

**Graphic Era**

**HILL UNIVERSITY**

**Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)**

**DEHRADUN CAMPUS**

**TERM WORK**

# MINI PROJECT

**B-Tech (CSE)**

# III- SEMESTER 2024-25

**SUBMITTED BY:**

Name: Pranjal Rawat

University Roll No:2319248

Class Roll No. : 51

Course / Section : B.Tech/B1

**SUBMITTED TO:** Dr.Prateek shrivastav

Dept. of Computer Sci. &Engg .

Graphic Era Hill University,

Dehradun.

A logo of a graphic era hill university

Description automatically generated

**Graphic Era**

**HILL UNIVERSITY**

**Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)**

**DEHRADUN CAMPUS**

**CERTIFICATE**

University Roll No. : 2319248 Class Roll No. : 51

This is to certify that \_\_\_ Pranjal Rawat\_\_\_ has satisfactory completed MINI PROJECT REPORT in this University. The course of the Mini Project in partial fulfillment of thus requirements at the third semester of B.Tech. (CSE) prescribed by the Graphic Era Hill University during the year 2024-2025.

DR.Prateek Shrivastav

## Faculty Incharge H.O.D

**Examiner**

A logo of a graphic era hill university

Description automatically generated

**Graphic Era**

**HILL UNIVERSITY**

**Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)**

**DEHRADUN CAMPUS**

**ACKNOWLEDGEMENTS**

I would like to express my special thanks and gratitude to my Data Structure Lab teacher "Dr.Prateek Shrivastav" for her able guidance and support in completing my Mini Project work file for the III sem. Thank all the teachers, lab assistants who supported me i n completing the term work.

## Student’s Signature

SYNOPSIS

1. **Titile of the Project:**

* Automated Resume Classification Using Machine Learning

1. **Objective of the Project:**

* Assist recruiters by quickly and accurately categorizing resumes, allowing them to focus on the most relevant candidates for a particular role.
* Reduce manual effort in categorizing resumes by automating the process using machine learning.
* Match resumes to job descriptions using classification models, ensuring candidates are aligned with the right opportunities.

1. **Project Category:**

* Machine Learning

1. **Language and tool used:**

* Front End : Tinker
* Backend :Python and it libraries

1. **Structure of the project:**

* **Purpose:**

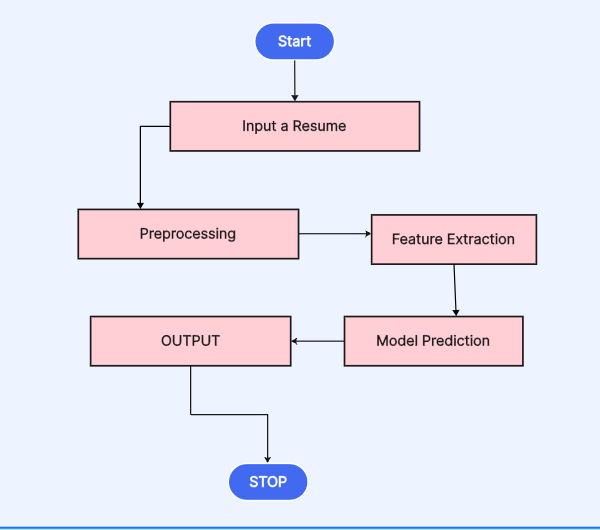
This project aims to provide a fast, accurate, and scalable solution for categorizing resumes into predefined job roles or categories, reducing the manual effort required by recruiters and improving hiring efficiency.

* **Project Description:**

User have to provide a resume/CV in .txt or PDF formate then this model predicts among the 25 job description which best suits to this resume.

