Comparative Analysis of Unsupervised Anomaly Detection Methods for Fraud Identification

Alex Rizvanov[#1], Sri Krishna Koushik Komanduru[*2]

[#]*Data Science Department, Roux Institute at Northeastern*
100 Fore St, Portland, ME 04101

[1]rizvanov.a@northeastern.edu

[2]komanduru.s@northeastern.edu

Abstract**—** Fraud detection is a challenging problem due to extreme class imbalance, evolving fraud patterns, and limited availability of labeled data. These challenges motivate the use of unsupervised anomaly detection methods that do not rely on fraud labels during training. This project presents a comparative study of three unsupervised techniques for fraud identification: Principal Component Analysis (PCA) reconstruction error, neural autoencoder reconstruction error, and density-based clustering using HDBSCAN. Each method is applied independently to the same preprocessed dataset, producing anomaly scores or flags that are later evaluated using known fraud labels for validation purposes only. The goal of this study is to examine how different unsupervised modeling assumptions influence anomaly detection performance and fraud enrichment. This report describes the dataset, preprocessing pipeline, and methodological framework used in the analysis. Experimental results and comparative evaluation are deferred to future work.

Keywords**—- Unsupervised learning, anomaly detection, fraud detection, PCA, autoencoders, HDBSCAN, density-based clustering**

I. Introduction

Fraud detection remains a critical challenge in many real-world applications, including financial transactions, online marketplaces, and digital platforms. Fraudulent observations are typically rare, representing a small fraction of all records, which leads to severe class imbalance. On top of that, fraud patterns often change over time, reducing the effectiveness of supervised models trained on historical labels. Unsupervised anomaly detection is commonly applied in fraud settings where labeled data are scarce or unreliable [1]

Unsupervised anomaly detection provides an alternative path by identifying observations that deviate significantly from normal behavior without requiring labeled examples during training. (Rather than directly predicting fraud, these methods estimate the structure of typical data and flag atypical observations as potential anomalies.

This project explores three distinct unsupervised anomaly detection approaches: reconstruction-based linear modeling, reconstruction-based non-linear modeling, and density-based clustering. It then compares their effectiveness for fraud identification. By applying each method independently under a shared preprocessing pipeline, this study aims to highlight the strengths and limitations of each approach in a realistic fraud detection setting.

II. Literature review

Early research on online recruitment fraud detection has mainly focused on supervised machine learning approaches, where job postings are classified as legitimate or fraudulent using labeled datasets. Ravi Kumar et al. apply several supervised classification algorithms, including Logistic Regression, Support Vector Machines, Naïve Bayes, and Random Forests, and show that these models can achieve high accuracy on the EMSCAD dataset when sufficient labeled data are available [2]. Their results demonstrate that traditional classifiers can be effective for detecting known patterns of job fraud; however, these methods depend heavily on predefined labels and may struggle when fraud patterns change over time or when labeled data are limited.

To overcome some of these limitations, later studies introduced hybrid approaches that combine unsupervised learning with supervised classification. Kim et al. propose a hierarchical framework that uses unsupervised clustering and autoencoder-based feature learning before applying supervised deep neural networks, allowing the model to capture both broad and fine-grained fraud patterns [3]. This approach highlights the benefit of learning underlying data structure before classification, but it also increases model complexity and computational requirements.

More recently, transformer-based models have been proposed to improve text representation in fake job detection. Taneja et al. introduce Fraud-BERT, a context-aware transformer model that leverages transfer learning to better capture semantic relationships in job postings, especially in highly imbalanced datasets [4]. Their results show that contextual language models outperform traditional TF-IDF-based methods; however, such models typically require labeled data and significant computational resources, which may limit their practicality in some settings.

In contrast to these supervised and hybrid methods, Ch'ng and Wong demonstrate that fraudulent job postings exhibit recurring linguistic and structural patterns that can be identified using fully unsupervised exploratory analysis and clustering, without relying on labeled outcomes [4]. Their work suggests that job fraud can be treated as an anomaly detection problem, where fraudulent postings deviate from typical job listing behavior. Motivated by this progression in the literature, the present project adopts an unsupervised anomaly detection approach that focuses on identifying deviations rather than predicting predefined classes. This design choice allows the model to remain adaptable to evolving fraud strategies while maintaining interpretability and scalability.

