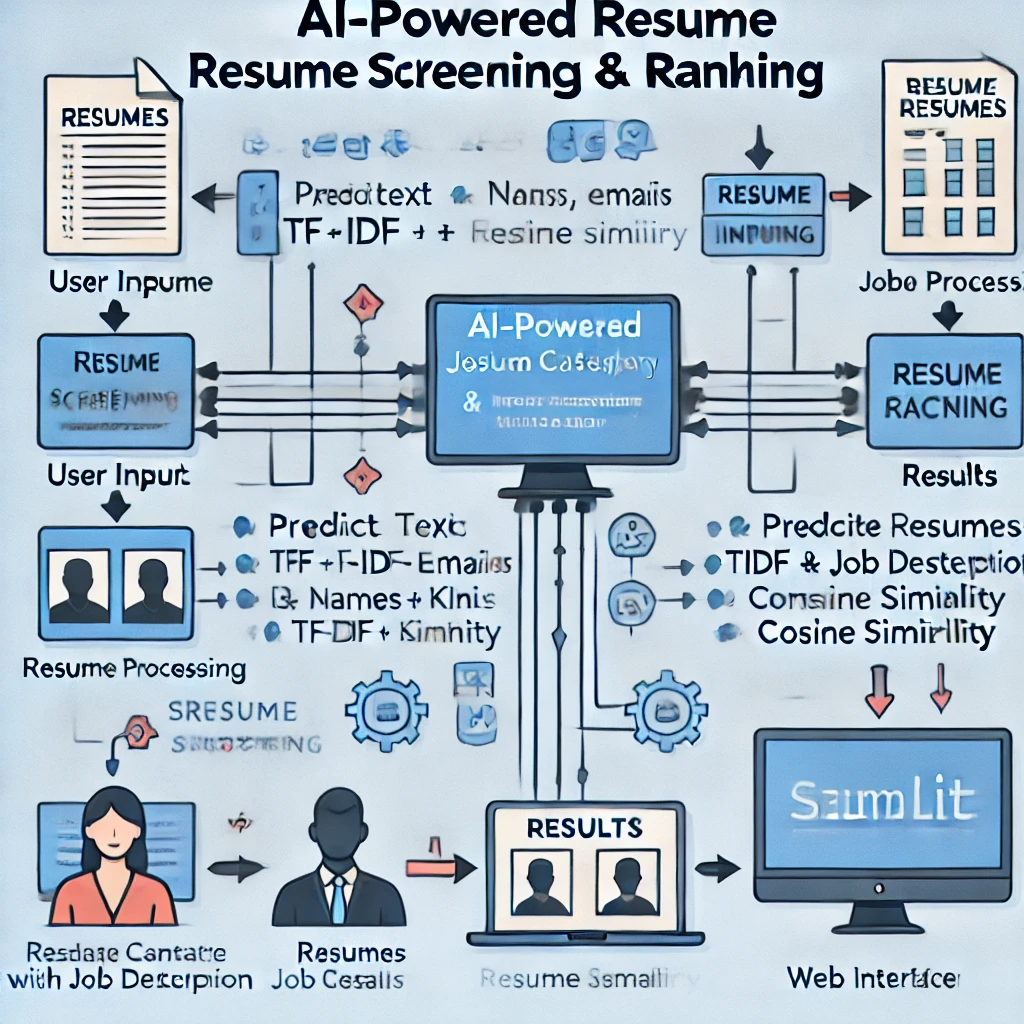
To address these limitations, our **AI-powered Resume Screening & Ranking System** combines **efficient ML techniques with a user-friendly interface**. The project utilizes **TF-IDF and Cosine Similarity for ranking**, ensuring computational efficiency, while **K-Nearest Neighbors (KNN) is employed for job category prediction**, providing an interpretable and scalable solution. Furthermore, the integration of these functionalities into a **Streamlit-based web application** bridges the gap between technical complexity and usability, enabling recruiters to seamlessly analyze and shortlist candidates. This approach ensures **fast, fair, and unbiased resume evaluation**, significantly enhancing the recruitment process.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

**System Workflow:**

1. **User Input**:
   * The system accepts multiple resume files (**PDF format**) and a **job description** as input.
2. **Resume Processing**:
   * Extracts text from PDF resumes using **PyPDF2**.
   * Identifies and extracts **Name and Email** using **Regular Expressions (Regex)**.
   * Cleans and preprocesses resume text for further analysis.
3. **Resume Screening (Job Category Prediction)**:
   * Uses **TF-IDF (Term Frequency-Inverse Document Frequency)** to convert resume text into numerical features.
   * A **K-Nearest Neighbors (KNN) classifier** is trained to predict the **job category** based on pre-labeled data from a dataset.
4. **Resume Ranking (Job Relevance Matching)**:
   * Converts both the **resume text** and the **job description** into **TF-IDF vectors**.
   * Computes **Cosine Similarity** to measure the relevance of each resume to the job description.
   * Assigns a **similarity score** to each resume.
5. **Result Generation**:
   * Ranks resumes based on **similarity scores** in **descending order**.
   * Displays a structured table with **Rank, Name, Email, Similarity Score, and Predicted Job Category**.
   * Allows recruiters to **download the results as a CSV file**.
6. **Streamlit UI**:
   * The system is integrated into a **Streamlit-based web application**.
   * Users can **upload resumes, input job descriptions, and view/download ranked results** in an interactive interface.
   * 
   1. **Requirement Specification**

The implementation of the AI-powered Resume Screening & Ranking System requires a combination of hardware and software tools to ensure efficient processing, model training, and a seamless user experience through a web-based interface.