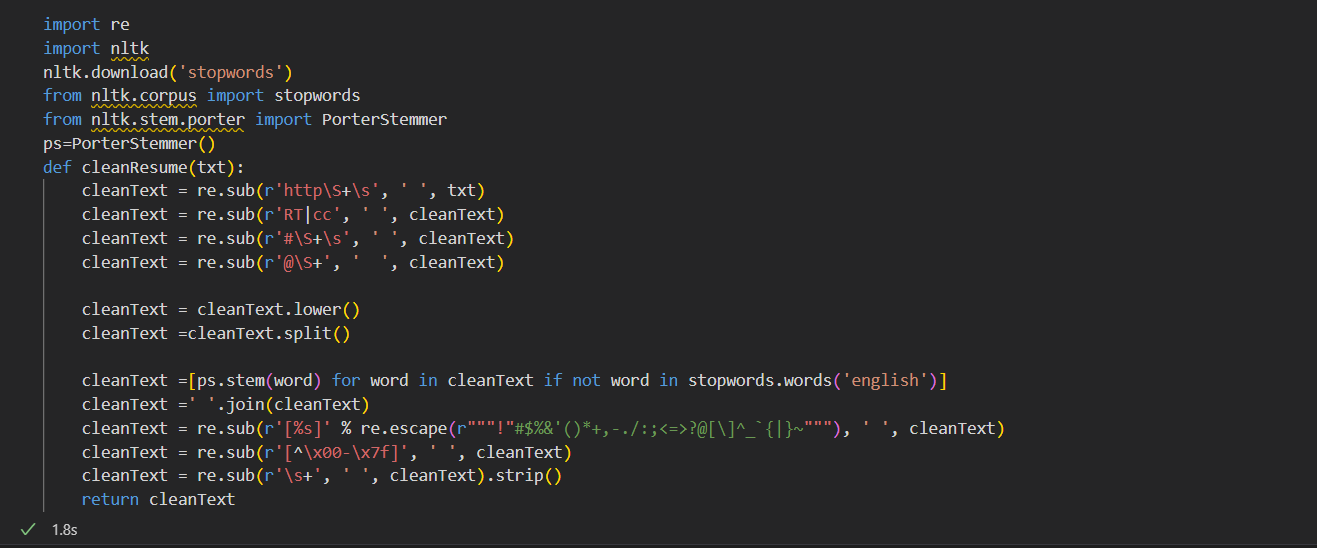
**How its works?**

**Design and Architecture**

****

1. **Resume Preprocessing:**

Uses advanced text preprocessing techniques to clean and standardize resume data. This includes removing URLs, stopwords, special characters, and stemming words to their base form.



1. **Text Vectorization:**

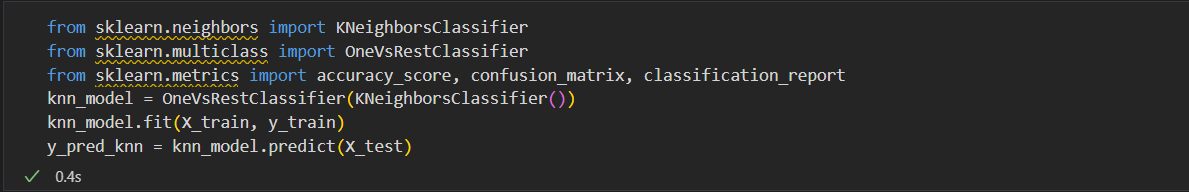
Converts unstructured text into numerical representations using TF-IDF (Term Frequency-Inverse Document Frequency), enabling machine learning models to process the data.

A black screen with colorful text

Description automatically generated

1. **Automated Classification:**

Implements a K-Nearest Neighbors (KNN) classifier to categorize resumes into job-related categories such as *Data Science, Marketing, Engineering,* etc.



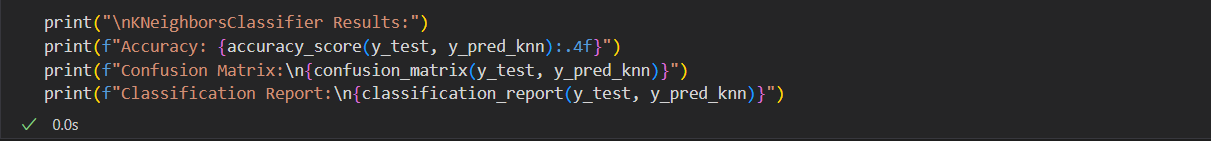
1. **Data Visualization:**

Provides visual insights into the distribution of resumes across categories using bar plots, pie charts, and count plots.



1. **Performance Evaluation:**

Evaluates the model's performance using metrics like accuracy, confusion matrix, and classification report to ensure reliability and effectiveness.



**Challenges and Solution**

**Challenges**

* The dataset may have noisy data (e.g., spelling errors, incomplete sentences and various symbols) or imbalanced categories, leading to biased predictions.
* High-quality labeled datasets for resume classification may not be collected , requiring manual effort to prepare data.

**Solution**

* **Data Cleaning:** Use techniques like spell correction, removing special characters, and stemming to clean the text data.
* Look for existing public datasets (eg Resume classifier Dataset from KAGGLE) that can serve as a base or starting point for training your model.

**FUTURE SCOPE OF THE PROJECT**

The future of the **"Automated Resume Classification Using Machine Learning"** project has significant potential for growth and innovation. With advancements in artificial intelligence (AI), natural language processing (NLP), and recruitment technology, this project can evolve to become an indispensable tool for modern hiring processes.

**References**

1.Previously built projects.(eg. Spam/Ham detection)

2. Python Official Documentation: https://docs.python.org/3/

3. Google.com to resolve syntax error.