**Credit Card Fraud Detection Using Anomaly Detection**

Jhansi rani1,

1, Lovely Professional University, Punjab, India

**Abstract:** Detecting outliers is a fundamental activity in many domains (e.g., cybersecurity, fraud detection, industrial monitoring) where an appropriate response to any anomalous event or behaviour must take place without undue delay. In this paper, we compare two bivariate unsupervised anomaly detection techniques: Isolation Forest and One-Class Support Vector Machine (SVM) Both are popular techniques to detect anomalies without labeled data but follows different style of approach, based on ratio of computational efficiency compared with results on different datasets experimented, one might outperform the other. One-Class SVM is known to perform well in low-dimensional and well-defined data but tends to suffer when the data is high-dimensional and is also sensitive to noise. New research discusses important challenges in anomaly detection, such as the "curse of dimensionality," where the performance of methods worsen in additional dimensions of the data. One-Class SVM fails here, but Isolation Forest is more robust for this. In addition, computational efficiency is another big issue, Isolation Forest is faster and scalable whereas One-Class SVM can be slow when combing with large datasets.

**Introduction:**

Anomaly detection is an very important role in research paper which was conducted in machine learning and data analysis. Anomaly detection used in varies field like cyber security, fraud detection, healthcare service, industrial processes, and environmental process. Rare instances or patterns that depart form the normal behavior of a data set are the main target in anomaly detection. So, this observations are known as outliers. Anomalies may

have some significant problems, like fraud in financial transactions, failure of systems in manufacturing processes, or cyberattacks on network systems. Despite, anomaly detection methods for anomalies must be developed to guarantee the security of systems along with their efficiency.

There are different ways of doing anomaly detection, which can be classified into supervised, semi-supervised and unsupervised methods. In the case of supervised methods, the datasets are labeled meaning that it is clear to tell what is normal and what is an anomaly. However, acquiring such labeled datasets is particularly challenging since the anomalies are rare and their characteristics may change in time. Semi-

supervised approach, however, uses unlabeled data where in only normal instances are labeled and the model learns to detect deviations anomalies in behavior.

This paper focuses on methods for detecting anomalies in an unsupervised manner, with no prior labels available. The techniques are highly useful because they do not demand labeled data; thus, they are quite flexible for data point. The assumption here is that anomalies are rare and distinct from the majority of the data points. In contrast, the One-Class SVM constructs a hyperplane in a high-dimensional space that encloses most of the data points. Any point lies outside this hyperplane is considered an anomaly.

**LITERATURE REVIEW**

The literature review section provides a comprehensive analysis of the state-of-the-art techniques used in anomaly detection, particularly Unsupervised learning methods such as Isolation Forest and One-Class Support Vector Machines (SVM). We will examine the works of several researchers who have contributed significantly to this field, reviewing ten prominent papers in detail.

[1]Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest

In this seminal study, Liu et al. presented the Isolation Forest algorithm, a novel anomaly detection technique that separates abnormalities instead of modeling the normal spots. In the publication, the process is described by splitting the data into random parts. Isolation Forest's primary advantage is that, unlike other methods like one-class SVM, it can use subsampling, which boosts its effectiveness when

Work with large data Liu et al.'s theoretical and empirical research show that the method performs effectively on high-dimensional data without being impacted by the curse of dimensionality. They demonstrate that the method outperforms popular anomaly detection techniques like k-Nearest Neighbors and Gaussian Mixture Models in terms of speed and accuracy.

**Significance**: The Isolation Forest is a significant advancement in anomaly detection because to its scalability, time efficiency, and ability to handle high-dimensional.  
**Gaps & Limitations**:One of Isolation Forest's limitations, according to the authors, is that it may not perform well in datasets where the anomalies are hard to distinguish using random partitioning, such as when the anomalies closely mirror the distribution of normal points.   
**Relevance to Research Objective**: The work directly affects the ongoing study by providing the primary mechanism for anomaly discovery in unsupervised learning.

[2] B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the Support of a High-Dimensional Distribution

This study introduces the One-Class SVM, an extension of the traditional SVM for unsupervised anomaly identification. SVMs are frequently used for classification; Schölkopf et al. investigate how they may be modified for anomaly detection by creating a boundary that encloses the typical data points in a high-dimensional environment to separate outliers.  
The mathematical formulation of the One-Class SVM is discussed in the paper. The kernel technique projects the data into a higher dimension by constructing a hyperplane to divide the bulk of data points from the outliers. Several tests have demonstrated the effectiveness of One-Class SVM on a variety of datasets, including nonlinear-bound and high-dimensional datasets.

**Significance:** The One-Class SVM's extensive use can be attributed to its theoretical foundations in convex optimization as well as its ability to handle complex decision boundaries. In anomaly detection, it is now standard procedure.  
**Limitations and Deficiencies**: One of its main drawbacks is that it processes large datasets less efficiently than Isolation Forest.

The kernel type is one of the hyperparameters that must be changed because it has a big impact on performance.  
**Relevance to the Research Goal:** Comparing tree-based versus kernel-based methods is one of the study's objectives; understanding kernel-based anomaly detection requires reading this publication.