Principal Component Analysis (PCA) has been extensively used for anomaly detection in high-dimensional settings, particularly through reconstruction-based approaches. By projecting data onto a lower-dimensional subspace defined by directions of maximum variance, PCA captures the dominant structure of normal observations. Anomalies are identified as instances that exhibit large reconstruction error when projected back into the original feature space. PCA reconstruction error is a widely used baseline for anomaly detection in high-dimensional data [5].

III. METHODOLOGY

A. Tools and Frameworks

This project was implemented using Python due to its extensive ecosystem of libraries for data science and machine learning. The pandas and NumPy libraries were used for data loading, cleaning, and numerical operations. scikit-learn was used extensively for preprocessing, feature extraction, dimensionality reduction, and classical machine learning techniques, including TF-IDF vectorization, Principal Component Analysis (PCA), and clustering algorithms.

For neural network–based models, TensorFlow (Keras API) was used to implement autoencoders for reconstruction-based anomaly detection. Visualization libraries such as Matplotlib and Seaborn were used to analyze distributions and model outputs. This combination of tools was selected because they are well-documented, widely adopted in academic and industry settings, and suitable for both classical and deep learning approaches.

B. Data and Preprocessing

1) Dataset Description

The dataset used in this project is the Employment Scam Aegean Dataset (EMSCAD), which is publicly available on Kaggle. It contains 17,880 job postings, of which 866 are labeled as fraudulent and 17,014 as legitimate, making the dataset highly imbalanced. Each job posting includes both structured and unstructured features.

Structured features include attributes such as employment type, required education, required experience, industry, and job function. Unstructured features consist of free-text fields such as job title, company profile, job description, requirements, and benefits. The presence of both text-based and structured features makes this dataset well suited for hybrid anomaly detection approaches.

2) Data Cleaning and Transformation

Initial preprocessing involved removing duplicate records and handling missing values. Non-informative or identifier-based columns were excluded from modeling. All text fields were concatenated into a single document per job posting to preserve contextual information across fields.

Text preprocessing steps included converting text to lowercase, removing punctuation, eliminating stop words, and tokenization. These steps reduce noise and improve the quality of the text representation. Structured numerical features were scaled using normalization to ensure that all features contributed equally to distance-based and reconstruction-based models.

3) Feature Engineering and Selection

Feature selection and engineering were guided by exploratory data analysis, domain relevance, and the requirements of unsupervised modeling, with an emphasis on interpretability and dimensionality control rather than optimizing a single predictive classifier. Structured numeric and binary features capturing posting legitimacy and structural characteristics were retained, including indicators of remote work, company branding, applicant interaction, and engineered length-based measures of descriptions, requirements, and company profiles. In addition, lightweight text-derived summary features, such as lexical richness of job descriptions, were engineered to capture linguistic patterns without introducing high dimensionality. Categorical variables representing job structure, seniority, education, industry, and geographic context were retained where cardinality was manageable, with grouping applied as an explicit feature engineering step to reduce sparsity. Features exhibiting high redundancy, extreme cardinality, inconsistent formatting, or substantial missingness were excluded or transformed, while raw text fields were preserved exclusively for downstream text vectorization rather than direct numeric inclusion.

4) Dimensionality Reduction

TF-IDF produces a high-dimensional feature space, which can negatively impact performance and computational efficiency. To address this, Truncated Singular Value Decomposition (SVD) was applied to the TF-IDF matrix to reduce dimensionality while preserving the most informative components.

C. Algorithms

1) PCA-Based Anomaly Detection (Structured Data)

Principal Component Analysis is a linear dimensionality reduction technique that projects high-dimensional data into a lower-dimensional subspace while preserving as much variance as possible. When used for anomaly detection, PCA functions as a reconstruction-based method in which data points that cannot be accurately reconstructed from the learned principal components are considered anomalous. Because the majority of job postings in the dataset are legitimate, PCA primarily learns patterns associated with normal behavior.

PCA was selected as a baseline due to its simplicity, interpretability, and computational efficiency. Because legitimate job postings dominate the dataset, PCA primarily learns their structure, causing fraudulent postings to produce higher reconstruction error.This method operates on structured numerical features derived from job posting metadata: encoded categorical variables and numerical attributes.

