**Resume Screening Application**

Natural Language Processing Mini Project

Submitted by

Rohan Khatode(Roll No : BA56)

Arnav Mahalpure(Roll No: BA63)

Aayush Bhalavi(Roll No: BA09)

Ashish Chandekar(Roll No: BA67)

FINAL YEAR COMPUTER ENGINEERING



Department of Computer Engineering

International Institute of Information Technology

Hinjawadi, Pune – 411057

SEMESTER II (AY 2024-25)

**TABLE OF CONTENTS**

**TITLE PAGE NO.**

1. ABSTRACT 1
2. INTRODUCTION 2
   1. SCOPE 2
   2. REQUIREMENT ANALYSIS 3
   3. SOFTWARE AND HARDWARE DETAILS 5
   4. LIBRARIES / PACKAGES USED 6
3. DATA SET DESCRIPTION AND LINK TO DATASET 7
4. OUTPUT 8
5. CONCLUSION 10

# ABSTRACT

In today's competitive job market, organizations face increasing challenges in efficiently screening and selecting candidates. Traditional manual resume screening is often time-consuming and prone to human error, leading to mismatches between candidates and job roles. To address this, we have developed an AI Resume Screening Application utilizing advanced data analytics and big data techniques. The application employs the K-Nearest Neighbors (KNN) algorithm to predict the most suitable job role based on resume data, significantly streamlining the recruitment process. Additionally, the integration with Langchain, a cutting-edge language processing platform, enables seamless interaction with a Language Model (LLM), offering personalized feedback and optimization tips to candidates for improving their resumes. This innovative solution not only enhances the efficiency of recruitment but also empowers candidates to better align their resumes with job expectations.

# INTRODUCTION

Recruitment is a critical function in organizations, yet the increasing number of applicants for each job opening has made the process of identifying the right candidates more challenging. Manual screening of resumes can be labor-intensive, time-consuming, and may lead to missed opportunities due to human bias or oversight. With the rapid advancements in artificial intelligence and machine learning, there is now an opportunity to automate the resume screening process, improving accuracy and efficiency.

Our AI Resume Screening Application leverages the power of the K-Nearest Neighbors (KNN) algorithm to automate candidate matching. By training on large datasets of resumes and job descriptions, the system predicts suitable roles for each candidate, drastically reducing the time required for initial screening. In addition to its predictive capabilities, the application integrates Langchain’s language processing capabilities to provide candidates with personalized resume feedback and optimization suggestions. This combination of AI-driven role prediction and interactive resume enhancement offers a more efficient and candidate-friendly approach to recruitment.

**2.1. SCOPE**

The AI Resume Screening Application aims to revolutionize the recruitment process for organizations by reducing the manual workload involved in resume screening and improving the accuracy of candidate-job matching. The system will:

* Utilize the KNN algorithm to predict the most appropriate job role for candidates based on resume analysis.
* Integrate with Langchain to leverage a Language Model (LLM) for generating personalized resume improvement tips and feedback.
* Provide recruiters with a streamlined and efficient workflow, significantly reducing the time spent on screening resumes manually.
* Offer candidates a user-friendly interface where they can interact with the application to optimize their resumes in real-time, increasing their chances of being shortlisted.
* Support scalability by handling large volumes of resumes and job descriptions, making it suitable for organizations of all sizes.
* Foster a better alignment between candidate skills and job requirements, leading to improved hiring outcomes.

**2.2. REQUIREMENT ANALYSIS:**

**Functional Requirements:**

1. Resume Upload:

The system allows users to upload resumes in various formats (e.g., PDF, DOCX).

Resumes are parsed and converted into structured data for analysis.

1. Candidate Role Prediction:

The application employs the K-Nearest Neighbors (KNN) algorithm to analyze resumes and predict the most suitable job role for each candidate. Based on the candidate's skills, education, and experience, the system recommends job roles.

1. Personalized Resume Feedback:

Integration with Langchain allows the system to analyze resumes and provide personalized feedback, such as formatting improvements or skill alignment suggestions. Feedback is displayed interactively within the user interface.

1. Interactive User Interface:

The system provides an intuitive dashboard where candidates can view their role predictions, job matches, and personalized feedback.

**Non-Functional Requirements:**

1. Performance:

The system should handle large volumes of resumes without significant latency.

Predictive role matching should be completed within seconds for each resume.

1. Scalability:

The system must be able to scale horizontally to accommodate an increasing number of users, resumes, and job descriptions. The application architecture should support load balancing and distributed computing.

1. Usability:

The user interface should be easy to navigate, with clear instructions and feedback.

1. Security:

Resumes contain personal information; therefore, the system must ensure data privacy and protection. All data transmissions must be encrypted using secure protocols (e.g., HTTPS).