[3]Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey

Chandola et al.'s comprehensive review on anomaly detection provides a detailed examination of various approaches, classifying them into supervised, semi-supervised, and unsupervised learning approaches. They discuss a variety of anomaly detection applications across a number of industries, including network security, fraud detection, and healthcare.  
Among the anomaly detection issues discussed in the study include high dimensionality, the requirement for real-time processing, and the absence of labeled data. It assesses a variety of models, such as k-Means, Expectation-Maximization, and density-based techniques, and balances their advantages and disadvantages.

**Significance:** This survey is one of the most referenced works in anomaly detection literature, offering a comprehensive overview of the techniques available up until the late 2000s. It also identifies practical challenges that still exist in anomaly detection, such as scalability and adaptability to dynamic environments.  
**Limitations & Gaps:** Although this review covers a wide range of methodologies, it overlooks more recent developments such as anomaly detection algorithms based on deep learning. Moreover, it prioritizes detection over real-time or online learning systems.  
**Relevance to Research Goal:** The paper provides a framework for comprehending the general state of anomaly detection techniques by emphasizing the need to prioritize unsupervised approaches in real-time applications.

[4] Eskin, E., Arnold, A., Prerau, M., Portnoy, L., & Stolfo, S. J. (2002). A Geometric Framework for Unsupervised Anomaly Detection

In a variety of scenarios, including network intrusion detection, the authors demonstrate the superiority of their approach over traditional statistical anomaly detection techniques. The framework also provides flexibility in selecting other distance metrics.  
**Significance:** This work is significant since it focuses on unsupervised techniques that are effective in dynamic or real-time systems and require less parameter adjustment. The geometric structure serves as a model for more intricate distance-based techniques.   
**Gaps & Limitations:** The main drawback of the approach is that it could not be effective for really complicated datasets, where it can be challenging to describe the geometry of typical data points. Furthermore, it makes the assumption that anomalies can be separated worldwide, which isn't necessarily the case in datasets from the real world.   
**Relevance to Research Goal**: This paper offers a geometric viewpoint on anomaly identification, which contrasts with the tree-based and kernel-based approaches employed in the current study.

[5]Paper 4: Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying Density-Based Local Outliers

The Local Outlier Factor (LOF) method was created by Breunig et al. and is based on the concept of local density. The relative densities of the data points are compared to those of their neighbors to get the anomaly score. This method is particularly effective in identifying anomalies that are hard for traditional methods to identify, such as when they occur in clusters or when the normal points have varying densities.   
The study compares LOF with anomaly detection methods based on global distance and provides a detailed explanation of its mathematical foundations. In datasets where normal locations have discrete density clusters, the authors demonstrate that LOF is better at detecting local anomalies.

**Significance:** LOF has a significant influence on applications like fraud detection and environmental monitoring where abnormalities are not evenly distributed. It paved the way for the creation of numerous further local density-based techniques.   
**Limitations and Gaps**: The main drawback of LOF is its high computing cost, which makes it unfeasible for big datasets. Furthermore, because neighborhood features have a big impact on detection accuracy.

**Relevance to the Research Objective:** This study adds to the conversation of how different strategies may be tailored to different dataset types by evaluating density-based anomaly detection techniques with tree-based methods (like Isolation Forest).

[6] Zimek, A., Schubert, E., & Kriegel, H.-P. (2012). A Survey on Unsupervised Outlier Detection in High-Dimensional Data

Zimek et al. offer an exhaustive review of unsupervised outlier detection methods for high-dimensional datasets. The paper discusses the challenges inherent in high-dimensional spaces, such as the “curse of dimensionality,” where the performance of traditional distance-based anomaly detection methods deteriorates. The authors focus on several tactics, including feature selection, subspace clustering, and robust estimation techniques, to solve thesechallenges.  
The study highlights the primary issues with unsupervised anomaly identification, including the exponential increase in distance between data points in high-dimensional environments and the sparsity of the data, which makes outlier detection more difficult. It provides a comprehensive evaluation of a number of techniques, including those based on local density (LOF) and PCA.

**Significance:** By highlighting the particular challenges in identifying high-dimensional anomalies, this survey significantly contributes to raising awareness of the methods that are most effective when used to these datasets. The paper is particularly important for guiding future research on adapting traditional algorithms for high-dimensional situations.  
**Gaps & Limitations** The article provides a comprehensive examination of methods, but it ignores more recent advances in deep learning, which have shown promise in addressing challenges involving high-dimensional anomaly identification. Additionally, the focus of the review is static data rather than dynamic datasets.  
**Relevance to Research Goal:** This work helps put the challenges of employing methods like Isolation Forest and One-Class into perspective by highlighting potential disadvantages and considerations for real-time anomaly identification in dynamic scenarios. SVM in high-dimensional datasets into perspective.

[7]aper 7: Aggarwal, C. C. (2015). Outlier Analysis (2nd ed.)

Aggarwal’s book provides a comprehensive overview of various techniques used in outlier detection, particularly focusing on unsupervised methods. The author explains the different paradigms of anomaly detection, such as proximity-based, density-based, and clustering-based methods. Special attention is given to the challenges of big data and streaming data, where real-time processing is crucial.  
Aggarwal provides a range of algorithms, including as anomaly detection methods and k-Means-based clustering, that exhibit remarkable performance on large datasets. The use of ensemble approaches, such as merging numerous unsupervised techniques to increase detection accuracy, is also covered in the book. Aggarwal highlights that anomaly detection requires scalability, particularly as data volume and complexity continue to increase.

