# "SKILLSync: BRIDGING TALENT AND OPPORTUNITY - AN INTELLIGENT RESUME AND JOB DESCRIPTION MATCHING SYSTEM"

Archana Deshpande, Chaitanya Lakshmi Chava, Prasad Bankar, Shravan Sailada, Under Prof. Lu Xiao

OVERVIEW

- Crucial Python libraries, covering data analysis, text processing, NLP, and machine learning, are imported to establish a robust foundation for diverse tasks.
- Text preprocessing functions are implemented to clean and prepare data, with a focus on skills extraction. Techniques like label encoding and TF-IDF vectorization are employed for effective data transformation.
- The K-Nearest Neighbors (KNN) algorithm is utilized for classification, accompanied by thorough evaluation metrics, including accuracy and a detailed classification report.
- Functions for text extraction from PDFs and computing cosine similarity between vectorized resumes and job descriptions are designed, forming a practical job matching system.
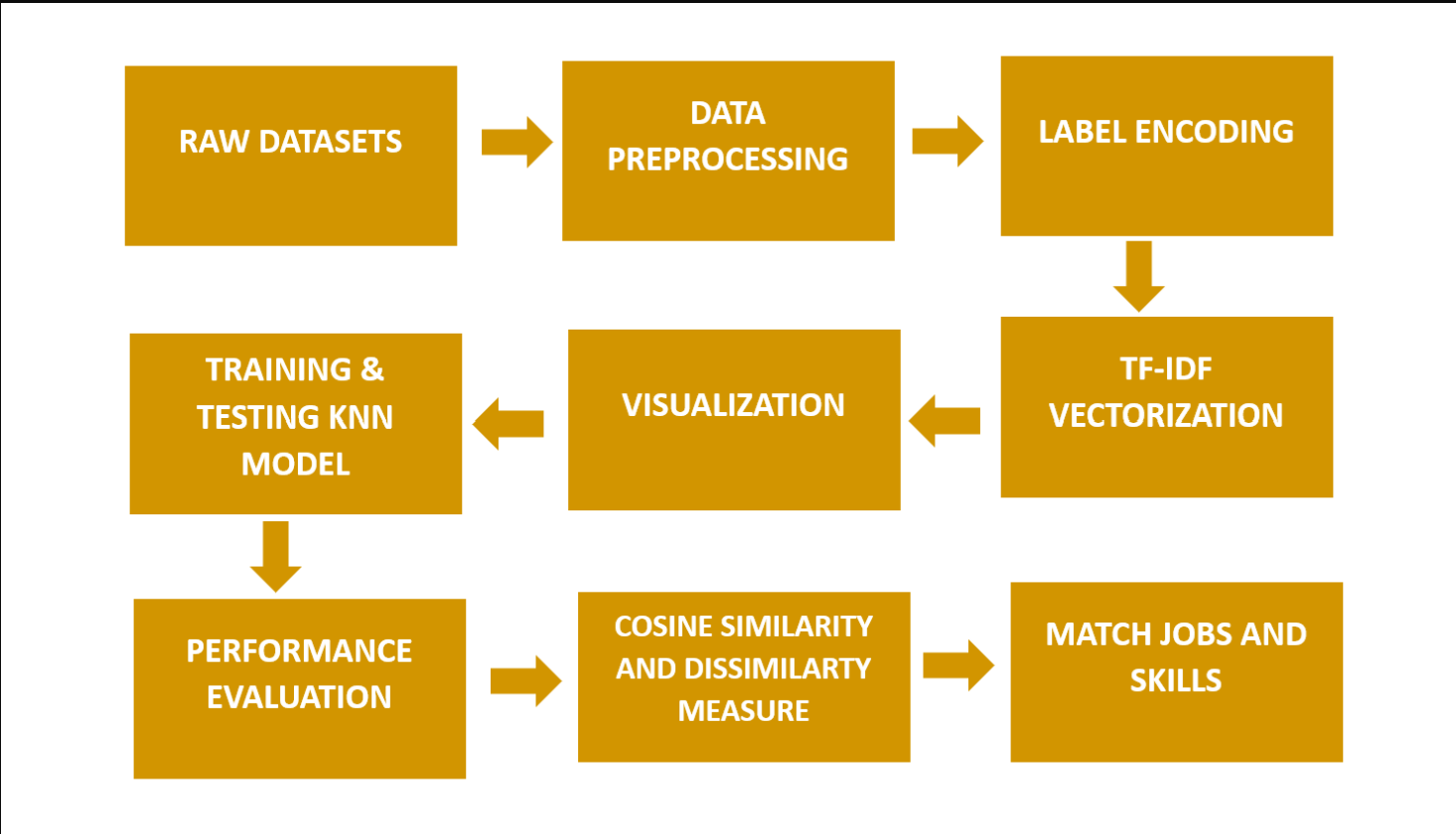
COMPREHENSIVE ANALYSIS