The primary tuning parameter in PCA is the number of principal components. Instead of selecting a fixed number, the model was tuned by retaining enough components to preserve 90–95% of total variance. This range balances information retention with noise reduction. Anomaly scores were computed using the mean squared reconstruction error for each sample

2) Autoencoder-Based Anomaly Detection (Structured Data)

An autoencoder is a neural network designed to reconstruct its input by passing it through a compressed latent representation. Unlike PCA, autoencoders are capable of modeling nonlinear relationships between features. In an anomaly detection setting, the model learns to reconstruct normal patterns in the data, while anomalous observations produce larger reconstruction errors. Autoencoders were chosen to capture complex, nonlinear interactions between structured features that PCA may fail to model. Fraudulent job postings may differ from legitimate ones through subtle combinations of attributes rather than single extreme values.

Hyperparameter tuning focused on the size of the latent representation, the depth of the network, and the learning dynamics of the training process. Smaller latent dimensions increased sensitivity to anomalies but risked underfitting, while larger latent spaces reduced the model's ability to distinguish between normal and abnormal observations. A balanced latent dimension was selected by analyzing reconstruction error separation between known fraudulent and legitimate postings.

The autoencoder was implemented using TensorFlow (Keras) with a symmetric encoder–decoder architecture. The model was trained using the Adam optimizer and mean squared error (MSE) loss. Early stopping was applied to reduce overfitting. Reconstruction error served as the anomaly score.

3) HDBSCAN-Based Anomaly Detection (Structured and Text Data)

HDBSCAN is a hierarchical density-based clustering algorithm that groups data points according to local density and identifies low-density points as noise. These noise points are interpreted as anomalies. Unlike traditional clustering algorithms, HDBSCAN does not require the number of clusters to be specified in advance and can identify clusters of varying shapes and densities.HDBSCAN was selected because fraudulent job postings are expected to be rare and structurally distinct, making them likely to appear in low-density regions of the feature space. Applying HDBSCAN to both structured data and text-based representations enables comparison of how anomalies manifest across different data modalities.

Fraudulent job postings are expected to be rare and irregular, making density-based clustering well suited for their detection. HDBSCAN was used on both structured and text data using TF-IDF features derived from job posting text. Applying HDBSCAN to both structured and textual representations allows evaluation of whether fraud is more strongly reflected in metadata irregularities or linguistic patterns.

Hyperparameter tuning focused on the minimum cluster size and the minimum number of samples required to form dense regions. Smaller parameter values increased sensitivity to anomalies but produced excessive noise, while larger values reduced the algorithm's ability to identify rare fraudulent postings. Multiple parameter settings were evaluated, and final values were selected based on cluster stability and the proportion of known fraudulent postings labeled as outliers.

IV. RESULTS

RECONSTRUCTION BASED RESULTS ON STRUCTURED DATA

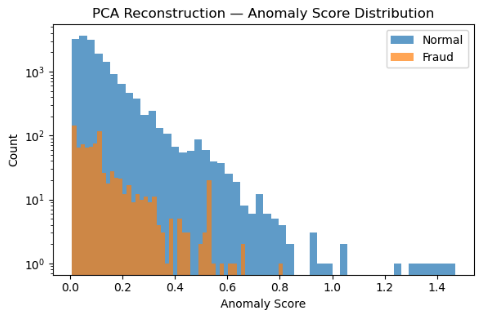

Figure 1. Distribution of PCA reconstruction anomaly scores for fraudulent and non-fraudulent observations.

Reconstruction-based methods, including PCA and an autoencoder, were applied to standardized structured features derived from job posting metadata. For both methods, anomaly scores were computed using reconstruction error, with larger values indicating greater deviation from learned normal patterns.

Substantial overlap is observed across nearly the entire range of anomaly scores (eg Figure 1). Fraudulent postings are not clearly isolated into a distinct region of the score space.While the autoencoder produces a wider range of reconstruction errors than PCA (eg Figure2), this increased spread does not translate into clear visual separation between fraudulent and legitimate postings.

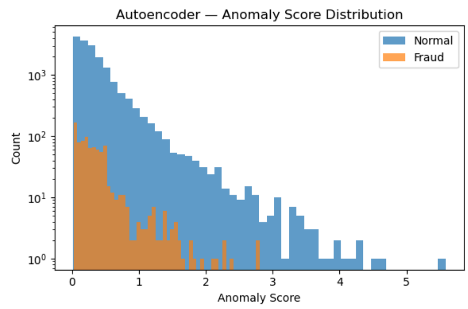

Figure 2. Distribution of PCA reconstruction anomaly scores for fraudulent and non-fraudulent observations.

