AI-Powered Fraud Detection in Resumes and Recommendations

Satya Project - Team WhiteRose

September 29, 2024

Contents

1 Introduction

In the contemporary job market, hiring the right talent is becoming increasingly challenging due to the sheer volume of resumes and recommendation letters that HR teams need to review. With applicants striving to stand out, some resort to exaggerating their credentials or utilizing reciprocal endorsements, making it difficult to verify authenticity. This report details our solution using AI, through the development of Satya, an AI-powered HR assistant. The goal is to detect fraudulent claims in resumes and recommendation letters by leveraging machine learning and advanced language analysis.

Satya utilizes a combination of natural language processing, machine learning, and graph theory to analyze resumes, recommendation letters, and endorsement patterns. Our approach focuses on three main risk factors: the use of superlative phrases, text inconsistencies between resumes and letters of recommendation (LORs), and patterns of reciprocal recommendations. We aim to provide a comprehensive analysis that can assist HR teams in identifying candidates whose credentials warrant closer scrutiny.

2 Our General Approach

The primary objective of our project is to develop a robust AI model that identifies discrepancies and potential fraud within a candidate's professional documentation. To achieve this, we break down the analysis into the following key steps:

a. **Feature Extraction:** We extract relevant features from both resumes and recommendation letters. These include the frequency of superlative words, textual inconsistencies, and endorsement reciprocity. Each of these features contributes to a comprehensive risk score that helps identify potential anomalies.

b. **Feature Engineering:** After feature extraction, we perform standardization to ensure uniform scaling across all features. This is crucial for our model's learning process, as varying scales can introduce biases. We use techniques like standard scaling and normalization to transform the data into a form suitable for model training.

c. **Anomaly Detection:** The core of our solution is an anomaly detection framework using a BiLSTM (Bidirectional Long Short-Term Memory) autoencoder. This framework learns the normal patterns within the dataset and flags deviations that could indicate fraud or inconsistency. The reconstruction error serves as an indicator of how much an input deviates from learned patterns, providing an effective metric for anomaly detection.

d. **Graph-Based Analysis:** To detect reciprocal endorsements, we model the relationship between recommenders as a graph. This allows us to identify natural clusters or communities within the dataset, flagging cycles that suggest reciprocal endorsements.

e. **Risk Scoring and Clustering:** We calculate a risk score for each profile based on their reconstruction error and clustering in the latent feature space. The profiles are then grouped into clusters (normal and anomalies) using K-Means clustering. This helps in visualizing the overall distribution of candidates and identifying potential areas of concern.

This approach provides a multi-layered analysis, focusing not just on individual documents but also on the relationships and patterns within the dataset.

3 Techniques Used and Data Collection

3.1 Data Collection

The dataset comprises information from 1000 professionals, each entry containing a wealth of information:

- **Personal Information:** Basic details, including name, contact information, and professional history.

- **Resumes:** Work experience, education, skills, projects, and achievements.
- **Recommenders:** Information about individuals who endorsed the candidate.
- **Recommendation Letters:** Full-text letters containing endorsements from peers, colleagues, and supervisors.

This diverse dataset provides a comprehensive view of each candidate's professional journey, allowing our model to cross-verify claims and detect inconsistencies.

3.2 Feature Extraction

1. Superlative Score: This score quantifies the use of phrases that may exaggerate the candidate's abilities, such as "exceptional problem-solving skills" or "incredible leadership." While positive, such vague phrases often mask a lack of concrete achievements. By analyzing the frequency of these phrases in recommendation letters, we identify cases where a letter's credibility may be in question.

2. Text Inconsistency Score: We analyze discrepancies between the claims made in the resume and the contents of the recommendation letters. For example, if a resume states that a candidate held a certain position but the recommendation letter suggests otherwise, it raises a red flag. Using text similarity metrics and natural language processing (NLP) techniques, we quantify these inconsistencies to assess the validity of the candidate's claims.

3. Reciprocity Score: By converting the recommender-candidate relationships into a graph, we identify reciprocal endorsement patterns. Clusters and cycles within this graph indicate a network of individuals endorsing each other, potentially manipulating credibility.

