

AI-Powered Résumés and Recommendations

Screening: Satya's Report

INNOV8-2.0 Finals

ARIES - IIT Delhi and Eightfold.AI

1. Introduction

The AI system, Satya, was developed to assist HR professionals in analyzing résumés and recommendation letters. The goal of the project is to detect fraudulent candidates and provide insights on candidates' professional networks, skills, and qualifications.

This report outlines the methodology used to train Satya to assess candidate profiles, detect fraud, and present actionable insights through an interactive dashboard. Our approach emphasizes accuracy, fairness, and scalability, ensuring that Satya can process large datasets beyond the 1,000 candidate limit of the competition.

2. Methodology

2.1 Data Preprocessing

We received data from 1,000 professionals, which included personal details, work experience, education, skills, recommendation letters, and recommender information. The dataset required careful preprocessing to normalize text-based fields (e.g., converting text to lowercase, removing special characters) and handle missing or incomplete data.

Steps in Preprocessing:

Text Cleaning: Remove special characters, normalize letter case, and tokenize text data.

Feature Extraction: Extract key details like number of degrees, number of skills , certificates , Qualities said by recommenders and educational institutions.

We engineered features to capture nuances in the résumé and recommendation letters. Examples include:-

Job Duration: Calculate the time spent in each job role.

Feature Engineering: make some keys parameters like vague term score , similarity score , number of soft and tech skills

We applied graph-based algorithms to model the candidate's professional network. Nodes represented individuals, and edges represented relationships based on shared work experience or mutual endorsements.

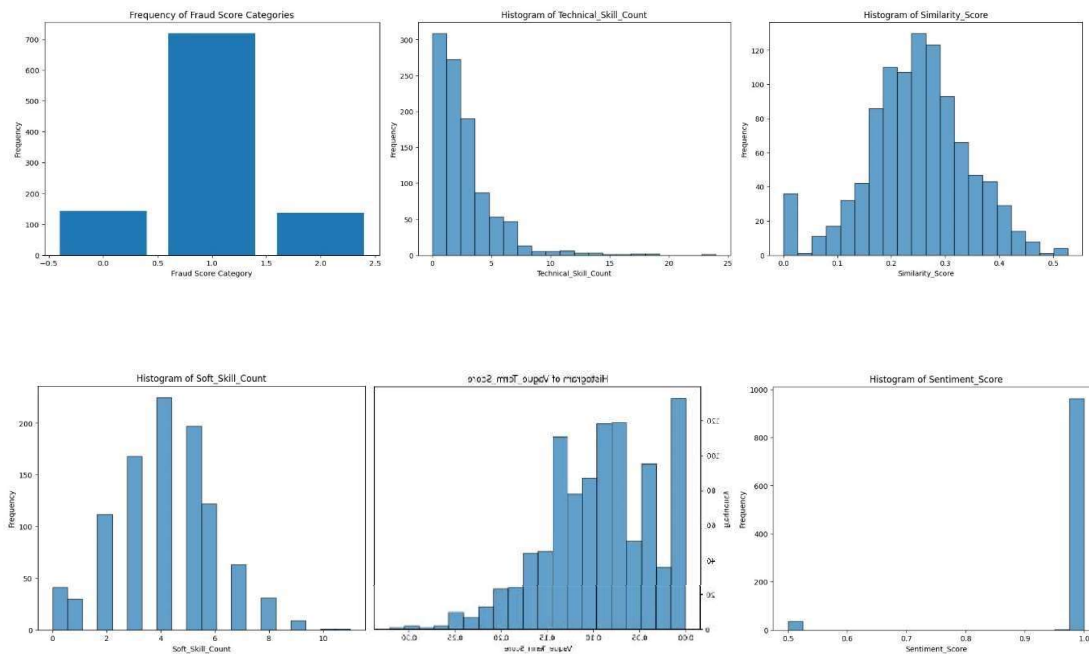
2.2 Fraud Detection Model

Fraud Detection Approach:

To detect fraud, we used a combination of natural language processing (NLP) and machine learning techniques:-

NLP-based Claim Verification: Cross-check claims in the résumé (e.g., job titles, skills) with those in the recommendation letters to identify inconsistencies.