SkillSync, a powerful job matching tool, employs a robust text processing pipeline to extract, transform, and analyze information from resumes and job descriptions.

KEY STEPS

PDF Parsing

- SkillSync utilizes PyPDF2 for parsing PDF files, extracting text from each page.

Preprocessing

- Extracted text undergoes thorough preprocessing.
- Techniques include lowercasing, removing stop words, and handling special characters.

Vectorization Techniques

- **TF-IDF Vectorization:** The preprocessed text is transformed into TF-IDF vectors, capturing the importance of terms relative to a document collection.
- This technique enhances the representation of document content.

Count Vectorization

- A complementary technique to TF-IDF, Count Vectorization creates a matrix representing term frequencies in the document.

DATA VISUALIZATION

In the SkillSync project, data visualization plays a crucial role in providing insightful representations of key information. The primary visualizations include:

1. Top 20 Skills Distribution:

- SkillSync generates a bar chart using seaborn to visualize the distribution of the top 20 skills based on their occurrence counts.
- The chart offers a clear overview of the prevalence of skills in the dataset, helping users identify the most sought-after skills.
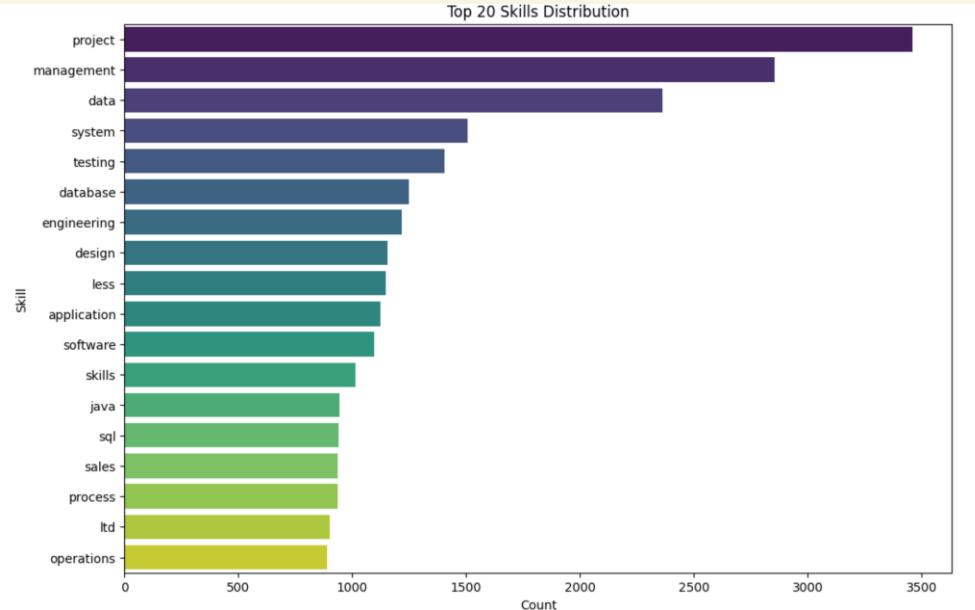
2. Resume Categories Distribution:

- A countplot is employed to visualize the distribution of resumes across different categories.
- Numeric category values on the y-axis are replaced with corresponding category names for improved interpretability.
- This visualization provides a comprehensive view of how resumes are distributed among various categories, aiding users in understanding the landscape of job opportunities.