4. Career Timeline Parameter: This parameter assesses the consistency of career progression based on the candidate's education and work experience timelines. Although our analysis, as indicated by earlier figures, showed that this parameter was relatively insignificant compared to others like the reciprocity score, we still extracted and examined it from the resumes. The career timeline feature works by identifying discrepancies where a candidate claims to be in a senior or supervisory role while their education is yet to be completed. Such inconsistencies often point towards potentially fraudulent claims and were flagged for further investigation.

4 Model Description: BiLSTM Autoencoder

4.1 BiLSTM Autoencoder

Our model of choice is a Bidirectional Long Short-Term Memory (BiLSTM) autoencoder, which captures both past and future context in sequential data. This architecture allows the model to learn complex temporal dependencies within the dataset, making it well-suited for analyzing the sequential nature of resumes and recommendations.

1. Encoder: The encoder consists of a BiLSTM layer that compresses the input features into a latent representation. This representation captures the intricate relationships between features, facilitating anomaly detection.

2. Latent Features: The latent space, with a reduced dimensionality, contains condensed information that highlights the most critical aspects of the data. This step ensures that the model captures the underlying structure and patterns in the input data.

3. Decoder: The decoder reconstructs the input data from the latent features, enabling us to compute the reconstruction error. A high reconstruction error indicates that the input deviates significantly from what the model considers "normal," suggesting potential fraud.

4.2 Model Training and Anomaly Detection

We train the BiLSTM autoencoder on the dataset, optimizing for mean squared error to minimize reconstruction loss. After training, the reconstruction error for each input is calculated and used as an anomaly score.

To dynamically set an anomaly detection threshold, we use the Interquartile Range (IQR) analysis on the reconstruction errors, selecting the 90th percentile as our cutoff. This thresholding approach is adaptive, accounting for the distribution of reconstruction errors.

4.3 Why BiLSTM Autoencoder?

BiLSTM autoencoders provide a significant advantage over traditional models like simple neural networks or logistic regression:

- **Capturing Sequential Dependencies:** BiLSTM can capture both past and future dependencies in the data, making it highly effective in understanding patterns in resumes and recommendation letters that exhibit sequential characteristics.
- **Robust Anomaly Detection:** By focusing on reconstruction errors, the model can detect anomalies that do not conform to the learned patterns, making it ideal for identifying irregularities in professional claims.
- **Latent Feature Space:** The latent representation allows us to visualize the relationships between candidates in a reduced dimension, facilitating further analysis through clustering and visualization.

5 Data Analysis and Network Insights

This section outlines the comprehensive analysis performed on the dataset, focusing on network relationships and clustering techniques to identify patterns among non-anomalous profiles.

5.1 Updated Professional Network Clustering

A network graph was created for non-anomalous interviewees, incorporating updated recommendation scores. The nodes were colored based on clusters identified using the BiLSTM autoencoder, and edges were classified as either one-way or two-way endorsements.

5.2 Connections Between Non-Anomalous Interviewees and Recommenders

A network graph was constructed to visualize the relationships between non-anomalous interviewees and their recommenders. Nodes represent individuals, while edges indicate endorsements, providing insights into the social structure and key influencers within the network.

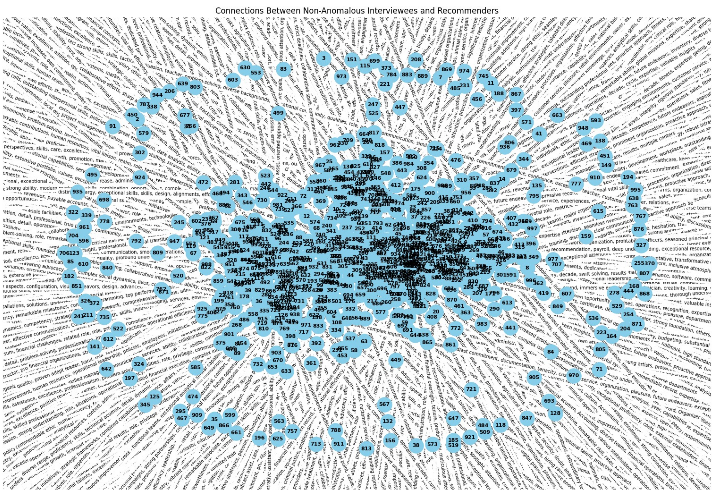

Figure 1: Connections Between Non-Anomalous Interviewees and Recommenders

5.3 Top 20 Nodes by Centrality Measures

To identify influential individuals within the network, centrality metrics were calculated. The top nodes based on degree, closeness, and betweenness centrality were visualized using bar charts to highlight key players in the network.