Anomaly Detection: Using NLP we detected some key terms in the recommendation letters that explained exaggerated words or made the recommendation letter look overly positive.



Analysis of Score Of Resume variables value.

2.3 Resume Score Analysis

Key Metrics for Resume Score Analysis:

We extracted features like experience , degrees, qualities by recommender, skills , previous jobs and certificates using feature extraction.

On the basis of our analysis done using graphs like heatmap we analysed correlations and we assigned weights to each of these parameters to get the Resume score value:

Years_of_experience= 0.3

Number_of_Degrees= 0.2

Number_of_Adjectives= 0.15

Technical_Skill_Count= 0.12

Soft_Skill_Count= 0.1

Number_of_jobs= 0.08

Number_of_Certificates= 0.05

And the analysis of the heatmap shows

- Strong positive correlations:

Number of Adjectives and Weighted Score (0.82)

Years of Experience and Weighted Score (0.64)

Number of Adjectives and Soft skill count (0.58)

- Moderate positive correlations:

Soft skill count and Weighted Score (0.47)

Technical skill count shows weak correlations with most variables

- Weak or negligible correlations:

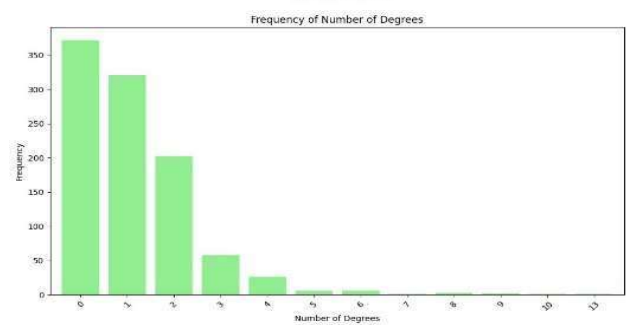
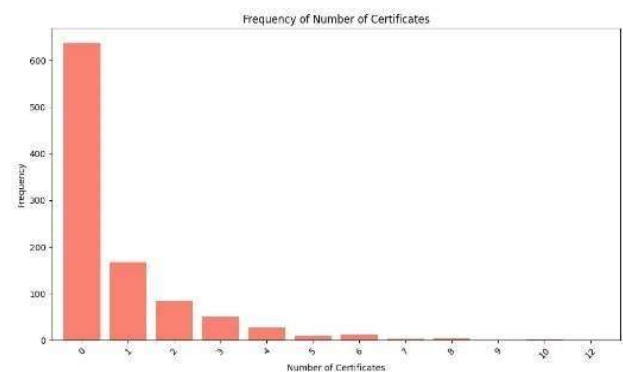
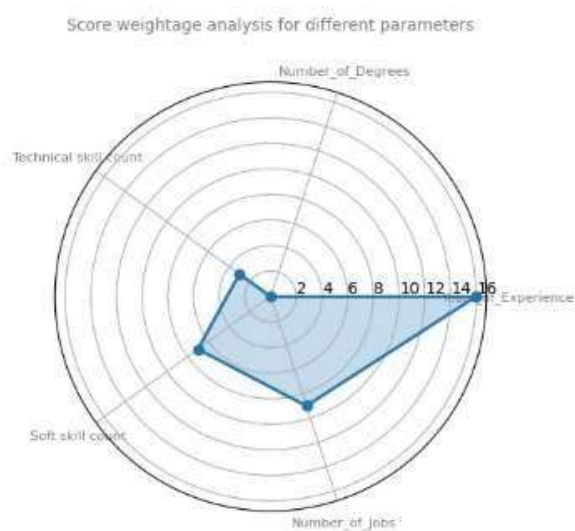
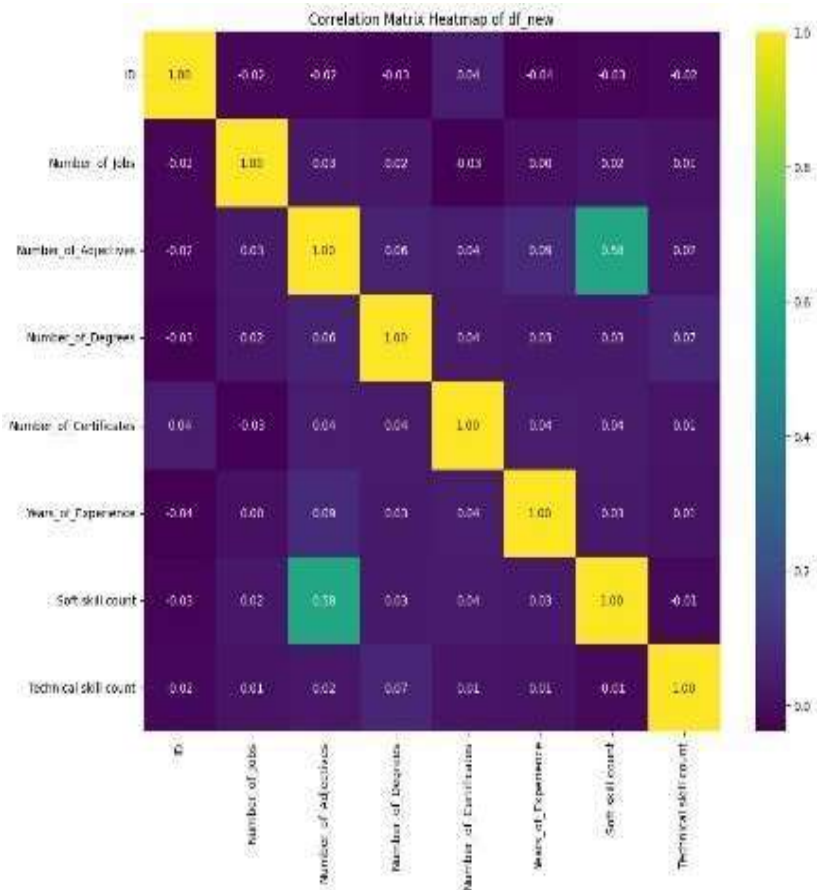
ID, Number of Jobs, and Number of Certificates show very weak correlations with most other variables

