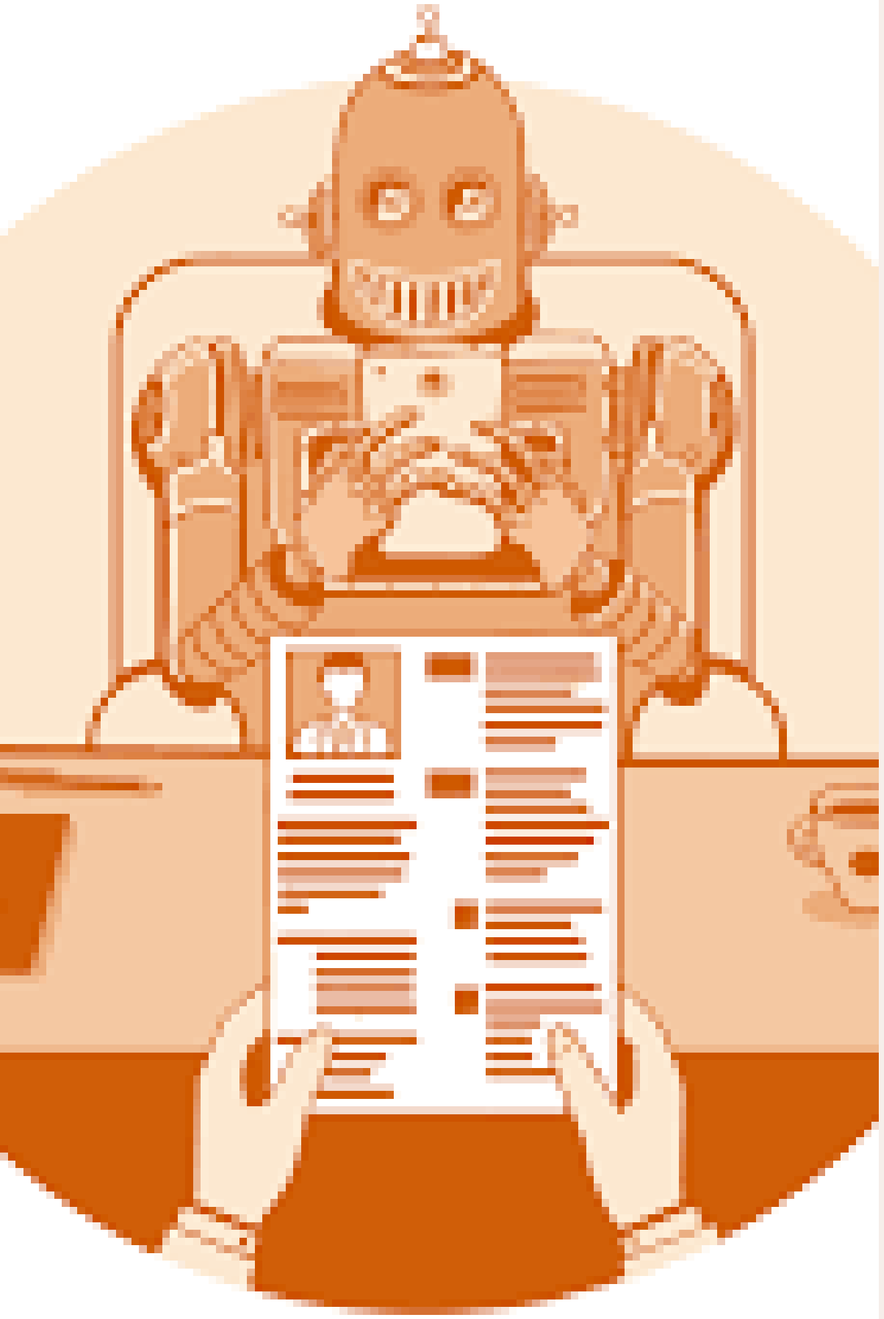
Automated Resume Filtering Using Machine Learning and NLP

AI Internship Project at Mentorness

NIDHI SHETTY

09-06-2024



INTRODUCTION

- Traditional resume screening is time-consuming and prone to human error.
- Objective: Develop an automated system for filtering and ranking resumes based on job descriptions using ML and NLP.

PROBLEM STATEMENT

- Manually reviewing resumes is inefficient and inconsistent.
- Need for a solution that can quickly and accurately filter resumes.

KEY FEATURES

Automated Database Search:
Identifies resumes matching job descriptions.

Intelligent Scoring Mechanism:
Ranks resumes based on relevance.

Efficiency:
Reduces manual effort required to review resumes.

Manual Verification:
Allows final stage manual review.

Advanced ML and NLP:
Uses TF-IDF vectorization and cosine similarity.

TECHNIQUES USED

- TF-IDF Vectorization: Converts text to numerical features.
- Cosine Similarity: Measures similarity between job descriptions and resumes.
- Natural Language Processing (NLP): Cleans and processes text data.

CODE OVERVIEW

- Libraries and Setup: Importing required libraries.
- Helper Functions: Extracting and cleaning text from PDFs.
- Directory Paths: Setting up directories for job descriptions and resumes.
- Text Extraction: Extracting and cleaning text from PDFs.
- Matching Resumes: Computing similarity scores and ranking resumes.

PERFORMANCE:

View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Code

```python
import os
import re
from PyPDF2 import PdfReader
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
import nltk
from nltk.corpus import stopwords

# Download NLTK stopwords
nltk.download('stopwords')

```

```python
# Function to extract text from PDF
def extract_text_from_pdf(file_path):
    reader = PdfReader(file_path)
```

CONCLUSION

- SUCCESSFULLY DEVELOPED AN AUTOMATED SYSTEM FOR RESUME FILTERING.
- LEVERAGED MACHINE LEARNING AND NLP TO ENHANCE RECRUITMENT EFFICIENCY.
- FUTURE IMPROVEMENTS COULD INCLUDE MORE SOPHISTICATED NLP TECHNIQUES AND A USER INTERFACE.

THANK YOU