DENSITY BASED RESULTS ON STRUCTURED DATA

HDBSCAN was applied to the same structured feature set to identify low-density observations labeled as noise. This resulted in a slightly better distribution with there being a large quantity of the listings being fraud after an anomaly score over 0.8 (eg Figure 3). However, both legitimate and fraudulent job postings are heavily concentrated at low anomaly score values, indicating that most observations reside in relatively dense regions of the structured feature space. While fraudulent postings appear across the full range of anomaly scores, they do not form a clearly separable group from legitimate postings.

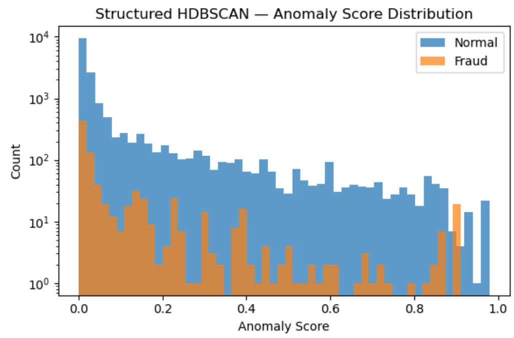

Figure 3. Distribution of HBSCAN anomaly scores for fraudulent and non-fraudulent observations.

Although some fraudulent postings attain higher anomaly scores, these regions are still dominated by legitimate postings in absolute count. This suggests that structural irregularity alone is insufficient to reliably distinguish fraudulent postings from normal ones using density-based clustering. Similar to the reconstruction-based methods, any threshold applied to the anomaly scores would involve a trade-off between identifying fraudulent postings and incorrectly labeling a large number of legitimate postings as anomalies.

Fraudulent postings were overrepresented among the noise points relative to their overall frequency in the dataset, indicating that density-based clustering is capable of isolating some structurally unusual postings. However, a substantial number of legitimate postings were also labeled as noise, particularly under more aggressive parameter settings. This reflects the sensitivity of density-based methods to hyperparameter choice and further highlights the difficulty of separating fraud based solely on structured attributes.

DENSITY BASED RESULTS ON TEXT-BASED FEATURES

While overlap between legitimate and fraudulent postings is still present, fraudulent postings display a clearer shift toward intermediate anomaly score values (eg Figure 4). In particular, fraudulent postings appear more frequently in the mid-range of the anomaly score distribution, whereas legitimate postings dominate both the very low and very high score regions. This pattern suggests that fraudulent job postings are linguistically distinct from typical postings, but not necessarily extreme outliers in the text feature space.

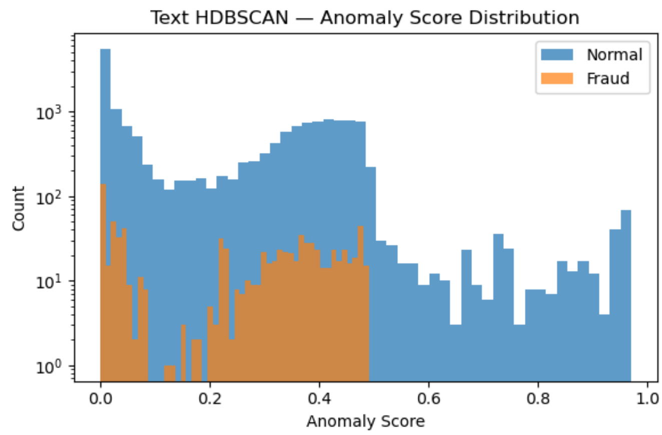

Figure 4. Distribution of HDBSCAN-text reconstruction anomaly scores for fraudulent and non-fraudulent observations.

In contrast to the structured data results, the text-based anomaly scores exhibit a noticeably different distributional pattern for fraudulent postings.

Unlike the structured HDBSCAN results, where fraudulent postings did not exhibit any consistent concentration relative to normal data, the text-based representation reveals regions of the anomaly score distribution where fraudulent postings are comparatively more prevalent. However, even in this case, legitimate postings remain numerically dominant across all score ranges, reflecting the strong class imbalance in the dataset.

