

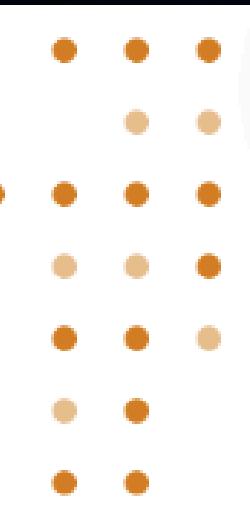
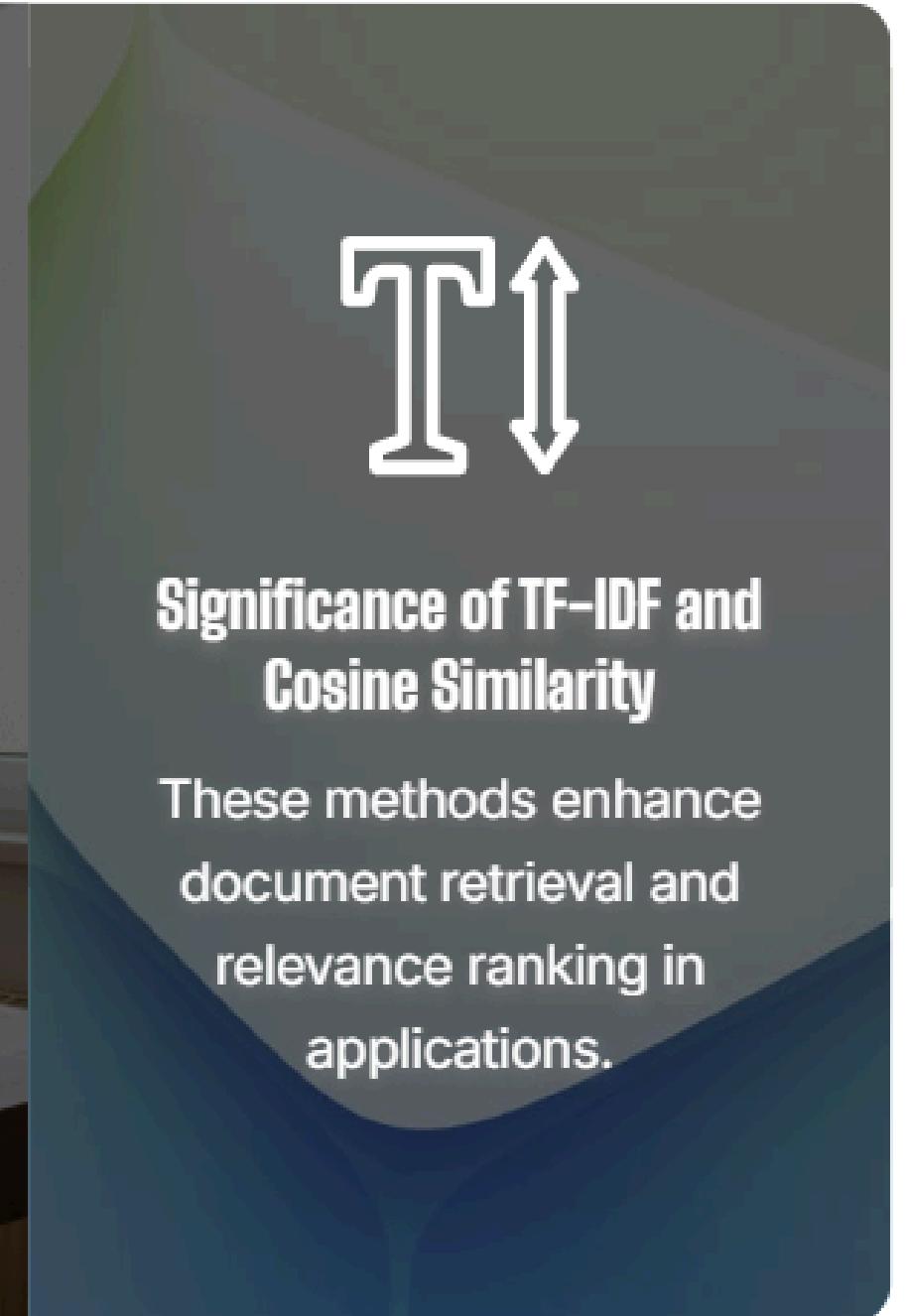