* + 1. **Hardware Requirements:**

The system is designed to run on standard computing devices without requiring specialized hardware. However, for optimal performance, the following specifications are recommended:

* Processor: Intel Core i5 or higher / AMD Ryzen 5 or higher
* RAM: Minimum 8GB (Recommended: 16GB for large-scale resume processing)
* Storage: At least 10GB of free disk space (Recommended: SSD for faster data processing)
* GPU (Optional): A dedicated GPU (e.g., NVIDIA GTX 1650 or higher) can accelerate certain NLP tasks but is not mandatory.
* Operating System: Windows 10/11, macOS, or Linux
  + 1. **Software Requirements:**

To develop, run, and deploy the system, the following software tools and libraries are required:

Programming Language & Environment:

* Python 3.8 or higher – Core development language
* Jupyter Notebook / VS Code / PyCharm – For development and debugging
* Streamlit – To create the interactive web-based application

Libraries & Frameworks:

* NLP & Text Processing:
  + nltk – For basic text processing
  + scikit-learn – For ML models (TF-IDF, KNN, Cosine Similarity)
  + PyPDF2 – To extract text from PDFs
  + re (Regular Expressions) – For email and name extraction
* Machine Learning:
  + TfidfVectorizer – Convert text into numerical representations
  + KNeighborsClassifier – Used for resume classification
  + cosine\_similarity – For ranking resumes based on job descriptions
* Data Handling & Visualization:
  + pandas – To manipulate and store structured data
  + numpy – For numerical computations
  + matplotlib/seaborn (optional) – For data visualization

Deployment & Execution:

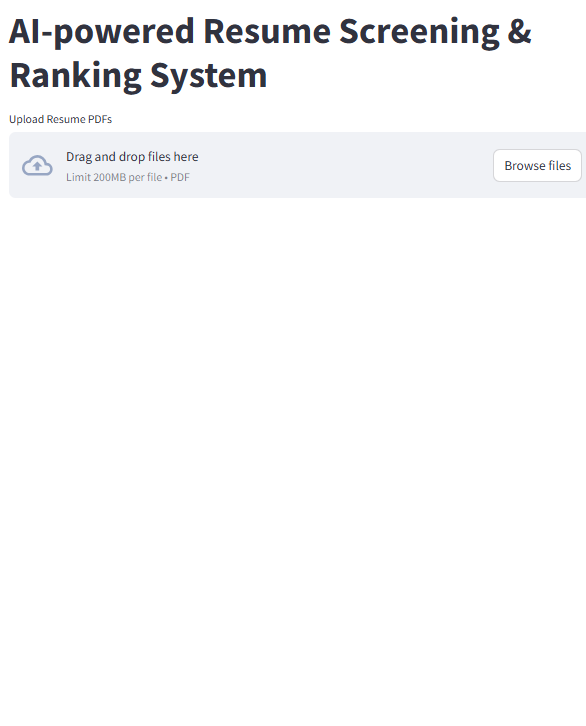
* pip (Python Package Installer) – To install required dependencies
* GitHub (Optional) – For version control and collaboration
* Streamlit Cloud / AWS / Azure (Optional) – For online deployment

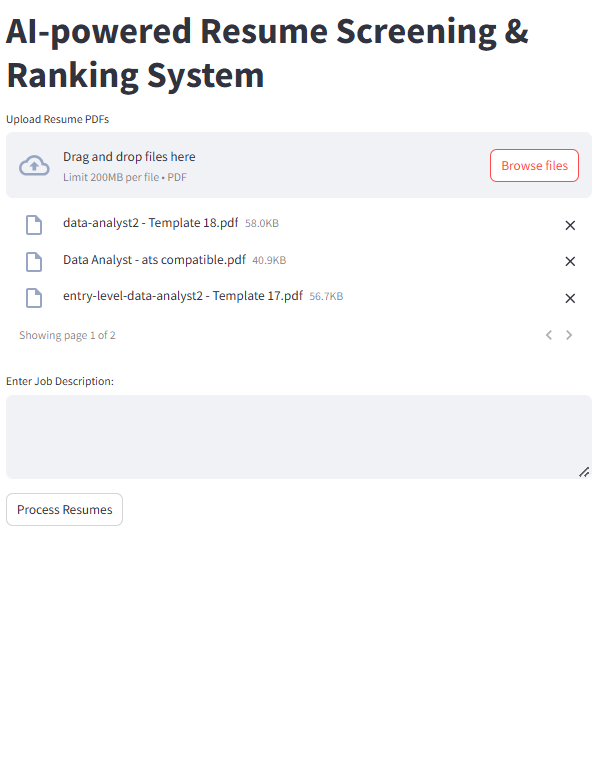
This setup ensures that the system can efficiently process resumes, classify candidates, and rank them for easy selection in a fast, scalable, and user-friendly manner.