The highest anomaly scores are not dominated by fraudulent postings, indicating that extreme linguistic sparsity or rarity does not reliably correspond to fraud. Instead, fraudulent postings tend to occupy moderately atypical regions of the text feature space, which may reflect deliberate attempts to mimic legitimate language while still exhibiting subtle inconsistencies or generic phrasing.

Comparison of Models

Across models, performance metrics indicate modest discrimination between fraudulent and non-fraudulent observations, with ROC-AUC values near 0.5 for most methods.

Table 1: Performance Comparison of Unsupervised Anomaly Detection Methods

Model	ROC-AUC	AP	Precision (k=500)	Mean Score	Std. Dev.
PCA Reconstruction	0.5186	0.0567	0.086	0.1095	0.1086
Structured HDBSCAN	0.5223	0.0603	0.076	0.0836	0.1702
Text HDBSCAN	0.5566	0.05169	0.008	0.212	0.2045
Autoencoder Reconstruction	0.4905	0.0471	0.058	0.3595	0.3938

Table 1. Reports ROC-AUC, average precision, Precision@500, and summary statistics of anomaly scores for PCA reconstruction, autoencoder reconstruction, and HDBSCAN.

Across methods, ROC-AUC values ranged from approximately 0.49 to 0.56. Average precision values were consistently low across all models, reflecting the extreme class imbalance present in the dataset. Precision@500 varied by method, with PCA reconstruction achieving the highest observed value, followed by structured HDBSCAN and the autoencoder. Text-based HDBSCAN yielded a Precision@500 of 0.008.

Anomaly score distributions differed across methods. PCA reconstruction scores were concentrated at lower values with a right-skewed tail, while autoencoder reconstruction scores exhibited a broader range and higher maximum values. HDBSCAN-based methods produced anomaly scores that were heavily concentrated near zero, with a small subset of observations assigned higher outlier scores. Differences in score scale and variance were observed between structured and text-based HDBSCAN.

1) *ROC curves*

To further characterize model behavior beyond scalar summary metrics, receiver operating characteristic (ROC) curves, precision–recall (PR) curves, and rank-based correlation analyses were computed for each unsupervised anomaly detection method. Fraud labels were used exclusively for post hoc evaluation.

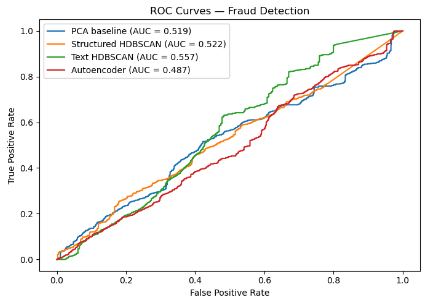

Figure 5: Receiver operating characteristic (ROC) curves for all unsupervised anomaly detection methods evaluated

Across models, the ROC curves exhibit similar overall shapes and remain close to the diagonal baseline, with area under the curve (AUC) values ranging from approximately 0.49 to 0.56. The text-based HDBSCAN method achieves the highest observed AUC, followed by structured HDBSCAN and PCA reconstruction, while the autoencoder exhibits the lowest AUC among the evaluated approaches. Differences between curves are most visible in the mid-range of false positive rates, where separation between methods is modest.

2) Precision-Recall Curve

Precision–recall curves plot precision as a function of recall and are particularly informative in settings with extreme class imbalance, such as fraud detection.

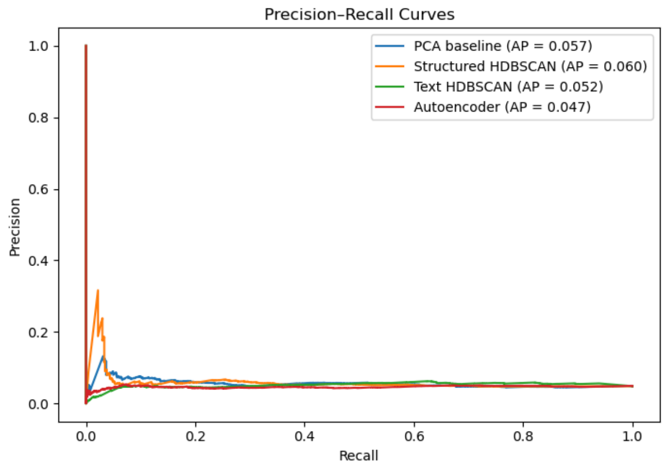

Figure 6. Precision–recall curves for PCA reconstruction, autoencoder reconstruction, structured HDBSCAN, and text-based HDBSCAN. All methods exhibit rapidly decreasing precision as recall increases, with average precision values ranging from approximately 0.047 to 0.060.