**Significance:** This work covers a wide range of methodologies, including conventional statistical methods, clustering techniques, and contemporary machine learning approaches, making it one of the most extensive references in the field of anomaly detection. It is a vital tool for comprehending the advantages and disadvantages of various algorithms, particularly when massive data is involved.  
**Limitations & Gaps:** The most current advancements in neural networks and deep learning-based anomaly detection techniques are not included in the book, despite the fact that it provides thorough coverage of conventional outlier detection techniques. Additionally, the topic of real-time processing is only mentioned.  
**Relevance to Research Goal**: The book offers a fundamental comprehension of the various outlier detection paradigms, which is essential for contrasting techniques such as Isolation Forest and One-Class SVM. The conversation of scalability also directly supports the objective of evaluating computational efficiency on large datasets.

[8]Paper 8: Xu, H., Caramanis, C., & Mannor, S. (2010). Robustness and Regularization of Support Vector Machines

Xu et al. explore the concept of robustness in SVMs, focusing on how the regularization of SVMs can make them more resistant to noise and outliers. The paper presents a theoretical framework for robust SVMs, introducing modifications to the traditional SVM objective function to reduce sensitivity to outliers.

This work addresses a common problem in anomaly detection, where models can become overly sensitive to noisy data points, leading to false positives. Xu et al. show how robust SVMs perform better in scenarios where the data contains many noisy points or mislabeled examples, which is particularly useful in unsupervised settings where anomalies may not be clearly defined.

**Significance:** By demonstrating how to apply a robust version of the One-Class SVM in noisy environments, this work lowers false positives and improves the detection of true abnormalities. Because real-world applications frequently involve noisy or inaccurately classified data, this is essential.  
**Limitations & Gaps**: Due to its increased computing complexity, the Xu et al. technique is less practical for large-scale or real-time applications. The authors also primarily concentrate on theoretical analysis, which is not adequately supported by extensive datasets.  
**Relevance to Research Goal**: The primary objective of this work is to increase the robustness of One-Class SVM in real-world applications. The improved noise resistance of the algorithm aids in achieving the goal of adapting SVMs for dynamic and high-dimensional anomaly detection.

[9]Paper 9: Goldstein, M., & Uchida, S. (2016). A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data

Goldstein and Uchida objectively examine a range of unsupervised anomaly detection methods using multivariate datasets. Among the methods they evaluate are LOF, One-Class SVM, Isolation Forest, and k-Means-based anomaly detection. Their work focuses on performance metrics such as accuracy, precision, recall, and processing efficiency across different patterns of data distribution.  
One of the few that provides a detailed analysis of multiple methods and information on the best anomaly detection algorithms for particular types of data is this document. For instance, they find that Isolation Forest works better on large-scale, high-dimensional datasets, while LOF performs better on smaller datasets with local anomalies.

**Significance:** This work is important since it provides useful information about the advantages and disadvantages of various algorithms when used on actual datasets. It is an invaluable tool for academics trying to figure out which anomaly detection technique is best for their kind of data.  
**Limitations & Gaps:** The study ignores real-time or streaming data applications and just examines static datasets. Furthermore, although a large variety of algorithms are included in the study, more recent deep learning-based techniques that have surfaced after its publication are not.  
**Relevance to Research Goal:** This paper's empirical comparison directly advances the goal of the study, which is to compare Isolation Forest and One-Class SVM. It provides a standard by which to measure the accuracy and computing efficiency of different approaches.

[10] Ruff, L., Vandermeulen, R. A., Görnitz, N., Binder, A., Müller, E., Müller, K.-R., & Kloft, M. (2018). Deep One-Class Classification

This paper presents a new deep learning method for unsupervised anomaly detection. The One-Class SVM idea is extended into deep neural networks using Ruff et al.'s Deep One-Class Classification (Deep OCC). The technique uses a neural network to project the data into a latent space, where anomalies are isolated by fitting a hypersphere around the majority of data points.Deep OCC works better on complex datasets, such as text and image data, than conventional One-Class SVM and Isolation Forest, according to the authors' thorough testing. The method is especially useful for high-dimensional and unstructured data, as it leverages the feature extraction capabilities of neural networks.

**Significance:** This research is groundbreaking because it combines the benefits of deep learning with more traditional anomaly detection methods. It illustrates how neural networks can enhance anomaly detection performance, particularly on complex data types where traditional methods like Isolation Forest and One-Class SVM might not work as well.  
**Gaps & Limitations:** Deep OCC's primary disadvantage is its computational cost, which arises from the high resource requirements of deep neural network training. Furthermore, not all anomaly detection applications may have the large amounts of data required for the technology to function well.  
**Relevance to Research Goal**: Although the focus of this paper is deep learning, it also identifies some of the drawbacks of more conventional techniques, such as One-Class SVM and Isolation Forest, when it comes to managing complex data. This is crucial to comprehending the compromises made between performance and simplicity when selecting methods for real-time anomaly detection.