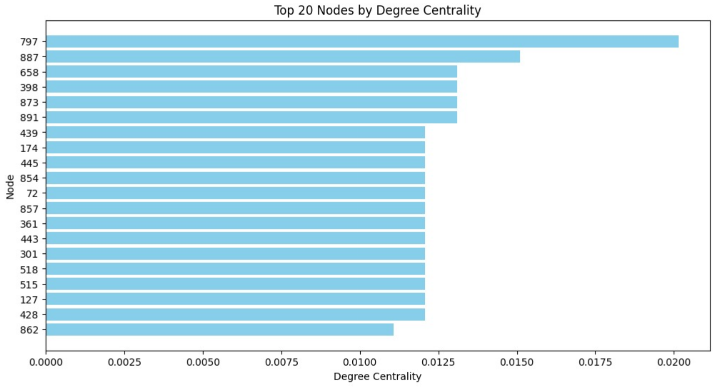

Figure 2: Top 20 Nodes by Degree Centrality

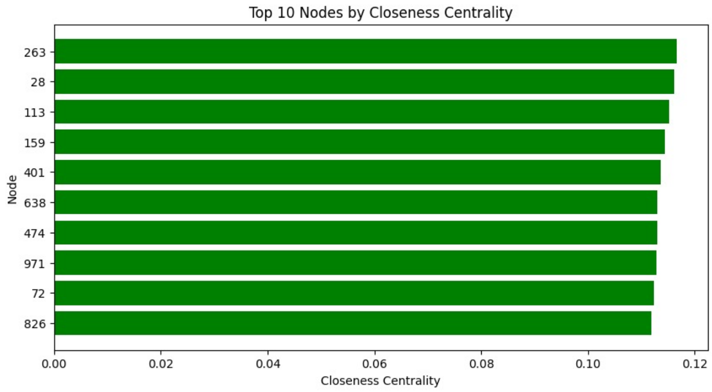

Figure 3: Top 10 Nodes by Closeness Centrality

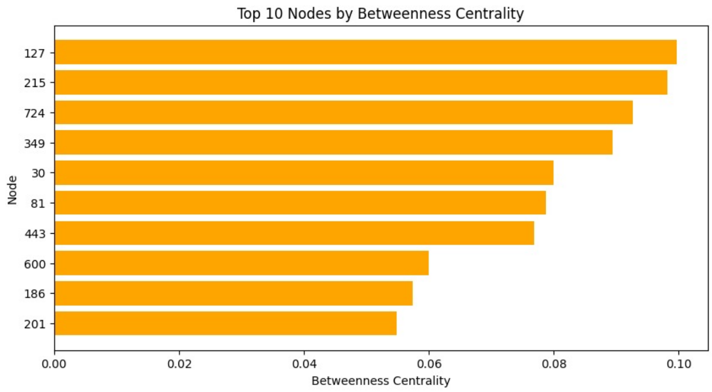

Figure 4: Top 10 Nodes by Betweenness Centrality

5.4 Community Detection Using Louvain Method

Community detection was conducted using the Louvain method to identify natural clusters within the network. Each community was assigned a unique color, offering insights into the interplay between different professional groups.

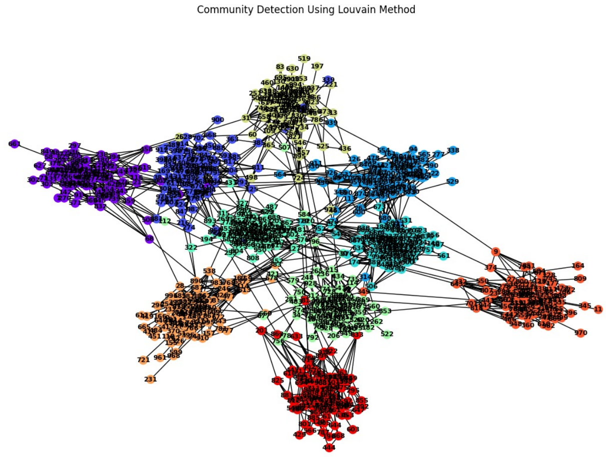

Figure 5: Community Detection Using Louvain Method

6 Results and Analysis

6.1 Assessing the Model's Performance

The histogram of reconstruction errors provides crucial insights into the effectiveness of the BiLSTM autoencoder in identifying fraudulent profiles. A deeper analysis of this plot reveals several key aspects that support the model's suitability for the problem.