Across methods, precision declines rapidly as recall increases, with all curves converging toward low precision values over most of the recall range. Average precision (AP) values are low for all models, ranging from approximately 0.047 to 0.060, as reported in (e.g. Table 1). Structured HDBSCAN achieves the highest observed AP, followed by PCA reconstruction and text-based HDBSCAN, while the autoencoder exhibits the lowest AP among the evaluated approaches.

At very low recall levels, short spikes in precision are observed for some methods, reflecting the presence of a small number of highly ranked fraudulent observations. However, as recall increases, precision stabilizes at low levels across all models. These curves summarize threshold-dependent performance under class imbalance and complement the ROC-based evaluation presented earlier.

3) Spearman Rank Correlation

The correlation matrix reveals substantial variation in agreement across model pairs. PCA reconstruction and autoencoder reconstruction exhibit a strong positive correlation of 0.79 (eg Figure 7), indicating a high degree of similarity in anomaly ranking between these two reconstruction-based methods. In contrast, correlations between HDBSCAN-based methods and reconstruction-based methods are low, with values close to zero. The correlation between structured and text-based HDBSCAN is modest, indicating limited alignment between anomaly rankings derived from structured features and those derived from textual embeddings.

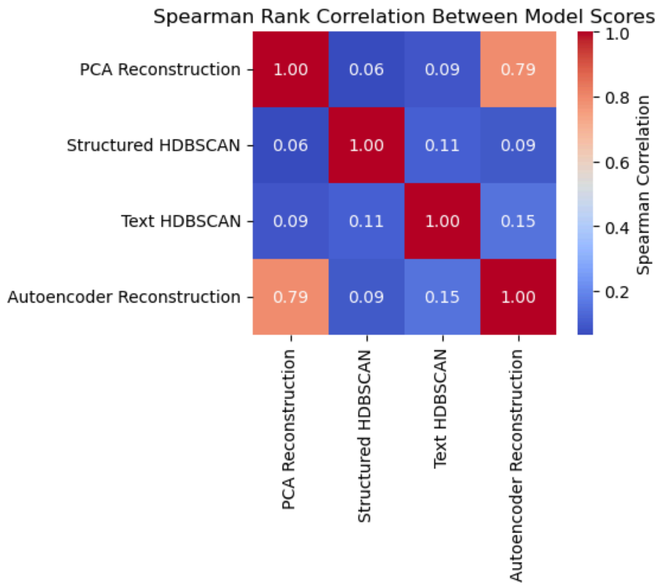

***Figure** 7 Spearman rank correlation matrix comparing anomaly score rankings across unsupervised anomaly detection methods.*

Higher values indicate greater agreement in relative anomaly ordering between model pairs.

V. Discussion

The results of this study highlight the challenges inherent in unsupervised fraud detection under large class imbalance. All evaluated models' performance metrics were modest, showing the difficulty of distinguishing fraudulent from non-fraudulent observations without access to labeled training data. Rather than identifying a single dominant approach, the results reveal meaningful differences in model behavior that align with their underlying assumptions.

Reconstruction-based methods exhibited distinct but related behaviors. PCA reconstruction error produced a relatively compact anomaly score distribution with a pronounced right tail for fraudulent observations, indicating sensitivity to linear deviations from dominant variance structure. The autoencoder generated a broader score distribution with substantially higher variance and maximum values, reflecting its ability to capture non-linear relationships but also increasing overlap between fraudulent and non-fraudulent observations. Despite this increased modeling flexibility, the autoencoder did not outperform PCA in ROC-AUC or precision-based metrics. The strong Spearman correlation between PCA and autoencoder scores suggests that, in this dataset, both methods rank observations similarly, even though the autoencoder assigns more extreme anomaly scores.