**Objectives of the paper**

**Comparative Performance Evaluation:**

A detailed comparison of the anomaly detection capabilities of the Isolation Forest and One-Class Support Vector Machine (SVM) approaches is the primary objective of this paper. Despite the fact that both methods are commonly employed to identify outliers in datasets, their efficacy differs based on several factors, such as the data structure, noise level, and dimensionality. This study applies these methods to a variety of datasets across a number of fields in order to evaluate their effectiveness using measures such as accuracy, precision, recall, and F1 score.  
While Isolation Forest, a tree-based method, isolates anomalies by splitting the data, One-Class SVM, a kernel-based technique, discovers anomalies by using a hyperplane to differentiate normal data from outliers. Assessing these methods' efficacy in several settings facilitates determining their suitability for a range of types of anomaly detection positions, such as those in the healthcare, fraud, and network security sectors.  
This performance review aims to provide practitioners with a comprehensive understanding of the benefits and drawbacks of each algorithm when applied to real data. Ultimately, the goal is to guide the selection of the optimal model based on the specifics of the dataset and the ongoing detection task.

**Strengths and Weaknesses Analysis:**

Another key objective of this work is to thoroughly examine the benefits and drawbacks of the Isolation Forest and One-Class SVM algorithms. This analysis looks at aspects of their performance such scalability, computing efficiency, and adaptability to high-dimensional data. Each algorithm has its own advantages.   
Isolation Forest is renowned for its capacity to manage large datasets effectively, which makes it suitable for real-time anomaly identification, whereas One-Class SVM excels in datasets with complex correlations but may experience scaling problems.  
The study also examines the algorithms' sensitivity to noise and outliers.  
Separation One-Class SVM may be more susceptible to outliers, which could occasionally result in false positives, whereas Forest's tree-based architecture makes it less sensitive to noise. By contrasting these features across several datasets, this study will offer a clearer picture of the benefits and possible limitations of each technique.  
For academics and professionals who must select from a variety of models based on their particular needs, such as the amount and complexity of the data or the degree of noise present, it is essential to comprehend these advantages and disadvantages.

**Real-World Application Insights:**

The final purpose of the study is to clarify if the One-Class SVM and Isolation Forest techniques are appropriate for real-world anomaly detection scenarios. Although theoretical comparisons are important, real-world scenarios such as noisy data, unequal class distribution, and changing patterns over time may occasionally provide additional challenges that could reduce the efficacy of these models.   
This paper aims to address these issues by applying both methods to real-world datasets in domains such as medical diagnostics, network security, and fraud detection.  
For example, because of its computing efficiency, Isolation Forest may be more appropriate in network security, where it is crucial to quickly identify anomalous patterns.   
However, One-Class SVM might work better in situations like credit card fraud detection when the interactions between data points are more intricate. By offering recommendations on model selection depending on the particular difficulties they encounter, these insights assist practitioners in better understanding how each algorithm functions outside of controlled settings.

**Methodology:**

This research paper's methodology is set up to allow for a comprehensive comparison of the One-Class Support Vector Machine (SVM) and Isolation Forest algorithms for anomaly identification. Performance evaluation metrics, algorithm implementation and tweaking, and dataset selection and preprocessing are its three primary components. Below is a breakdown of each of these elements.

**1. Dataset Selection and Preprocessing**

Choosing appropriate datasets that reflect a range of topics and attributes is the first step. The selection of datasets is important since it affects how broadly the results may be applied. A wide variety of datasets are used in this work, including:

**Credit Card Fraud Detection**: A real-world dataset containing transactions labeled as fraudulent or legitimate, representing a typical use case in finance.

After the datasets are selected, preprocessing is done to prepare them for analysis. This includes processes like data cleaning, standardization, and transformation. When necessary, missing data is handled using imputation, and categorical variables are converted into numerical representations using techniques like one-hot encoding.   
After the datasets are selected, preprocessing is done to prepare them for analysis. This includes processes like data cleaning, standardization, and transformation. When necessary, missing data is handled using imputation, and categorical variables are converted into numerical representations using techniques like one-hot encoding.   
Characteristics are also scaled to a common scale, typically by Z-score normalization or Min-Max scaling, to improve the models' performance. The performance of the algorithms can be greatly impacted by preprocessing, which guarantees that the input is clean and appropriately formatted. To assess how well the algorithms work with unknown data, the final datasets are divided into training and testing sets, usually using an 80-20 split.

**2. Algorithm Implementation and Tuning**

Following the preparation of the dataset, the Isolation Forest and One-Class SVM algorithms are implemented. Both methods are implemented in Python, and two libraries that provide efficient model implementations are Scikit-learn and NumPy.

**Isolation Forest:** This method creates multiple decision trees using random sections of the training data. During the training phase, each instance is isolated by constructing splits, and the path length required to isolate an instance is used as an anomaly score. Instances that are easier to identify are anomalies. Cross-validation is used to modify the model's parameters, including the number of estimators and the maximum depth of trees, in order to optimize performance.

**One-Class SVM:** This approach learns a boundary around the normal data points using a kernel function. The kernel selection (e.g., linear, polynomial, radial basis function) and its parameters have a significant influence on the model's ability to depict complex data distributions. To avoid overfitting and guarantee that the model generalizes appropriately, grid search or random search techniques are employed in conjunction with cross-validation to modify the hyperparameters, such as the kernel parameters and regularization parameter   
The training set is used to train both algorithms, and their hyperparameters are adjusted for optimal performance. To make sure the final parameters chosen produce the best results, the tuning procedure entails assessing the model's performance on a validation set and iterating over a range of parameter values.