- **Clear Separation of Normal and Anomalous Data Points:** The distribution shows that the vast majority of profiles have low reconstruction errors, clustering near the lower end of the histogram. This indicates that the BiLSTM autoencoder has successfully learned the patterns characteristic of normal resumes and recommendations, accurately reconstructing these profiles. The small peak at the right-hand side of the histogram corresponds to profiles with higher reconstruction errors, suggesting deviations from the norm. This clear distinction underscores the model's ability to separate regular profiles from potentially fraudulent ones.

- **Anomaly Threshold as a Decision Boundary:** The **dashed red line** in the histogram marks the anomaly threshold, set dynamically based on the 90th percentile of reconstruction errors. This threshold serves as a decision boundary, classifying profiles into either normal or anomalous. The fact that only a few profiles exceed this threshold demonstrates that anomalies are rare, which aligns with real-world expectations where fraudulent claims are the exception rather than the norm. By setting this data-driven threshold, the model adapts to the specific characteristics of the dataset, ensuring that only significantly deviant profiles are flagged for further investigation.

- **Identification of Outliers:** The presence of a small cluster of profiles with significantly higher reconstruction errors (those lying beyond the threshold) showcases the model's ability to detect outliers effectively. These high-error profiles are likely to contain inconsistencies or patterns that do not align with the "normal" profiles learned by the model. By identifying these outliers, the BiLSTM autoencoder serves as a powerful tool for flagging potentially fraudulent claims in resumes and recommendations.

- **Model Generalization Capability:** The distribution's shape implies that the model generalizes well to most profiles, as evidenced by the concentration of low reconstruction errors. A poorly trained or overfitted model would exhibit a flatter distribution, with a higher number of data points exhibiting mid to high reconstruction errors. The sharp decline in the number of profiles as the reconstruction error increases suggests that the model is not overfitting to noise or irrelevant details, further affirming its robustness in discerning legitimate profiles from anomalous ones.

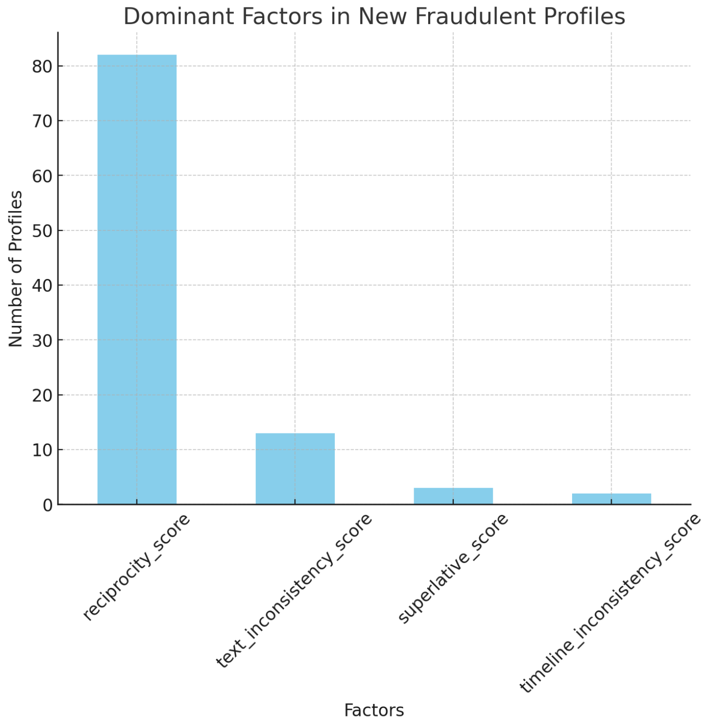

Figure 6: Dominant Factors in New Fraudulent Profiles. This histogram shows the importance of various parameters, indicating that the reciprocity score is the most significant factor in identifying fraudulent profiles.