KNN PERFORMANCE EVALUATION

At the heart of SkillSync's intelligent skill matching lies the K-Nearest Neighbors (KNN) algorithm, a robust foundation that underpins the platform's recommendation system. Through numerical representation and cosine similarity, KNN adeptly gauges the closeness of skills, allowing for precise matches between user profiles and job opportunities. This algorithmic core ensures that SkillSync not only identifies the most relevant job options but also tailors recommendations based on the unique skill portfolios of each user, offering a personalized and effective job-seeking experience.

```
--------------------------------
| Training Accuracy  :- 98.05% |
--------------------------------
| Validation Accuracy :- 93.26% |
--------------------------------
| Overall Model Score  :- 95.66% |
--------------------------------
```

TEXT SIMILARITY MEASUREMENT (COSINE SIMILARITY)

Cosine similarity is used to compare the similarity of different text documents. Cosine similarity is a widely-used metric in NLP for tasks like document similarity analysis, where the focus is on understanding how similar two pieces of text are in terms of their content or context.
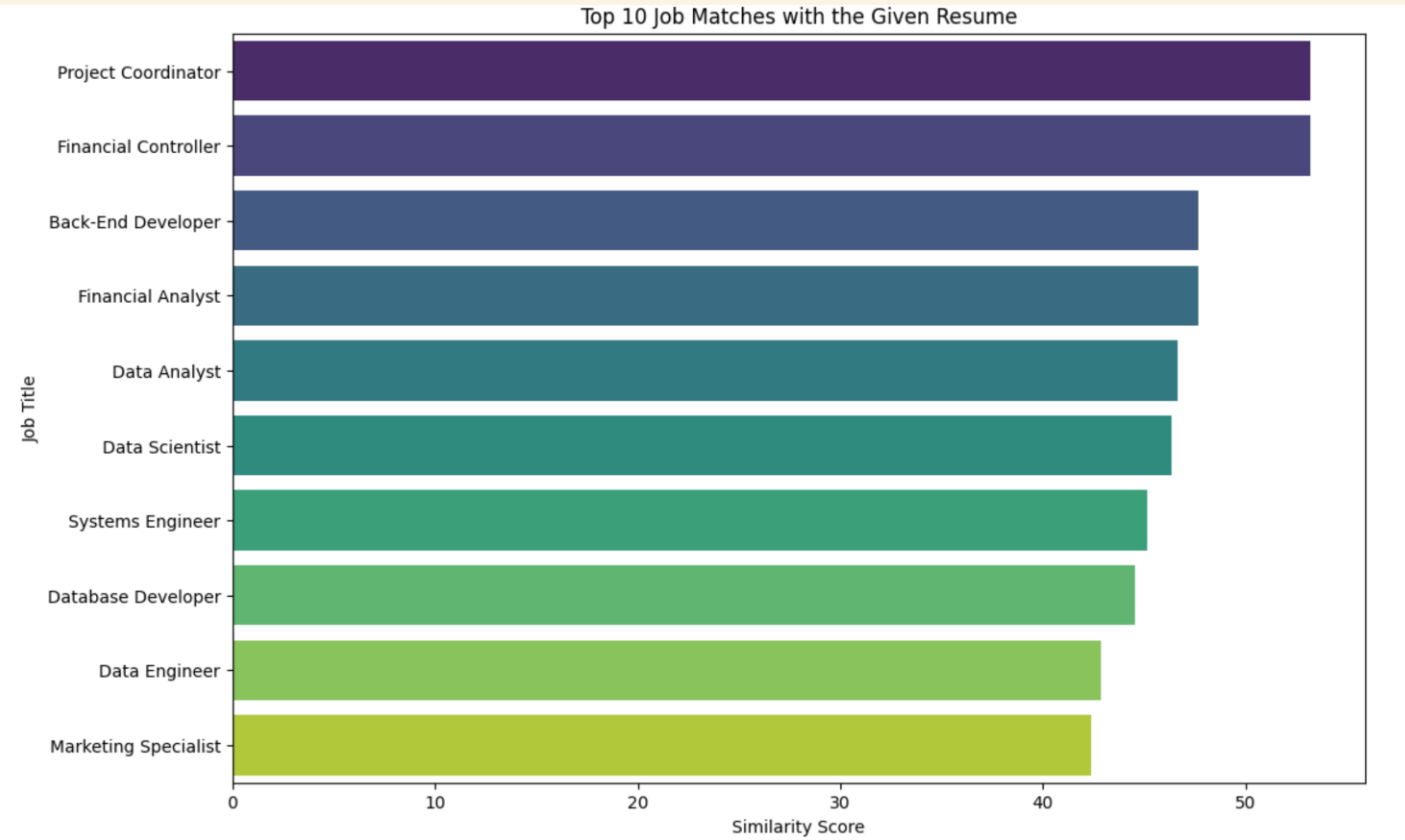