1. Compatibility :

The application should be compatible with major web browsers (Chrome, Firefox,

Safari, etc.).

**2.3. SOFTWARE AND HARDWARE REQUIREMENTS:**

**Software Requirements:**

* + 1. Operating System: Windows 11
    2. Programming Languages: Python
    3. Development Tools: Jupyter notebook
    4. Browser: Latest versions of Chrome

**Hardware Requirements:**

Processor: Intel i5.

RAM: 8 GB (minimum) – 16 GB

GPU: A dedicated GPU (e.g., NVIDIA GTX/RTX series) is recommended for developers working on machine learning models locally.

Storage: At least 256 GB SSD (minimum) – 512 GB SSD or more

**2.4. LIBRARIES / PACKAGES USED:**

1. **Pandas**
2. **Numpy**
3. **Matplotlib**
4. **Seaborn**
5. **RE**
6. **Scikit-learn:**
   * LabelEncoder
   * TfidfVectorizer
   * Model\_Selection
   * KNeighborsClassifier
   * Metrics
   * SVM

# DATASET DESCRIPTION AND LINK OF DATASET

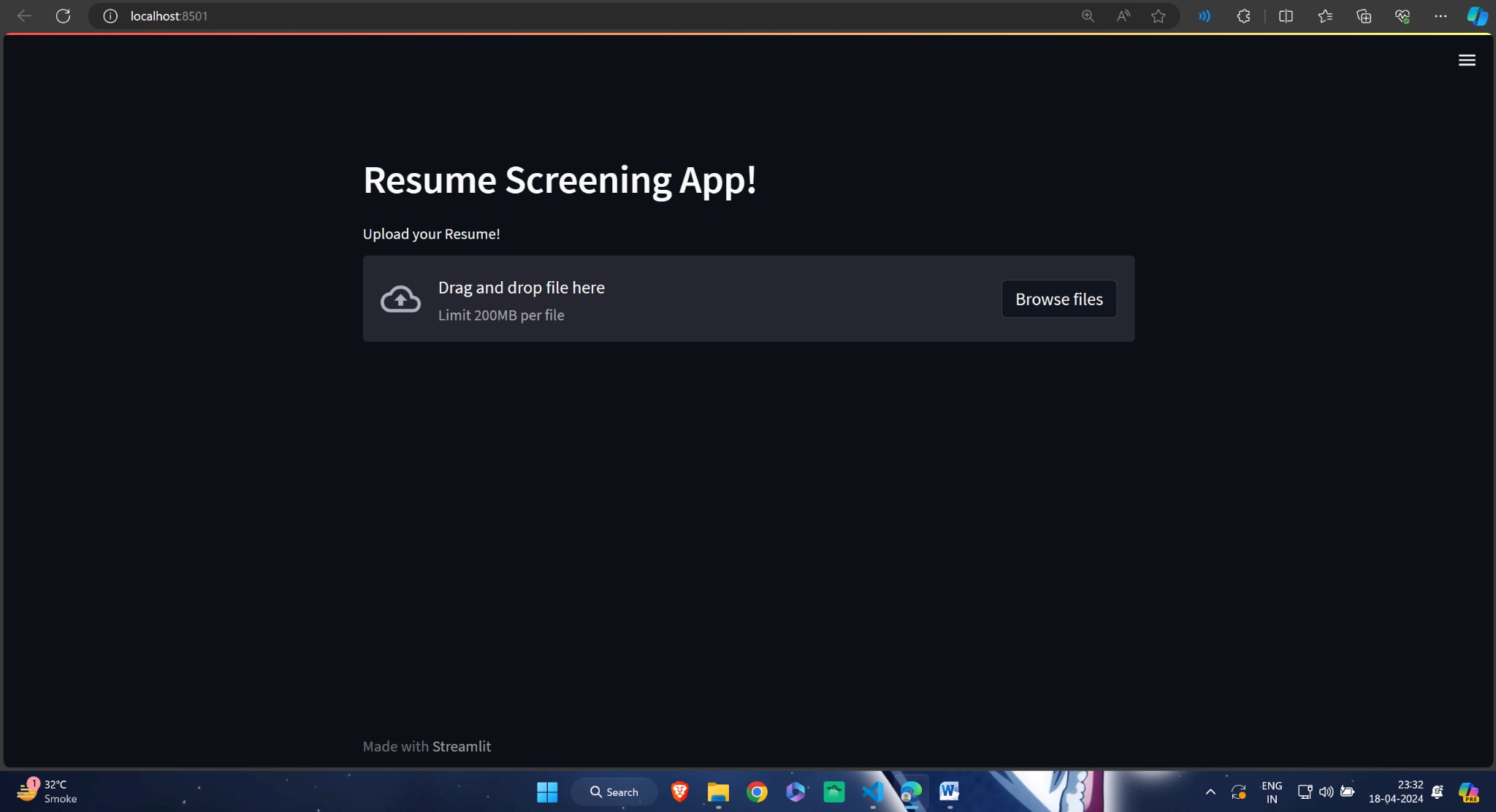
**Description**

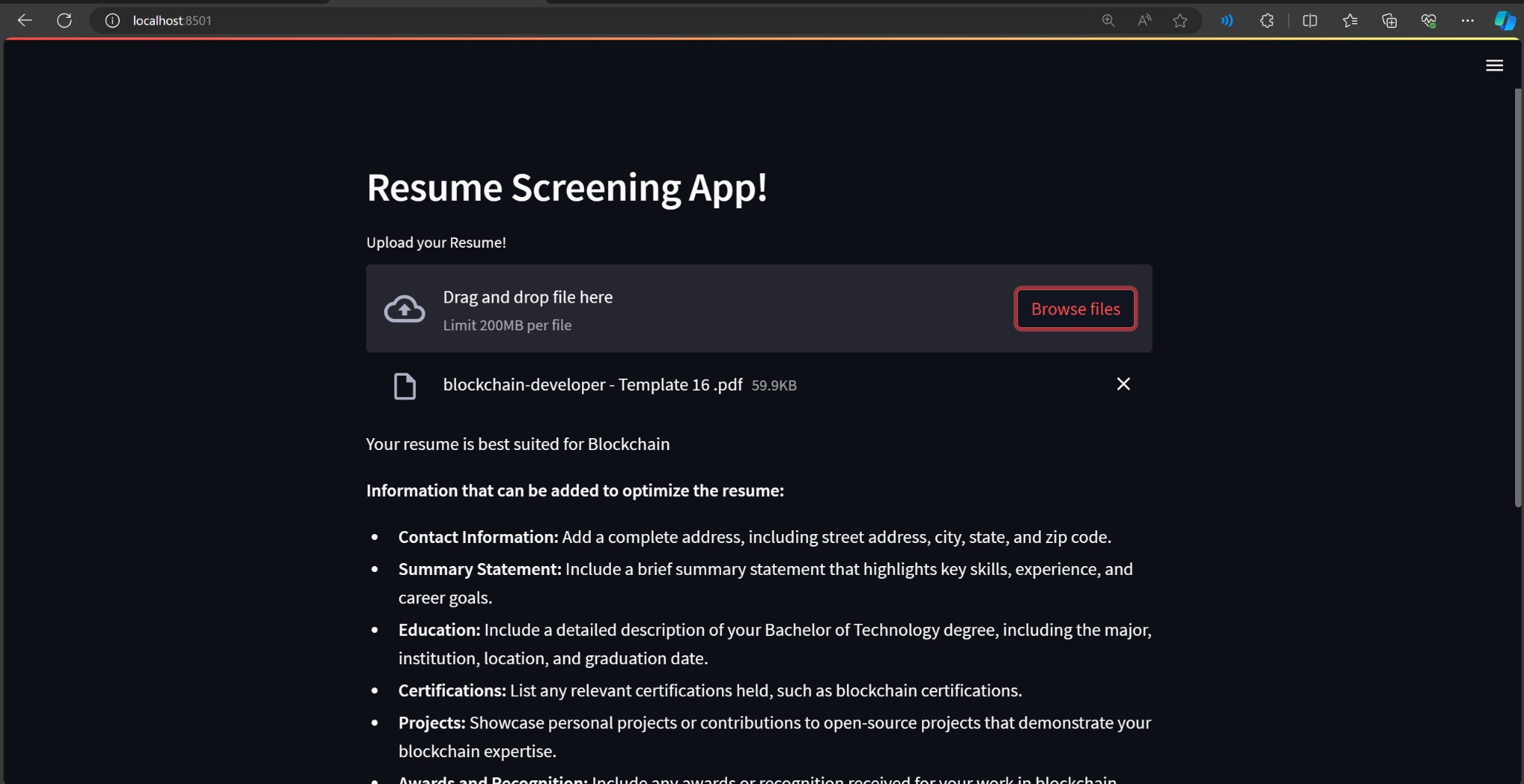
The "Resume Screening Dataset" on Kaggle is designed for developing and testing machine learning models for automated resume screening. The dataset consists of resumes labeled with attributes like education, experience, skills, and more, enabling the development of models to rank or classify candidates based on job requirements. It can be particularly useful for tasks like text classification, natural language processing, and building recommendation systems in recruitment. The dataset provides insights into the key elements recruiters look for when evaluating candidates and can be used to streamline hiring processes.