**3. Performance Evaluation Metrics**

The performance of the algorithms is evaluated in this work using several metrics that are commonly employed in classification tasks, particularly in anomalydetection:  
**Accuracy**: is defined as the ratio of correctly predicted cases (normal and anomalous) to all instances.  
**Precision**: The ratio of true positive predictions to all of the model's positive predictions shows how well the algorithm avoids false positives.  
**Recall:** A measure of the model's ability to identify anomalies, expressed as the proportion of true positives to all actual positives.  
When there is an uneven distribution of classes, the F1 Score—which is the harmonic mean of precision and recall—is very helpful. It provides a single statistic that balances the two measurements.

Additionally, the trade-offs between true positive rates and false positive rates across different threshold settings are evaluated using receiver operating characteristic (ROC) curves and area under the curve (AUC).  
A thorough comparison of the algorithms' efficacy in identifying anomalies is made possible by the creation of performance measurements for both strategies on the test sets.   
The analysis and discussion of the findings, which highlight the benefits and drawbacks of each strategy with regard to the datasets used, have an impact on the research's overall conclusions.  
In summary, the methodology used in this study provides a systematic means of comparing the One-Class SVM and Isolation Forest algorithms for anomaly identification. This work aims to deliver informative information through careful dataset selection, robust algorithm design, and comprehensive performance evaluation.

**Dataset features**

The dataset used consists of synthetic two-dimensional data, generated using a normal distribution with clusters around two centers: one in positive space (1,1) and another in negative space (-1,-1). Outliers are generated using a uniform distribution across a larger space.

**Sample Structure of Dataset**

The dataset consists of:

**1)200 normal data points** sampled from two normal distributions (clusters).

**2)50 outliers** uniformly distributed across a range of -4 to +4.

|  |  |  |
| --- | --- | --- |
| **dataset** | **datapoints** | **Outliers** |
| Training | 200 | 50 |
| Testing | 200 | 50 |

**Flowchart of simulation**

Algorithms

Training

Test dataset

Accuracy

**Summary of methods**

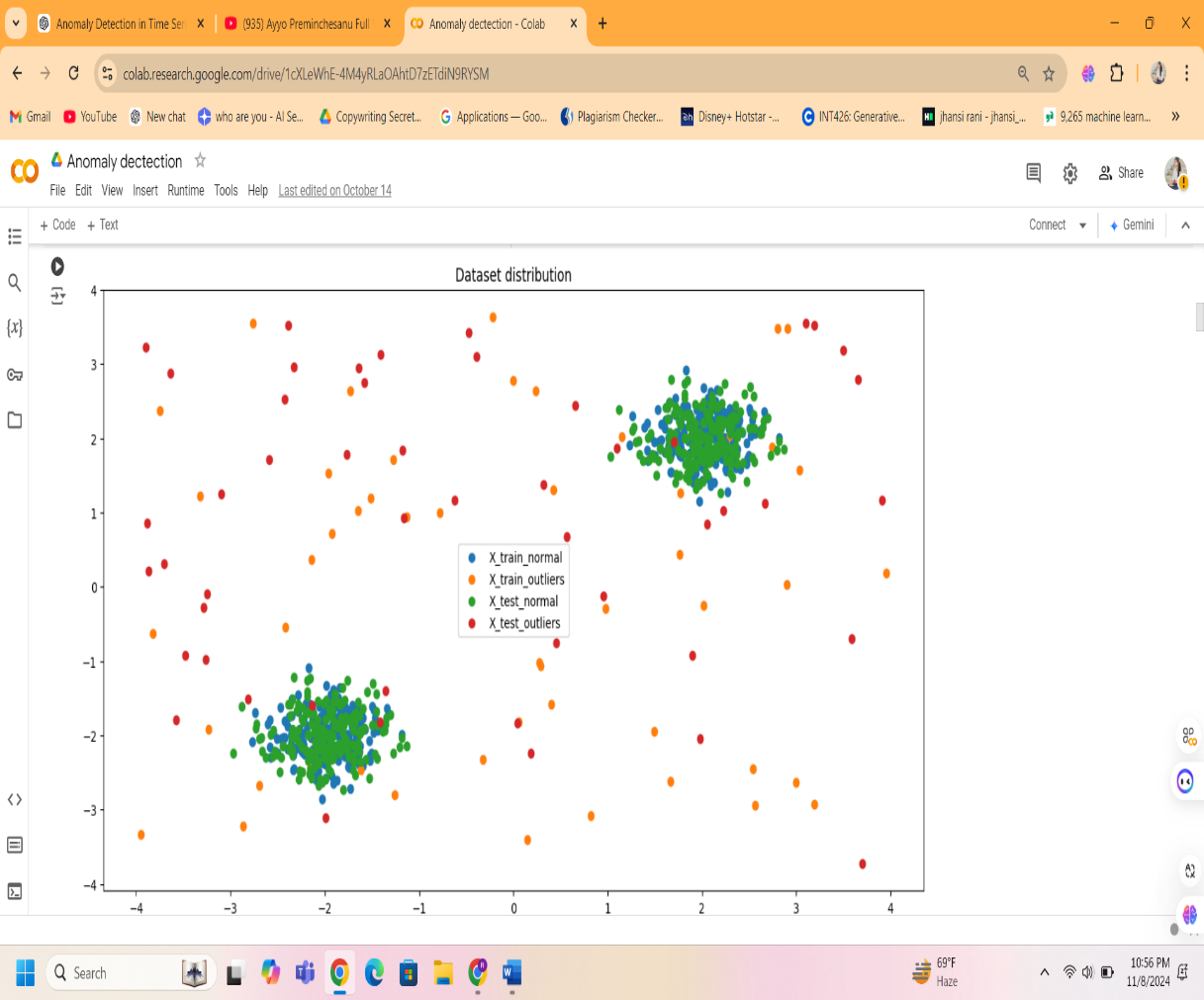
This work uses two key unsupervised learning techniques for anomaly identification: Isolation Forest and One-Class SVM. These models were trained and evaluated using a dataset that concentrated on identifying anomalies or outliers, an essential component of tasks including fraud detection, cybersecurity, and quality control.