Density-based methods showed different behavior. Structured HDBSCAN exhibited conservative anomaly scoring. Most observations received close to zero scores, with only a small subset identified as outliers. This behavior aligns with its relatively higher precision (e.g. Table 1) . In contrast, text-based HDBSCAN achieved the highest ROC-AUC among evaluated approaches, indicating improved global ranking performance. However, this did not translate into higher precision at the top-ranked anomalies, suggesting that density-based isolation in reduced text embedding space does not necessarily correspond to fraud enrichment at strict cutoffs.

Precision–recall curves further emphasize the limitations of unsupervised methods in highly imbalanced settings. All models exhibit rapid declines in precision as recall increases. This means that high recall is achieved primarily at the cost of a large number of false positives. Short-lived spikes in precision at very low recall levels indicate that each method can identify a small number of highly suspicious observations, supporting the use of anomaly scores as a prioritization mechanism rather than a standalone decision rule.

Spearman rank correlation analysis provides additional insight into model complementarity. The low correlations observed between reconstruction-based methods and HDBSCAN-based methods indicate that these approaches capture different notions of anomalous behavior. Similarly, limited agreement between structured and text-based HDBSCAN suggests that linguistic and structural feature spaces highlight distinct types of irregularities. These findings support the decision to evaluate models in parallel rather than rely on a single anomaly detection strategy.

VI. Conclusion

This project evaluated multiple unsupervised anomaly detection techniques for fraud identification in a highly imbalanced transactional dataset, focusing on PCA reconstruction error, autoencoder reconstruction error, and density-based anomaly detection using HDBSCAN applied to both structured and text-derived features. All methods were trained without access to fraud labels and assessed using post hoc evaluation to reflect realistic deployment constraints.

The results demonstrate that no single unsupervised approach consistently outperforms others across all evaluation metrics. Reconstruction-based methods exhibited similar anomaly ranking behavior, as evidenced by strong rank correlation, while density-based methods identified anomalies according to fundamentally different criteria. Text-based HDBSCAN achieved the highest overall ranking performance as measured by ROC-AUC, whereas PCA reconstruction and structured HDBSCAN produced higher precision among top-ranked anomalies. Precision–recall analysis further highlighted the challenges of fraud detection under extreme class imbalance, with all methods exhibiting rapidly declining precision as recall increased.

Overall, these findings support the use of unsupervised anomaly detection as a prioritization and risk-ranking tool rather than a standalone classification solution. Future work may explore integrating temporal or relational features, incorporating limited supervision, or combining model outputs in a principled manner to improve fraud prioritization while maintaining adaptability to evolving patterns.

VIII. Individual Reflection

ALEX

Through this project, I developed a deeper understanding of unsupervised anomaly detection and its practical limitations in real-world fraud detection settings. Implementing and evaluating multiple methods in parallel highlighted how different modeling assumptions, such as reconstruction error versus density estimation, lead to fundamentally different anomaly ranking behaviors. I learned the importance of choosing evaluation metrics carefully under extreme class imbalance and gained experience interpreting ROC and precision–recall curves beyond headline scores. Additionally, conducting rank-based correlation analysis reinforced the value of comparing model agreement rather than relying solely on performance metrics. Overall, this project strengthened my ability to design defensible machine learning experiments, communicate results clearly, and critically assess model behavior in the absence of labeled training data.

Sri Krishna Koushik Komanduru

This project really changed how I think about machine learning. Getting ROC-AUC scores around 0.5 was frustrating at first, but I learned that sometimes honest results are more valuable than inflated ones. The biggest lesson was realizing that "anomaly" doesn't automatically mean "fraud", PCA and autoencoders found statistical outliers while HDBSCAN found rare observations, but neither strongly matched actual fraudulent postings. Working with such imbalanced data (only 4.8% fraud) taught me that you can't just optimize for accuracy; you need to think carefully about precision-recall trade offs and what false positives mean for real people. I also discovered that no single method works perfectly, different algorithms see different things, which is why real fraud detection systems combine multiple approaches with human review. Most importantly, I learned to report what I actually found, even when the results weren't impressive, because that's what honest research looks like.

VII. References

[1] E. S. L. Kaggle, “Fake Job Postings Dataset,” Kaggle, 2012. [Online]. Available: https://www.kaggle.com/shivamb/real-or-fake-fake-jobposting-prediction

[2] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” ACM Computing Surveys}, vol. 41, no. 3, pp. 1–58, 2009

[3] ‘ B. J. M. Ravi Kumar, M. B. Eba, and R. Mohammed, “Predicting fake job posts using machine learning models,” Int. J. Creative Res. Thoughts (IJCRT), vol. 12, no. 4, Apr. 2024, Art. no. IJCRT2404876.