AUTOMATED SKILL IMPROVEMENT SUGGESTION

Generates a vivid histogram showcasing the top 10 missing skills between a resume and a job description. By calculating dissimilarity percentages, it identifies and visually emphasizes crucial skills lacking in the resume based on job requirements. This offers actionable guidance for users to enhance their resumes, ensuring a more effective alignment with job prerequisites. The dynamic visual insight contributes to an overall improvement in the quality of the resume.
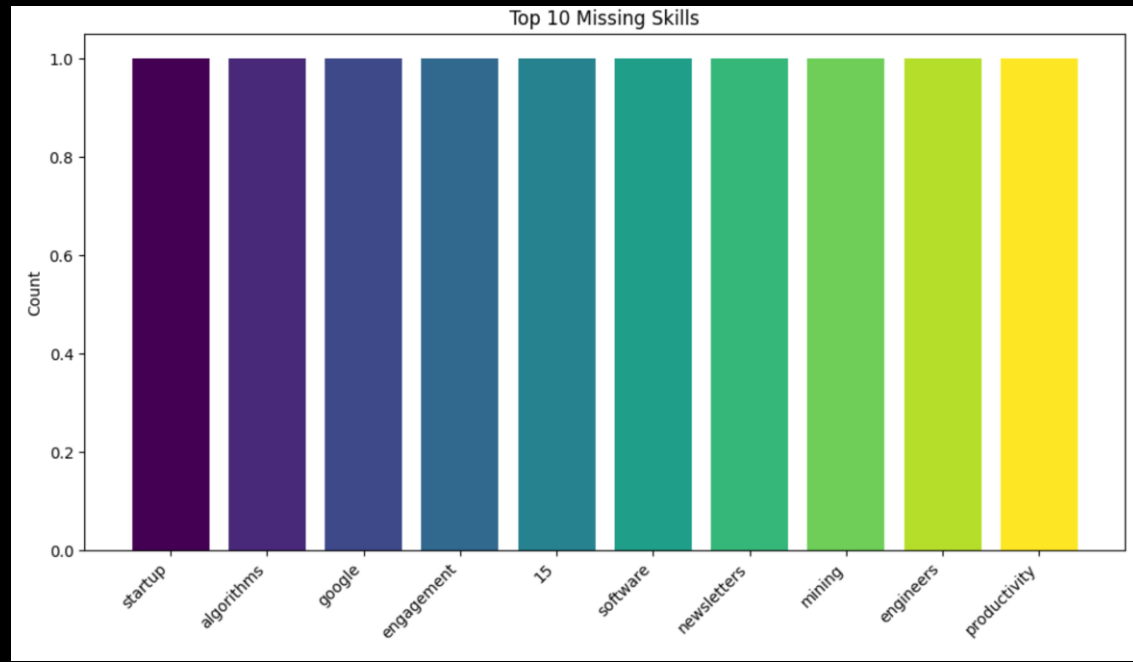
Skillsync also creates a heatmap that visually represents the presence or absence of each skill in a resume compared to a job description. Using distinct colors, it clearly distinguishes between matched and unmatched skills, providing a comprehensive overview of how well the resume aligns with the job requirements. The custom colors enhance the interpretability of the heatmap, offering a quick and clear snapshot of skill alignment.
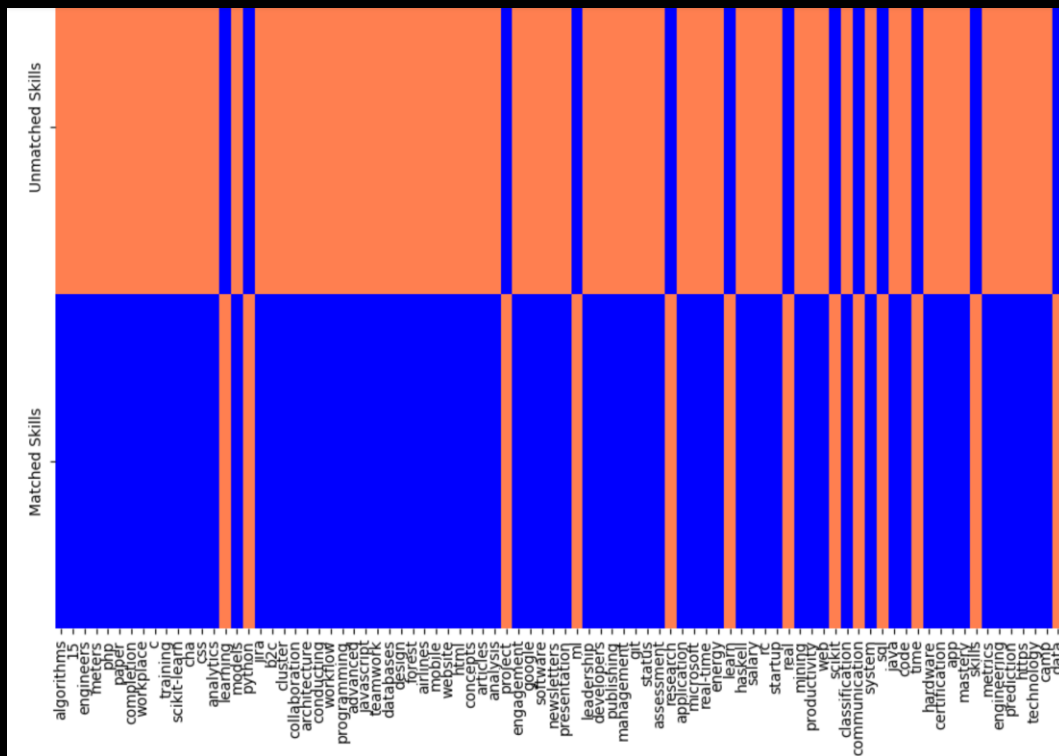
OPERATIONAL INSIGHTS

In managing Skillsync's operations, focus on scalability, real-time capabilities, and data privacy. Regular model updates, user feedback, and continuous monitoring ensure system relevance, user satisfaction, and optimal NLP performance. Prioritizing these aspects ensures a reliable and efficient platform for resume and job description analyses.

REFERENCES

1. Sajid et al., "Resume Parsing Framework for E-recruitment," 2022. [IEEE Xplore]
2. Ali et al., "Resume classification system using NLP and ML techniques," 2022. [Informit](https://search.informit.org/doi/10.3316/informit.263278216314684)
3. Bharadwaj et al., "Resume Analysis Using NLP," 2023. [Springer](https://doi.org/10.1007/978-981-99-1624-5_40)
4. Tejaswini K et al. (2022) achieved 85% parsing and 92% ranking accuracy using cosine similarity and KNN. [Global Transitions Proceedings](https://doi.org/10.1016/j.gltp.2021.10.002)
5. G. Rafiei et al. (2021) presented "Automating HR Recruiting" at the 5th NCAEA in Mashhad, Iran, achieving 43-47. [IEEE Xplore](https://doi.org/10.1109/NCAEA54556.2021.9690504)
6. P. Singh et al., "Automatic Candidature Selection by Artificial NLP," Springer, 2023. DOI