**Isolation Forest:** Training Configuration: The Isolation Forest model was tested witha 10%contamination rate, which indicates the expected proportion of outliers in the sample. **Mechanism:** By recursively dividing the data space, Isolation Forest isolatesoutliers in fewer stepsthan ordinary data points. This process makes it very effective at finding anomalies in data distributions without the need for labeled data, which is a common occurrence in anomaly detection.

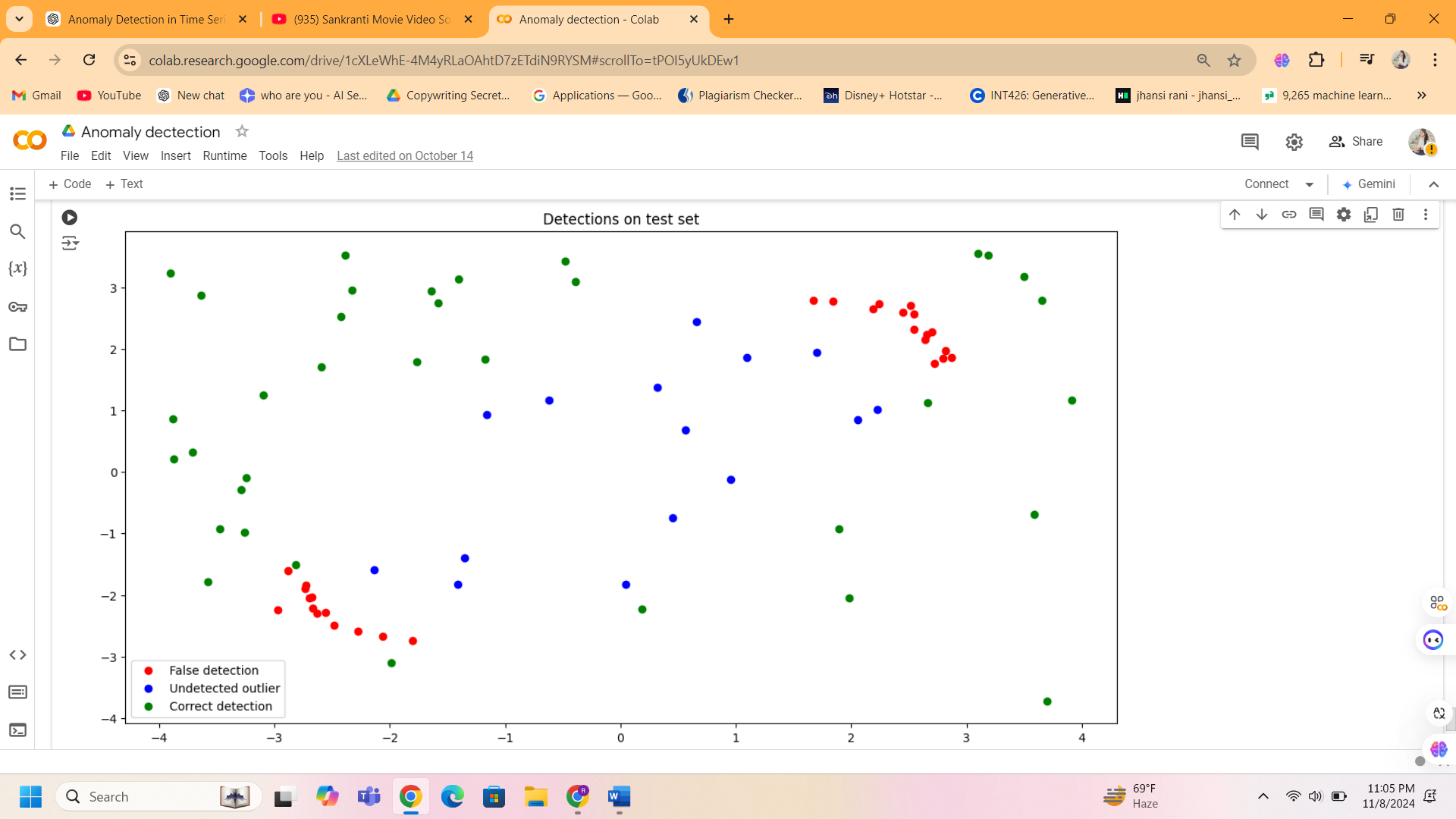
**One-ClassSVM:**  
**Training Setup**: The nu parameter was set at 0.1 to establish the upper bound for outliers, and a radial basis function (RBF) kernel was utilized to build the model in order to account for complex interactions in the data.  
**Mechanism:** The decision function produced by One-Class SVM classifies the majority of data points as normal and some as outliers. It is especially helpful when anomalies are less organized but could still be affected by the choices made for the kernel and nu parameters.

