

RESUME CATEGORY PREDICTION

PRESENTED BY

STUDENT NAME: VALLABH VINAYAK SANGAR

COLLEGE NAME:: DKTE SOCIETY'S TEXTILE AND
ENGINEERING INSTITUTE, ICHALKARANJI

DEPARTMENT: ARTIFICIAL INTELLIGENCE AND DATA
SCIENCE ENGINEERING

EMAIL ID: VALLABHSANGAR412@GMAIL.COM

AICTE STUDENT ID: STU662F3FD556D471714372565



VectorStock

VectorStock.com/25623021

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Manual resume screening is time-consuming and error-prone. Recruiters often struggle to categorize resumes accurately due to varying formats and content, especially with large application volumes.

PROPOSED SOLUTION

The proposed system aims to automate the classification of resumes into predefined job categories using machine learning and natural language processing (NLP). This addresses the challenge of manual filtering by providing fast, consistent, and scalable resume screening.

Data Collection:

- Utilized a labeled resume dataset containing resumes categorized into fields like Data Science, HR, DevOps, etc.
- Accepted resumes in various formats including PDF, DOCX, and TXT for real-world applicability.

Data Preprocessing:

- Extracted text using libraries: PyPDF2, python-docx, and encoding-handled text reads.
- Cleaned resumes to remove URLs, special characters, emojis, stopwords, and non-ASCII characters using regex.
- Applied lemmatization and case normalization to improve text consistency.

Machine Learning Algorithm:

- Used TF-IDF Vectorizer to transform the resume text into numerical feature vectors.
- Trained a Support Vector Machine (SVM) classifier for multi-class classification.
- Encoded the output labels using LabelEncoder for structured prediction.

Deployment:

- Developed an interactive web app using Streamlit, allowing users to:
- Upload resume files (PDF/DOCX/TXT)
- View extracted resume content
- Receive the predicted job category instantly
- Backend models (clf.pkl, tfidf.pkl, and encoder.pkl) are pre-trained and loaded using pickle.

Evaluation:

- Model performance evaluated using accuracy score and classification report during training.
- Real-time predictions tested with unseen resume samples.
- Future versions may include cross-validation, multi-label classification, or model ensembles.

a

SYSTEM APPROACH

The system uses:

- Python, Streamlit (for the web app)
- scikit-learn (for ML)
- TF-IDF for text vectorization
- SVM for classification
- PyPDF2, python-docx (for text extraction)

ALGORITHM & DEPLOYMENT

Algorithm Used:

Support Vector Machine (SVM) for classifying resumes into categories like HR, Data Science, DevOps, etc.

•Input Features:

Cleaned resume text (PDF, DOCX, TXT) transformed using **TF-IDF Vectorizer**.

•Training:

SVM model trained on labeled resume dataset.

TF-IDF and LabelEncoder used for text and label preprocessing.

Model saved as clf.pkl for deployment.

•Prediction Process:

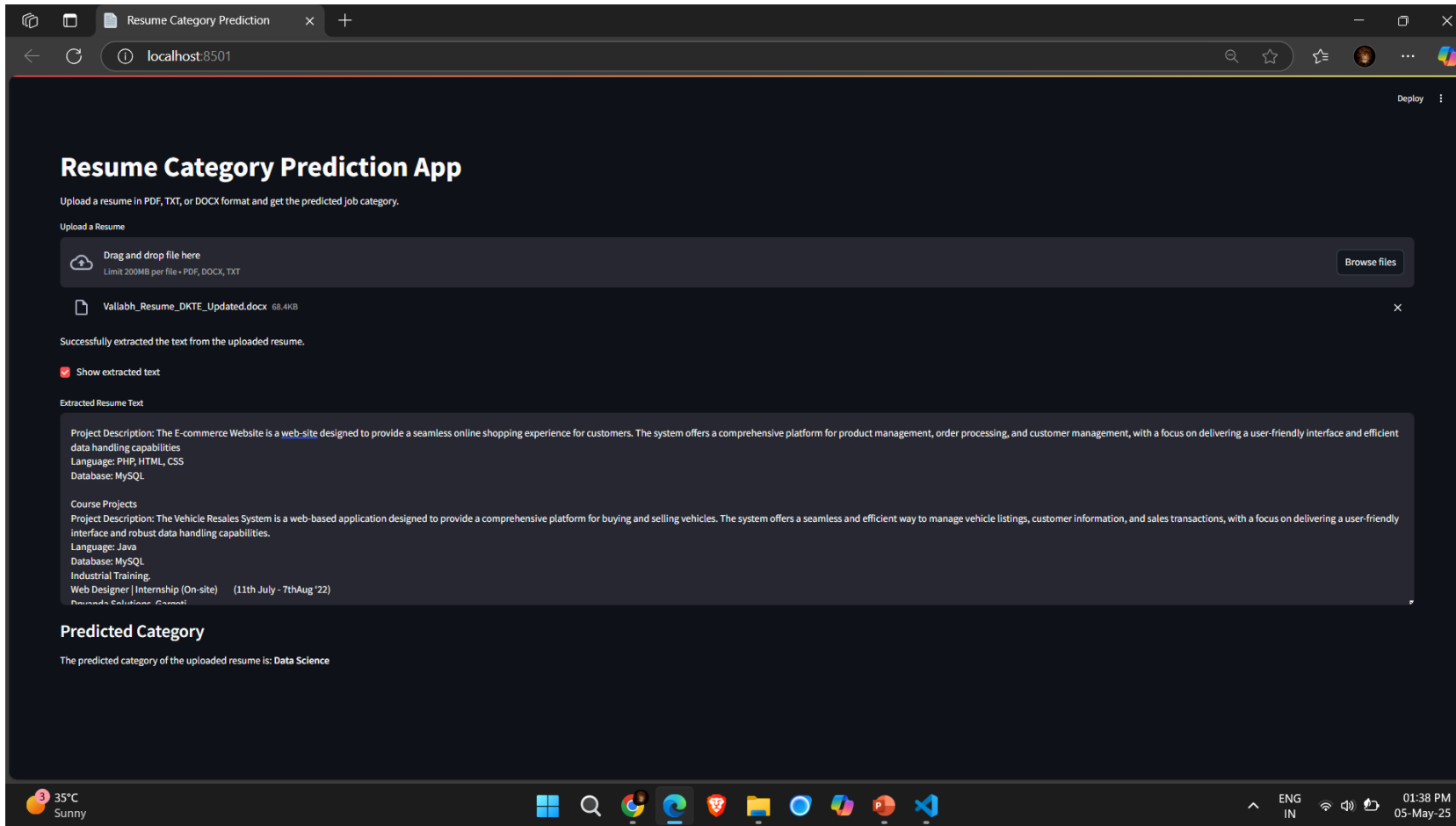
Resume → Text Extraction → Cleaning → TF-IDF → SVM → Category Prediction.

•Deployment:

Deployed using **Streamlit**.

Users upload resumes and receive real-time category predictions via a web app.

RESULT



CONCLUSION

The resume category prediction system leverages machine learning and NLP techniques to automate resume classification into predefined job categories. By implementing robust data preprocessing, training an accurate classifier, and deploying a user-friendly web application, the system significantly reduces manual effort and enhances hiring efficiency. With promising evaluation results, future improvements may include multi-label classification, cross-validation techniques, and ensemble models to further refine accuracy and scalability. This project presents a practical and scalable solution to streamline the resume screening process in recruitment.

FUTURE SCOPE

- **Enhanced Accuracy with Advanced Models:** Implement deep learning techniques like transformers and ensemble methods to improve classification precision.
- **Multi-Label Classification:** Enable resumes to be tagged with multiple relevant job categories, offering more nuanced classification.
- **Cross-Domain Learning:** Train the model using diverse datasets across industries to enhance adaptability.
- **Integration with Applicant Tracking Systems (ATS):** Seamlessly embed the classifier into recruitment platforms for automated screening.
- **Real-Time Recommendations:** Develop a feature that suggests job roles based on extracted resume content.
- **Scalability and Cloud Deployment:** Utilize cloud computing for handling large-scale resume data efficiently.
- **Bias Mitigation and Fairness:** Incorporate fairness-aware AI techniques to ensure unbiased categorization across demographic factors.

REFERENCES

GitHub Link: <https://github.com/vallabhsangar12/ResumeCategoryPrediction.git>

Thank you