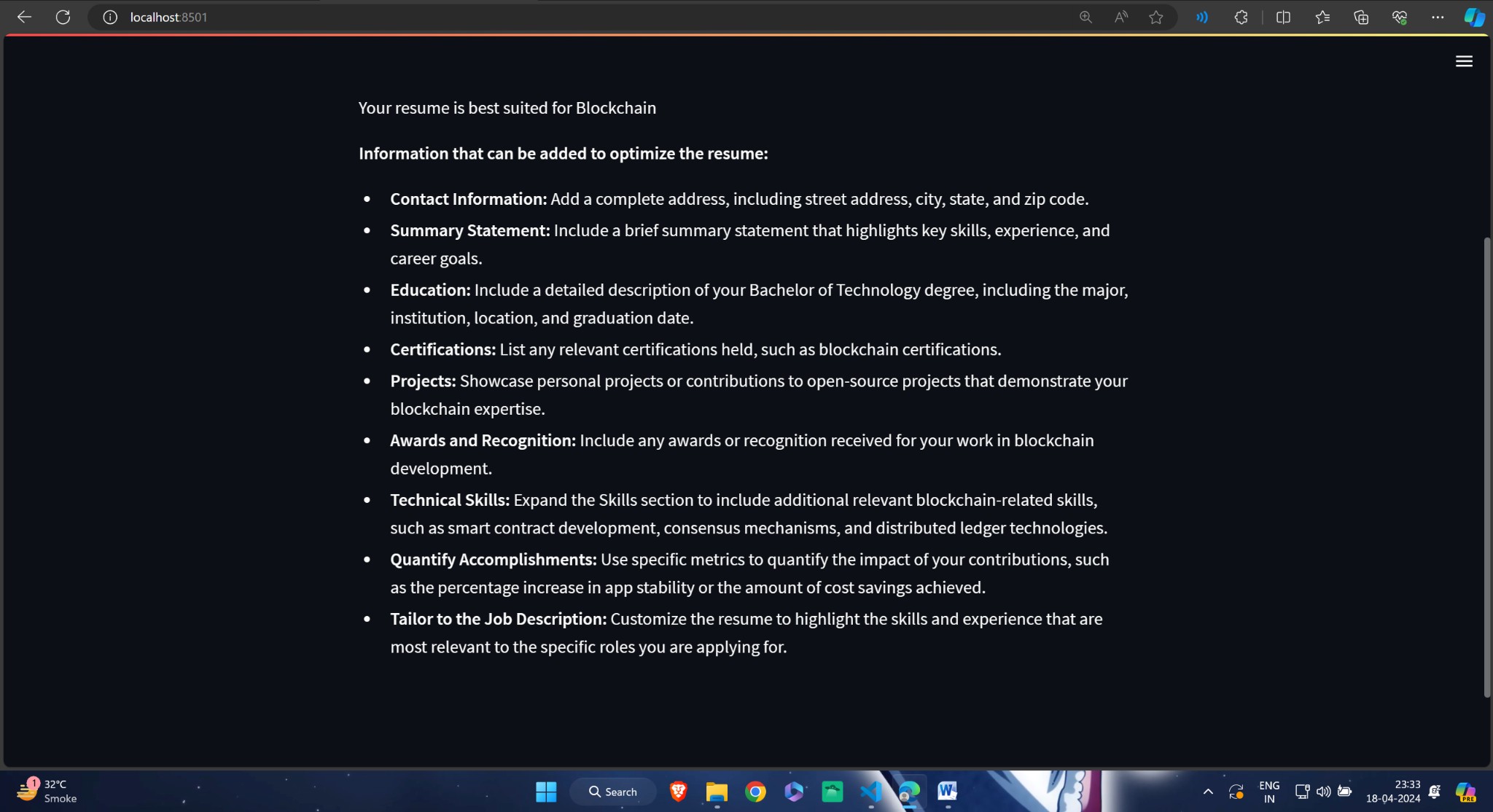
**Link To Dataset**

<https://www.kaggle.com/datasets/serkanp/resume-screening-dataset>

# OUTPUT







# CONCLUSION

In conclusion, the AI Resume Screening Application significantly enhances the recruitment process by utilizing the K-Nearest Neighbors (KNN) algorithm to accurately predict suitable job roles based on candidates' qualifications and experiences. The integration of Langchain allows for personalized resume feedback, empowering candidates to optimize their applications and increasing their chances of being shortlisted.

As the job market evolves, this application addresses the need for efficient screening solutions, setting a new standard for leveraging artificial intelligence in hiring practices. Future developments will focus on expanding features, enhancing data analytics, and integrating with job platforms, ultimately benefiting both recruiters and candidates.