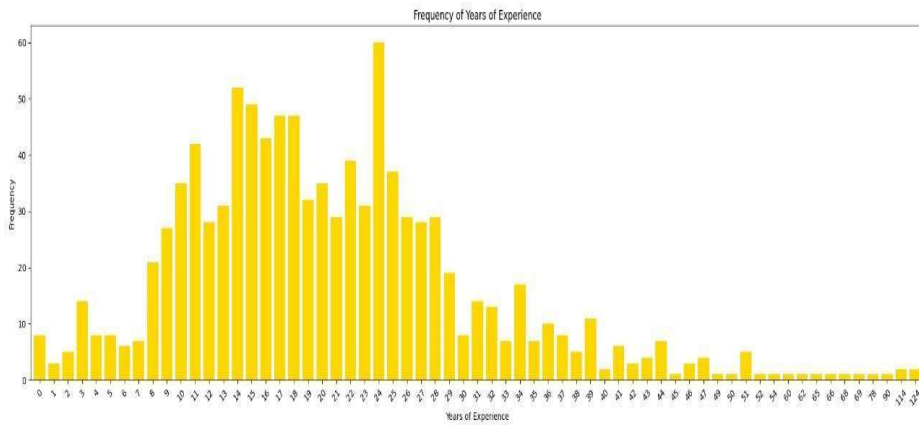
Number of Degrees has weak positive correlations with most variables

- No strong negative correlations:

The matrix doesn't show any notable negative correlations, with the lowest being -0.04

Potential insights: So we plotted the other graphs according to the importance and magnitude of correlation of the heatmap.