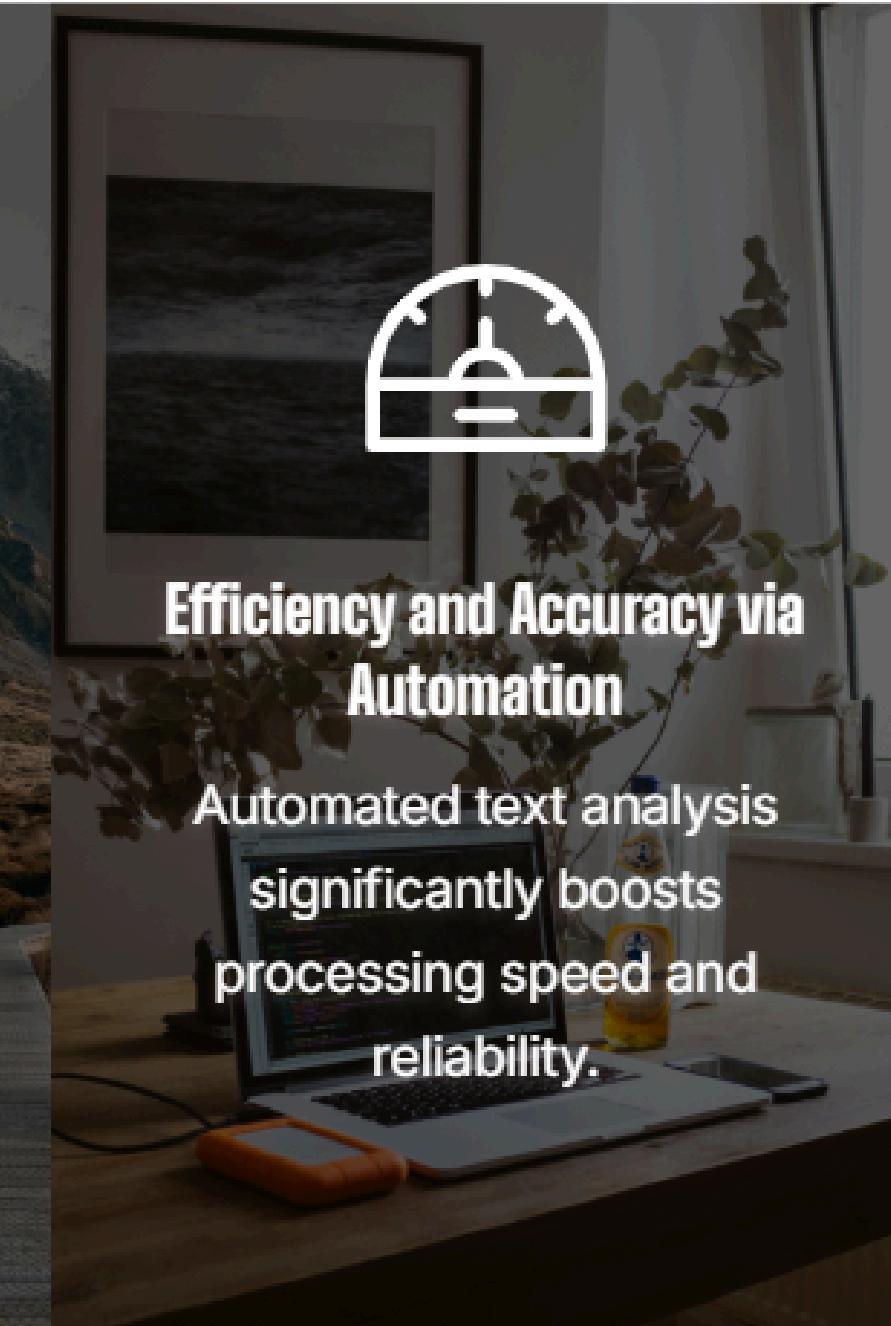
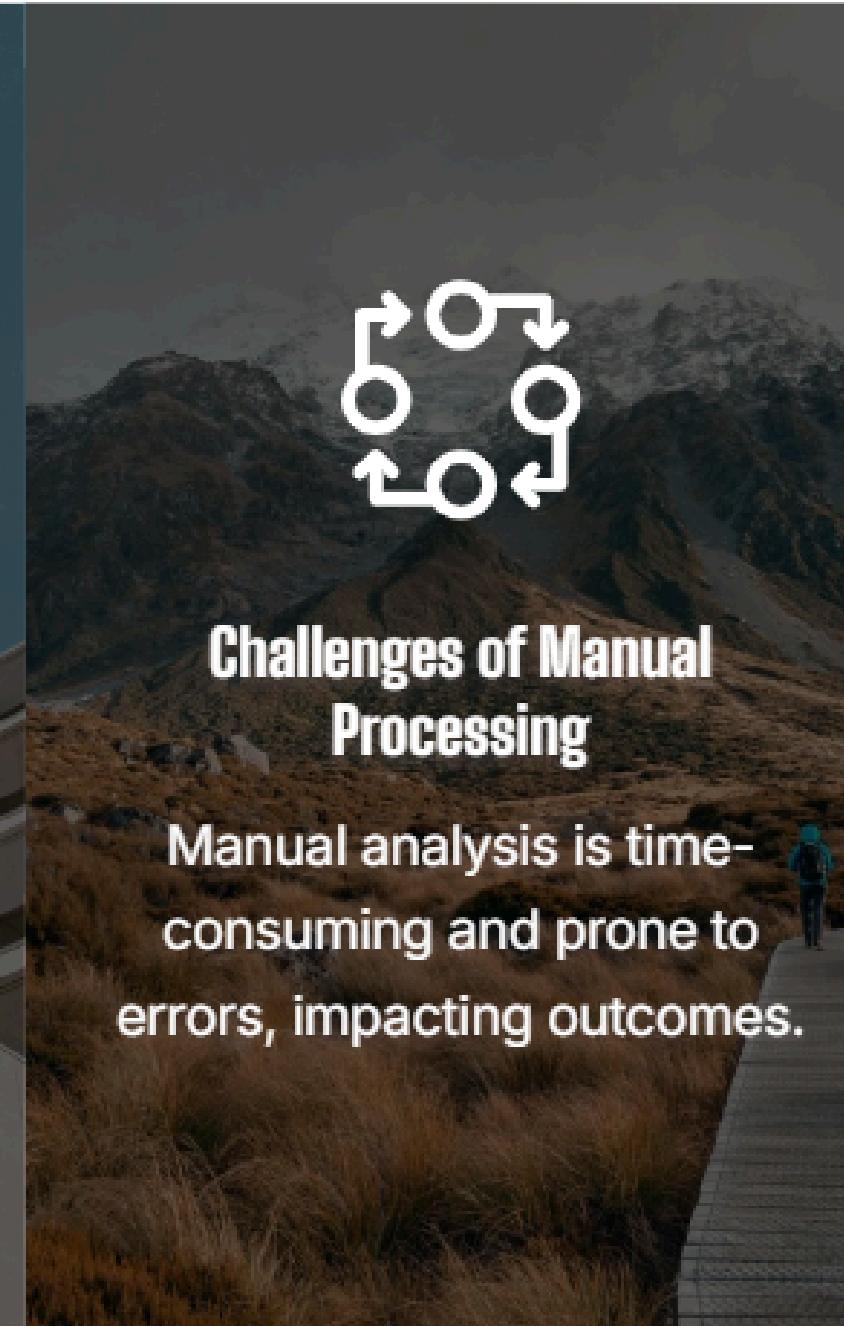
Automated Text Analysis in Resume Screening and Research Paper Classification