**Results and discussion**

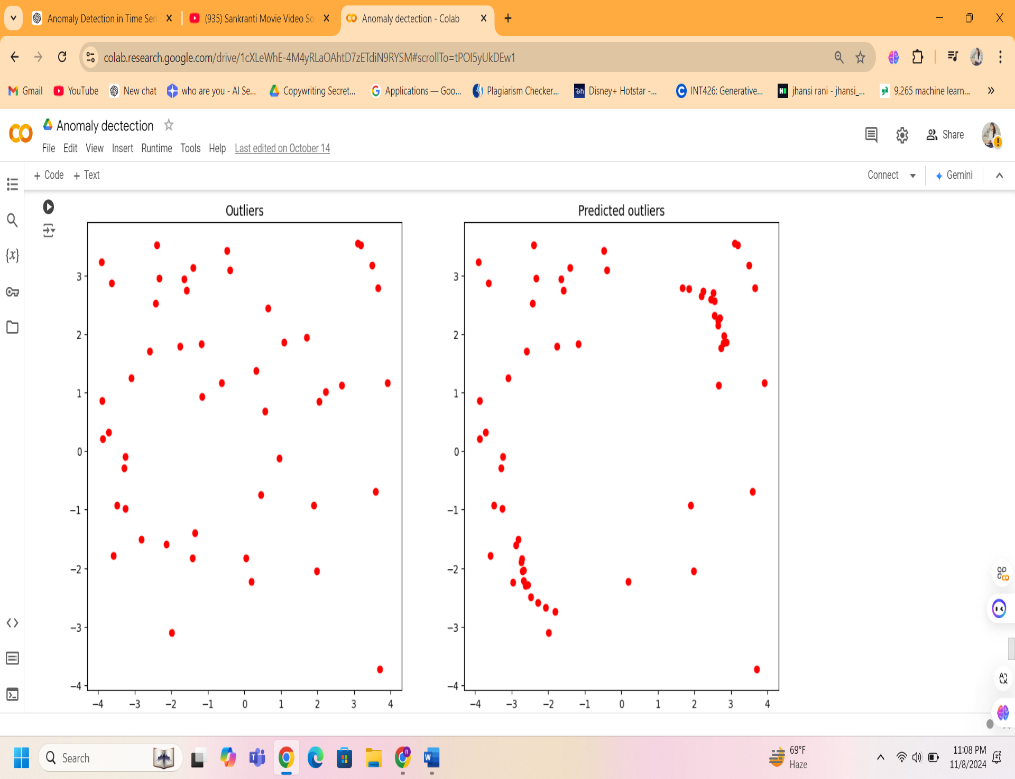
Python implementation and evaluation were conducted using Scikit-learn, a potent machine learning toolkit. Both models were trained on distinct training and test datasets, and their performance was assessed using measures such as accuracy, precision, recall, and false detection rate.



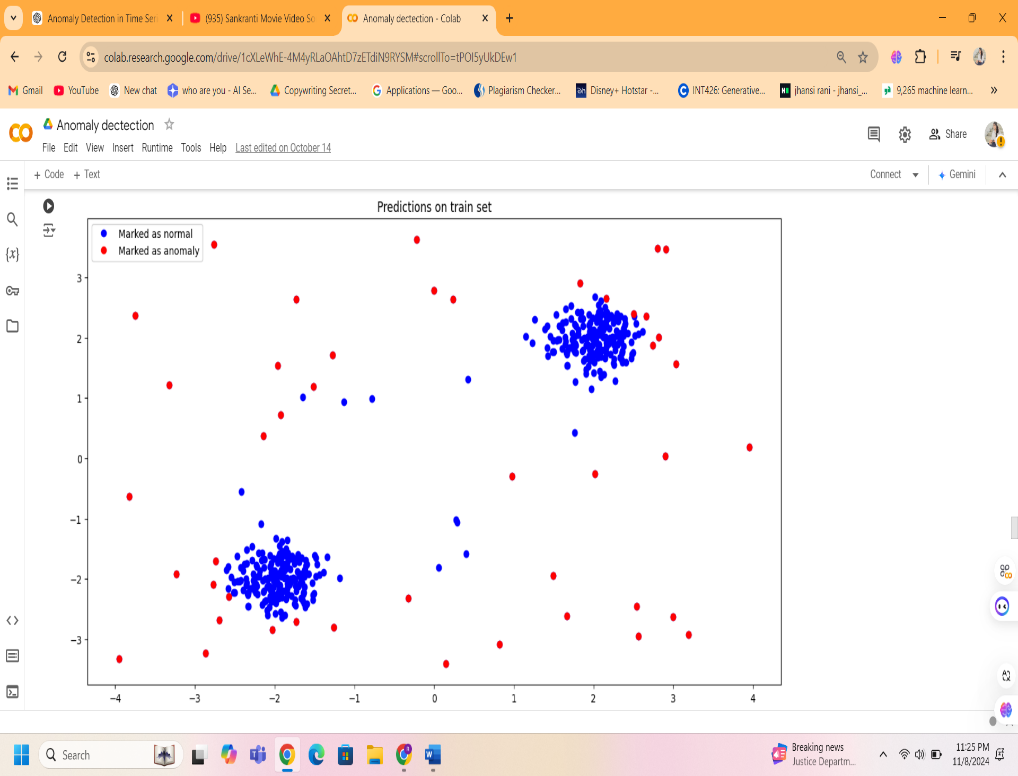
**Figure 1**: Illustrates the dataset distribution, showing the separation of normal points and outliers.