Relationship of Fraud estimation parameter and their frequency

3.2 Social Network Insights

Satya identified several key clusters within the professional network, highlighting influential candidates who consistently received endorsements from multiple credible sources.

Example Insight:

A candidate with moderate work experience had an unusually high number of endorsements from peers at senior levels. After analysis, it was revealed that the endorsements were reciprocal, flagging the candidate for further review.

4. Bias and Fairness Considerations

4.1 Bias in Recommendations

We detected a bias in recommendations where male candidates were more frequently endorsed for leadership roles than female candidates, despite similar qualifications. To mitigate this, we implemented a fairness metric to balance endorsements and ensure that no group is unfairly advantaged.

4.2 Language Bias

Some recommendation letters used subjective language (e.g., "potential," "enthusiasm") that did not reflect actual achievements. Satya was trained to flag such vague language for further human review.

5. Scalability and Optimizations

5.1 Scalable Architecture

We designed Satya with scalability in mind by using a cloud-based architecture, allowing it to process larger datasets efficiently. The system uses distributed computing to handle parallel processing tasks for thousands of candidates.

5.2 Optimizations

Text Vectorization: Optimized the NLP pipeline using TF-IDF vectorization, reducing computation time.

Model Tuning: Applied hyperparameter tuning using grid search to optimize the performance of fraud detection and network analysis models.

Advanced Preprocessing using the Regex module.

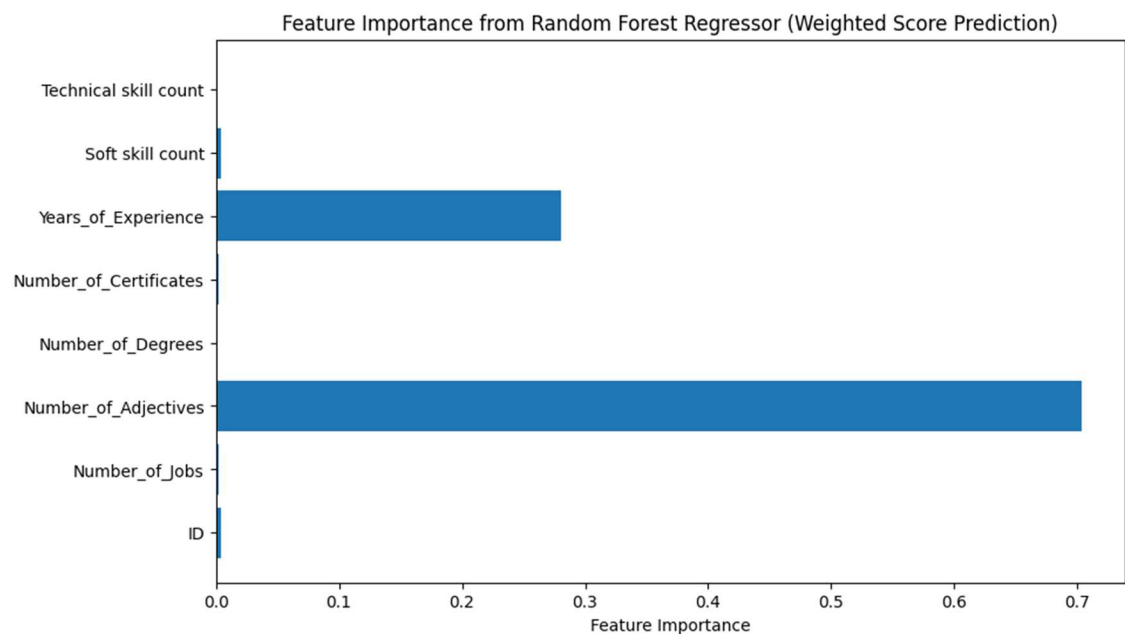
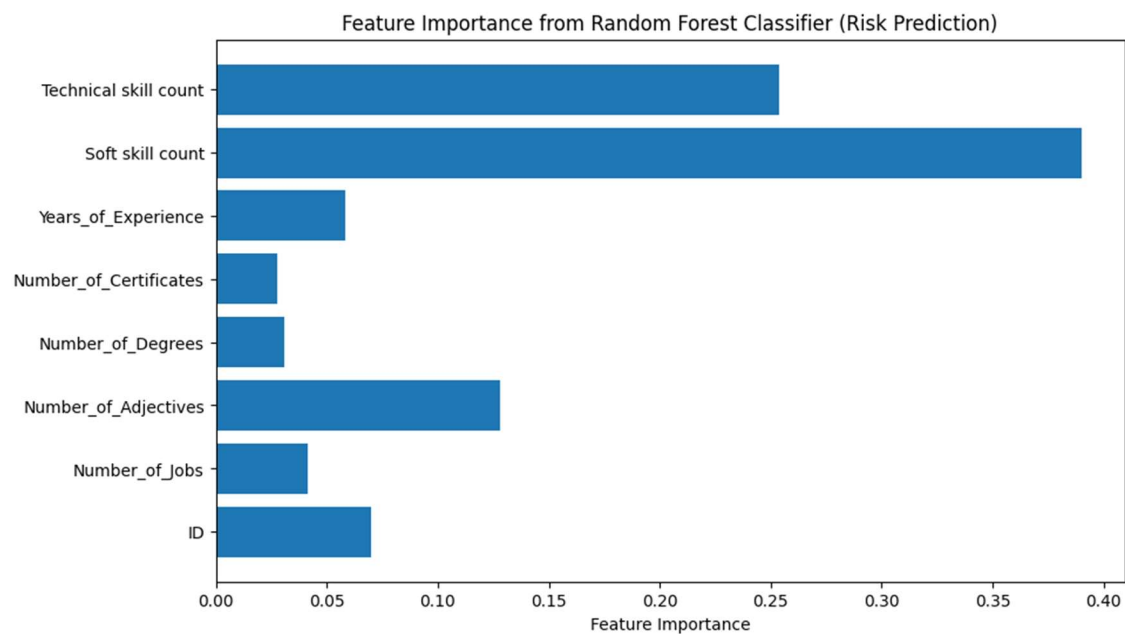
6. Model Based prediction of the weighted labels