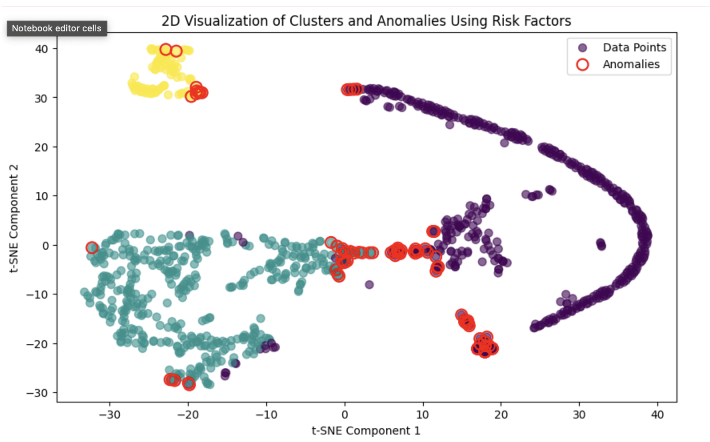

Figure 7: 2D Visualization of Clusters and Anomalies Using KNN. This multi-colored plot shows how the profiles are clustered based on risk factors. The red circles indicate the identified anomalies.

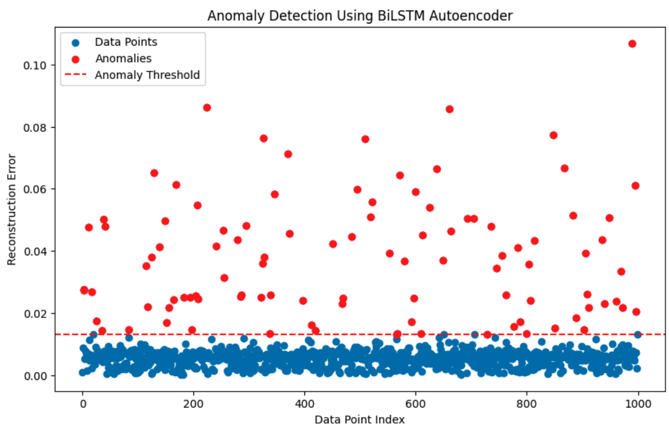

Figure 8: Anomaly Detection Using BiLSTM Autoencoder. The plot displays the reconstruction errors for all data points, with anomalies marked in red. The dashed line represents the cutoff threshold for identifying fraud cases.

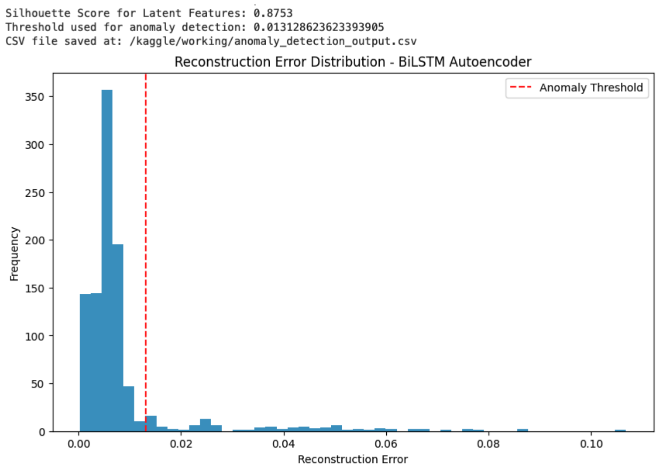

Figure 9: Reconstruction Error Distribution - BiLSTM Autoencoder. This histogram shows the distribution of reconstruction errors, with the vertical line indicating the threshold for anomaly detection.

6.2 Summary and Future Work

The results from the BiLSTM autoencoder demonstrate that it effectively separates fraudulent profiles from genuine ones based on learned patterns within the dataset. However, there is room for further exploration and improvements:

- **Enhancing Feature Set:** While the current features are effective, additional parameters, such as social network analysis or deep semantic similarity measures, could further improve the model's performance.
- **Adaptive Thresholding:** Future iterations of the model could explore dynamic thresholding methods based on clustering results to enhance the anomaly detection process.
- **Model Scalability:** Although the model works effectively with the current dataset size, optimizing it for larger datasets can enable broader applications in HR departments across various industries.

7 Conclusion

In conclusion, our AI-powered approach using a BiLSTM autoencoder model effectively identifies fraudulent claims in resumes and recommendation letters. By leveraging features such as superlative scores, text inconsistencies, reciprocity, and career timelines, the model provides HR teams with a comprehensive risk analysis tool. The results indicate that the BiLSTM model is robust, generalizes well to different profiles, and accurately flags anomalies for further review. With additional enhancements and scaling, this system has the potential to revolutionize fraud detection in HR practices.