[4] C. K. Ch’ng and X. Y. Wong, “Visualizing and clustering fake job postings: Data-driven insights for fraud detection,” BAREKENG: J. Math. & App., vol. 20, no. 1, pp. 865–880, Mar. 2026, doi: 10.30598/barekengvol20no1pp0865-0880.

[5] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” Philosophical Transactions of the Royal Society A}, vol. 374, no. 2065, 2016.

[6] J. Kim, H.-J. Kim, and H. Kim, “Fraud detection for job placement using hierarchical clusters-based deep neural networks,” Appl. Intell., vol. 49, no. 7, pp. 2552–2571, Jul. 2019, doi: 10.1007/s10489-019-01419-2., 1997.

IX. Appendix

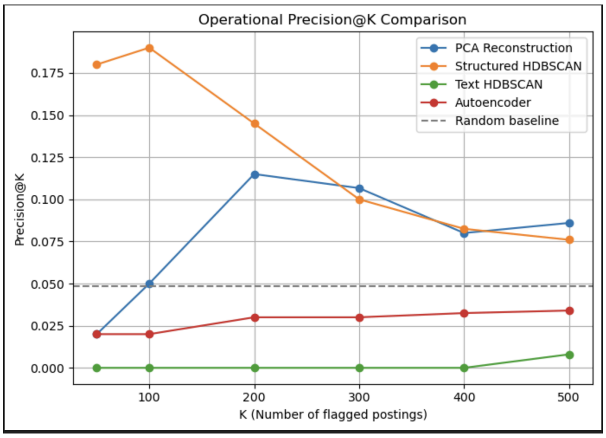

Fig. A.1. *Operational Precision@K comparison across unsupervised anomaly detection methods, including PCA reconstruction error, structured-feature HDBSCAN, text-based HDBSCAN, and an autoencoder baseline. The dashed line indicates the expected precision under random selection.*

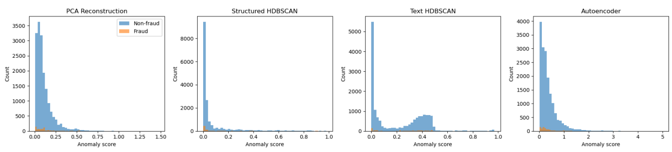

Fig. A.2. *Anomaly score distributions for fraudulent and non-fraudulent postings under PCA reconstruction error, structured-feature HDBSCAN, text-based HDBSCAN, and an autoencoder. Differences in distributional overlap highlight the varying discriminatory behavior of each method.*

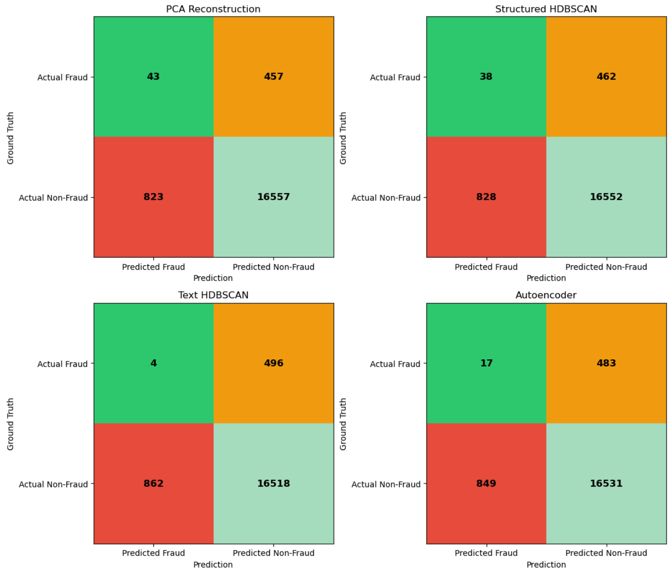

Fig. A.3. *Confusion matrices for unsupervised anomaly detection methods evaluated at a fixed decision threshold, including PCA reconstruction error, structured-feature HDBSCAN, text-based HDBSCAN, and an autoencoder. Predicted labels are derived by thresholding anomaly scores to flag a fixed number of postings as fraudulent.*