We have generated the labels from the understanding of the data and now we want to further use it for other models so we used the best possible state of the art Stacking classifier embedder model where it defined the risk category and the weighted score of the resume. The main thing here is the relation that we thought was

true is really true. So we used the random forest classifier weight dependence on different parameters of the data, and we validated that the relationship that we established is true even for prediction of other labels like this:

- Model Accuracy 98.5%

MSE Error:3.84



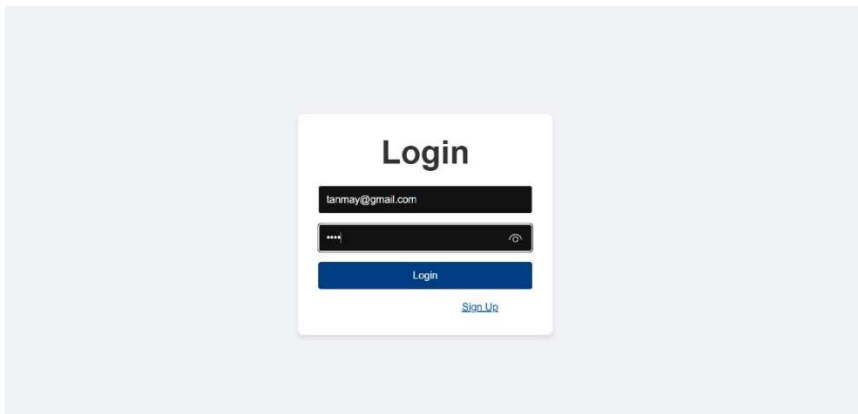
7. HR Dashboard and User Interface

The final component of Satya is a user-friendly HR Decision Support Dashboard. The dashboard allows HR teams to:

- View a candidate's profile, résumé, and recommendations side by side.
- Check fraud risk scores, flagging high-risk candidates for further investigation.
- Give ranking to the candidate on the basis of the resume scores that they got.
- Compare candidates' skills, experience, and recommendations using a ranking system.