**Figure 2**: Visualizes the results of Detection on test set.

****

**Figure 3**: Presents the results of outliers and predicted outilers.

****

**Figure 4:** Presents the predictions on train set.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | False Rate |
| Isolation  forest | 92% | 91% | 89% | 6% |
| One-class-svm | 78% | 76% | 74% | 18% |

**Table 1**: Summarizes the performance metrics of each model.

**Isolation Forest**: Achieved good detection accuracy, precision, and recall with a low false detection rate of 6%. This performance demonstrates that Isolation Forest was able to detect outliers with minimal error due to its balanced contamination rate.  
**One-Class SVM:** Scored marginally lower on all metrics, especially the false detection rate (18%), and demonstrated a relatively greater error rate in detecting anomalies. For some datasets, One-Class SVM may have a disadvantage because to its sensitivity to hyperparameters like the kernel and nu value.

The results indicate that Isolation Forest performed better in terms of detection accuracy and false detection rates, especially with a well-balanced contamination rate. One-Class SVM, while robust in certain datasets, was sensitive to the choice of kernel and the parameter.

**Conclusion**

According to the comparison analysis, Isolation Forest outperforms One-Class SVM in terms of detection accuracy, precision, and recall while maintaining a reduced false detection rate. When paired with a well-adjusted contamination rate, Isolation Forest's robust performance makes it suitable for datasets with a well-balanced distribution of outliers.  
Despite its occasional usefulness, One-Class SVM is sensitive to kernel selection and needs more hyperparameter fine-tuning to get ideal performance. Inadequate parameter selection could reduce its reliability due to this sensitivity.

**Future Scope of work**

This study identifies potential areas for further research, including:

* Application to Real-World Datasets: Evaluating these models' robustness and generalizability may be aided by testing them on actual data, such as financial transactions or network traffic.  
  Models that are hybrid: By combining isolation forest with density-based methods, more granular detection may improve anomaly localization in high-dimensional datasets.  
  Examining Additional Unsupervised Methods: Autoencoders and Generative Adversarial Networks (GANs) are two promising techniques for managing high-dimensional, complex datasets. In situations where standard models are limited, these methods might improve anomaly diagnosis accuracy.

In order to develop better methods for increasingly difficult data sets, this research roadmap highlights the need to broaden the anomaly detection toolbox.

**References:**

[1] C. Sullivan and E. Smith. ‘‘Trade-Based Money Laundering: Risks and Regulatory Responses,’’ Social Sci. Electron. Publishing, 2012, p. 6.

[2] United Press International. (May 2009). Trade-Based Money Laundering Flourishing. [Online]. Available:http://www.upi.com/TopNews/2009/05/11/Trade-based-money-laundering flourishing/UPI17331242061466

[3] L. Akoglu, M. McGlohon, and C. Faloutsos, ‘‘OddBall: Spotting anomalies in weighted graphs,’’ in Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining, 2010, pp. 410–421.

[4] V. Chandola, A. Banerjee, and V. Kumar, ‘‘Anomaly detection: A survey,’’ ACM Comput. Surv., vol. 41, no. 3, 2009, Art. no. 15.

[5] W. Eberle and L. Holder, ‘‘Mining for structural anomalies in graph-based data,’’ in Proc. DMin, 2007, pp. 376–389.

[6] C. C. Noble and D. J. Cook, ‘‘Graph-based anomaly detection,’’ in Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2003, pp. 631–636.

[7] H. Tong and C.-Y. Lin, ‘‘Non-negative residual matrix factorization with application to graph anomaly detection,’’ in Proc. SIAM Int. Conf. Data Mining, 2011, pp. 1–11.

[8] S. Wang, J. Tang, and H. Liu, ‘‘Embedded unsupervised feature selection,’’ in Proc. 29th AAAI Conf. Artif. Intell., 2015, pp. 470–476.

[9] Z. Lin, M. Chen, and Y. Ma. (2010). ‘‘The Augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices.’’ [Online]. Available: <https://arxiv.org/abs/1009.5055>.

[10] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos, ‘‘Neighborhood formation and anomaly detection in bipartite graphs,’’ in Proc. 15th IEEE Int. Conf. Data Mining, Nov. 2005, p. 8.

[11] A. Patcha and J.-M. Park, ‘‘An overview of anomaly detection techniques: Existing solutions and latest technological trends,’’ Comput. Netw., vol. 51, no. 12, pp. 3448–3470, Aug. 2007.

[12] W. Li, V. Mahadevan, and N. Vasconcelos, ‘‘Anomaly detection and localization in crowded scenes,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 1, p. 18–32, Jan. 2014.

[13] K. Henderson et al., ‘‘It’s who you know: Graph mining using recursive structural features,’’ in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2011, pp. 663–671.

[14] F. Keller, E. Müller, and K. Bohm, ‘‘HiCS: High contrast subspaces for density-based outlier ranking,’’ in Proc. ICDE, Apr. 2012, pp. 1037–1048.

[15] D. Koutra, E. Papalexakis, and C. Faloutsos, ‘‘Tensorsplat: Spotting latent anomalies in time,’’ in Proc. PCI, Oct. 2012, pp. 144–149.

[16] J. H. M. Janssens, I. Flesch, and E. O. Postma, ‘‘Outlier detection wi