Meeting the demand for efficient text analysis in
resume screening & research paper classification

**PRADHUMN SHARDA
DHRUV JAIN
ASHTON MENDONCA**

The Growing Need for Automated Text Analysis

Enhancing Efficiency in Resume Screening and Research Classification



TEXT ANALYSIS TECHNOLOGIES

Key Technologies: TF-IDF and Cosine Similarity

Importance of TF-IDF and Cosine Similarity in Text Analysis



TF-IDF Algorithm

TF-IDF calculates the importance of a term in a document relative to a collection of documents. It helps in identifying the most relevant words in a document, essential for tasks like resume parsing and document classification.



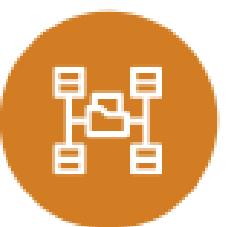
Cosine Similarity

Cosine similarity measures the similarity between documents by calculating the cosine of the angle between their feature vectors. It is crucial for assessing document relevance, aiding in tasks like resume screening and research paper classification.



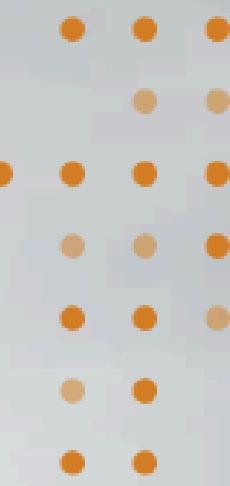
Resume Screening

Using TF-IDF and cosine similarity enhances resume screening by accurately matching candidate skills with job requirements, streamlining the recruitment process for HR departments.



Document Classification

TF-IDF and cosine similarity play a vital role in categorizing research papers efficiently based on content relevance, enabling quick retrieval and analysis in academic research.



TEXT ANALYSIS BENEFITS

Literature Review: Applications in Resume Screening

Utilizing TF-IDF and cosine similarity for enhanced text analysis in resume screening and research paper classification.

95%

Real-time Matching

- The application of TF-IDF and cosine similarity in resume screening has shown a 95% real-time matching rate, significantly improving the efficiency of candidate selection processes.

\$23

High Accuracy

- The use of TF-IDF and cosine similarity in resume parsing resulted in a cost reduction of \$23 per applicant due to the high accuracy achieved, minimizing manual efforts and errors.

TEXT DATA PROCESSING CHALLENGES

Problem Statement

Challenges in processing large volumes of text data and the need for automated systems in resume screening and document classification.

Efficient Text Data Processing

Processing large volumes of text data presents challenges in accuracy and manual effort reduction.



Automated Systems Importance

Automated systems are crucial to enhance accuracy and reduce manual effort in tasks like resume screening and document classification.



Resume Screening Challenges

Specific issues exist in resume screening due to the volume of data and the need for precise matching.



Document Classification Issues

Challenges arise in document classification processes requiring automation for efficiency and accuracy.



Text Analysis Methodology

Methodology Overview

Understanding TF-IDF and Cosine Similarity in Text Analysis

Feature Extraction with TF-IDF

Transforms text into numerical vectors for analysis.

Cosine Similarity Measurement

Evaluates relevance by comparing vector angles.

Text Preprocessing Steps

Includes tokenization, stop-word removal, and stemming.

TF-IDF Calculation Process

Involves computing term frequency and inverse document frequency.

Cosine Similarity Computation

Measures similarity between document vectors for relevance.



Research papers used

"NLP Based Resume Parser Analysis" by Mhaske et al. (2023). The proposed system uses TF-IDF for feature extraction and Support Vector Machine (SVM) for classification

"Using TF-IDF to Determine Word Relevance in Document Queries" by Juan Ramos (2003) The paper discusses how TF-IDF can effectively identify the most relevant words in a document by weighing term frequency against the inverse document frequency across a corpus.



Research papers used

"**Research Paper Classification Systems Based on TF-IDF and LDA Schemes**" by **Kim and Gil (2019)** introduces a hybrid system combining TF-IDF with Latent Dirichlet Allocation (LDA) for the classification of research papers.

“AN AUTOMATED RESUME SCREENING SYSTEM USING NATURAL LANGUAGE PROCESSING AND SIMILARITY ”
by **Chirag Daryania, Gurneet Singh Chhabrab, Harsh Patel , Indrajeet Kaur Chhabrad, Ruchi Patele (2020)**

The techniques used here include TF-IDF and cosine similarity (the methods we used for our project). Other than that, the text preprocessing techniques used here include tokenization, lemmatization, POS tagging, vectorization.



Research papers used

“Resume Classification System using Natural Language Processing and Machine Learning Techniques” by Irfan Ali, Nimra Mughal, Zahid Hussain Khan, Javed Ahmed integrates NLP with ML models. For NLP- BoW technique, N-Gram Model, and TF-IDF vectorizer were used to extract the features. Then ML models like SVM, logistic regression and Naive Bayes were used for classifying the resumes.



Conclusion

In conclusion, automating resume screening using TF-IDF and cosine similarity offers a robust and efficient approach to handling large volumes of applicant data. By leveraging TF-IDF, the system effectively captures the importance of keywords and their relevance within resumes, while cosine similarity provides a quantitative measure of how closely a resume aligns with job descriptions. This method not only streamlines the screening process but also enhances the accuracy of matching candidates with job requirements.