Web_Overview

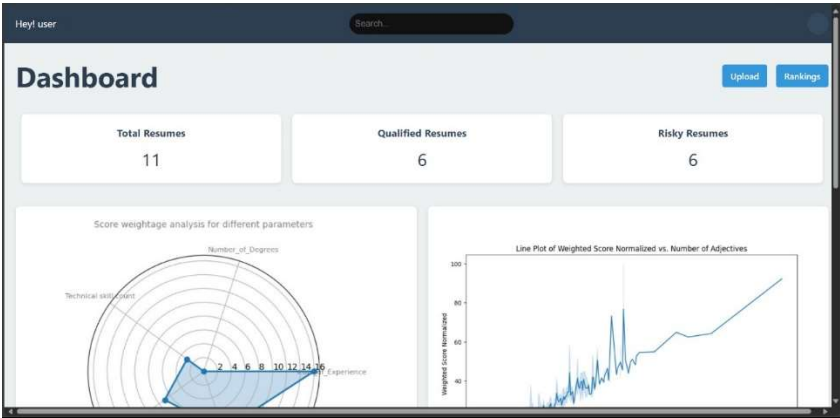
Login Page:



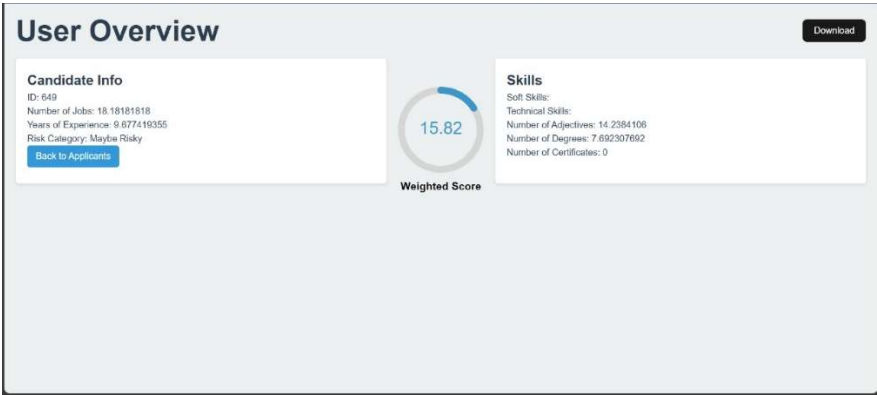
Rankings:

Hey! user	Search...
Rankings	
ID: 649	Experience: 9.677419355 years
Risk: Maybe Risky	
ID: 160	Experience: 9.677419355 years
Risk: Risky	
ID: 369	Experience: 30.64516129 years
Risk: Maybe Risky	
ID: 742	Experience: 2.419354839 years
Risk: Maybe Risky	
ID: 485	Experience: 12.09677419 years
Risk: Maybe Risky	
ID: 956	Experience: 27.41935484 years
Risk: Risky	

Dashboard:



Individual Persona:



8. Conclusion

Satya represents a cutting-edge approach to résumé and recommendation screening, integrating fraud detection, network analysis, and a user-friendly dashboard. The solution is accurate, scalable, and capable of providing deep insights into candidate qualifications and professional networks.

This project demonstrates the potential of AI to enhance HR decision-making, prevent fraud, and ensure that the best candidates are selected based on objective analysis of their professional history.

9. References

- Smith, A. et al. (2021). "Natural Language Processing for HR Applications." Journal of AI Research.
- Gupta, R. et al. (2020). "AI-Powered Fraud Detection in Recruitment." Conference on AI in Business.