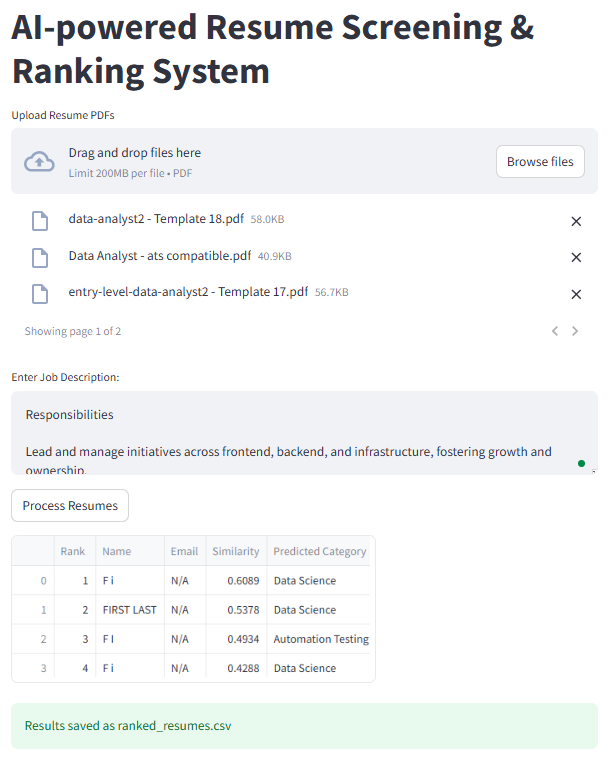
**CHAPTER 4**

**Implementation and Result**

* 1. **Snap Shots of Result:**

****

****

****

* 1. **GitHub Link for Code:**

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

While the **AI-powered Resume Screening & Ranking System** effectively automates resume evaluation and ranking, there are several areas for improvement and future enhancements:

* **Support for Multiple File Formats:** Currently, the system processes only PDF resumes. Future versions can include support for **DOCX, TXT, and scanned image-based resumes** using **OCR (Optical Character Recognition)** techniques like Tesseract.
* **Advanced Named Entity Recognition (NER):** The current method extracts names and emails using regex-based approaches. Implementing **pre-trained NLP models** such as **BERT-based NER models** can enhance accuracy in extracting structured information.
* **Deep Learning-Based Ranking:** While TF-IDF and Cosine Similarity provide efficient ranking, future versions could integrate **Transformer-based models like BERT or SBERT** for **context-aware resume-job description matching**, leading to improved ranking accuracy.
* **Improved Job Category Prediction:** The K-Nearest Neighbors (KNN) classifier can be replaced with **more sophisticated models** such as **Random Forest, XGBoost, or a fine-tuned deep learning classifier** to enhance job category prediction accuracy.
* **Skill-Based Matching:** Future versions can extract **technical and soft skills** from resumes and compare them with **required skills in job descriptions** to provide a more **comprehensive ranking**.
* **Bias Reduction & Fairness:** To minimize bias in AI-based screening, the system can incorporate **Fair AI techniques** that prevent discrimination based on gender, age, or ethnicity by **ensuring unbiased feature selection**.
* **Web Deployment & Integration:** The Streamlit application can be **hosted online** using **AWS, Azure, or Streamlit Cloud** to make it **accessible to recruiters globally**. Additionally, integration with **ATS (Applicant Tracking Systems)** can further streamline hiring processes.
* **Interactive Dashboard & Filters:** Implementing an **advanced visualization dashboard** with filters (e.g., experience level, skills, job category) would help recruiters make **more informed hiring decisions** efficiently.

By implementing these enhancements, the system can evolve into a **fully scalable, intelligent recruitment assistant**, making hiring **faster, more accurate, and unbiased**.

**Conclusion**

The **AI-powered Resume Screening & Ranking System** successfully automates and optimizes the recruitment process by leveraging **Natural Language Processing (NLP) and Machine Learning (ML)**. By integrating **resume screening and ranking into a single unified system**, it enhances **efficiency, accuracy, and fairness** in candidate selection. The system extracts **key details (name, email)**, predicts **job categories**, and ranks resumes based on **textual similarity to job descriptions** using **TF-IDF and Cosine Similarity**.

The **Streamlit-based web application** provides a **user-friendly interface**, allowing recruiters to **upload multiple resumes, analyze candidate suitability, and download ranked results**, thereby reducing manual effort and hiring time. Unlike traditional recruitment methods that are prone to **bias and inefficiency**, this system ensures **objective and data-driven candidate evaluation**.

The project contributes significantly to **modernizing the hiring process**, making it **scalable, unbiased, and time-efficient**. While the system performs well in automating resume evaluation, future enhancements such as **deep-learning-based ranking, skill-based matching, and ATS integration** could further refine its effectiveness. Overall, this project serves as a **valuable tool for recruiters**, helping them make **informed hiring decisions quickly and accurately** in an increasingly competitive job market.

**REFERENCES**

1. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.