ABNORMAL PREDICTION
IN SURVEILLANCE CAMERA

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

SATTU SAI PRANEETH	**(2103A52034)**
VANAMAMALAI SAI BABU	**(2103A52038)**
SABBANI UDAY	**(2103A52183)**

Under the guidance of

Mr. D. Ramesh

Assistant Professor, Department of CSE.

SR UNIVERSITY

Department of Computer Science and Artificial Intelligence

CERTIFICATE

This is to certify that project entitled **“ABNORAMAL PREDICTION LEARNING ALGORITHMS” is** the BonafIde work carried out by **SATTU SAI PRANEETH, VANAMAMALAI SAI BABU, SABBANI UDAY**

As a course project of the partial fulfilment to award the degree **BACHELOROF TECHNOLOGY in ARTIFICIAL INTELLIGENCE AND MACHINELEARNING** during the academic year **2022-2023** under the guidance and supervision.

Mr. D.Ramesh
Asst**.** Professor,
SR university ,
Ananthasagar, Warangal.

Dr. M Sheshikala
Asst.Prof .&HOD(CSE) ,
SR University,
Ananthasagar, Warangal

ACKNOWLEDGEMENT

We express our thanks to Course co-coordinator **Mr.D.Ramesh, Asst. Prof** for guiding us from the beginning through the end of the course Project . We express our gratitude to Head of the department CS&AI, **Dr. M. Sheshikala** , **AssociateProfessor** for encouragement , support and insightful suggestions,. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School and Artificial Intelligence , **DrC.V.Guru Rao** , for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

ABSTRACT

Anomaly detection, also known as abnormal detection, is a common task in machine learning that involves identifying patterns in data that deviate significantly from the norm. Anomalies can be caused by a wide range of factors, including errors in data collection, equipment failures, or fraudulent activity.

In the field of machine learning, various approaches can be used for abnormal detection, including supervised and unsupervised learning techniques. In supervised learning, the model is trained on labeled data to identify the patterns that distinguish normal data from abnormal data. In unsupervised learning, the model is trained on unlabeled data to identify patterns that are statistically significant and different from the normal data.

The performance of an anomaly detection model is typically evaluated based on its ability to accurately classify new data points as normal or abnormal. This can be measured using metrices such as precision, recall, and f1-score.

Anomaly detection has a wide range of applications in various industries, including finance, healthcare, and cybersecurity. It can help identify potential fraud, detect equipment malfunctions, and identify patients at risk of developing certain diseases.

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

Anomaly detection is the process of identifying patterns in data that deviate significantly from the norm. In other words, it involves finding data points that are significantly different from the majority of the data points. Anomaly detection has a wide range of applications, including fraud detection, network intrusion detection, predictive maintenance, and medical diagnosis.

Anomaly detection can be divided into two categories: supervised and unsupervised. In supervised anomaly detection, a model is trained on labeled data, where each data point is labeled as normal or abnormal. The model then uses this labeled data to identify anomalies in new data. In unsupervised anomaly detection, the model is trained on unlabeled data, where there is no prior knowledge of normal or abnormal data points. The model uses statistical techniques to identify data points that are significantly different from the majority of the data.

Anomaly detection can be challenging because anomalies can take many different forms, and they may be rare or occur in unexpected ways. Therefore, it is important to use appropriate techniques for anomaly detection, depending on the type of data and the specific application.

Machine learning algorithms such as decision trees, clustering, and neural networks are commonly used for anomaly detection. These algorithms can learn patterns in the data and use them to identify anomalies. However, it is important to evaluate the performance of these algorithms on the specific dataset and application, as different algorithms may be more or less effective depending on the context.

Overall, anomaly detection is a critical task in many industries, and machine learning techniques play an important role in identifying anomalies and helping organizations take appropriate action to address them.

1.1-PROBLEM STATEMENT:

Problem Statement:

The goal of this project is to develop a machine learning model that can accurately predict the abnormal detection based on their behavioral characteristics. Artificial intelligence has revolutionized the way we detect abnormalities. AI-based techniques are used to detect anomalies in data, detect malicious activities, and identify potential threats. AI can be used to detect abnormalities in medical images, detect fraud in financial transactions, and identify suspicious activities in security systems.This presentation will discuss the various applications of AI in abnormal detection, the advantages and disadvantages of using AI for abnormal detection, and the challenges and opportunities associated with using AI for abnormal detection.

The main motive is to prove the prediction accuracy using the different classification models and compare which model performs better regarding the problem.

1.2-EXISTING SYSTEM:

There are various existing solutions for anomaly detection, which can be broadly categorized into statistical, and hybrid approaches.

Statistical approaches involve using statistical techniques to identify anomalies in data. These techniques include control charts, time series analysis, and hypothesis testing. Control charts involve plotting data points over time and identifying points that fall outside of control limits, indicating a potential anomaly.

Hybrid approaches combine statistical and machine learning techniques to improve the accuracy of anomaly detection. For example, a hybrid approach might use statistical techniques to identify potential anomalies and then use machine learning algorithms to classify them as normal or abnormal.

1.3-PROPOSED SYSTEM:

With the assist of dataset obtained we are try to develop a model to detect anomalies in the data and theyare using ISOLATION FORESTS ALGORITHM.
We examined the outcomes of accuracy and found that model performs is best**.**

1.4-OBJECTIVES:

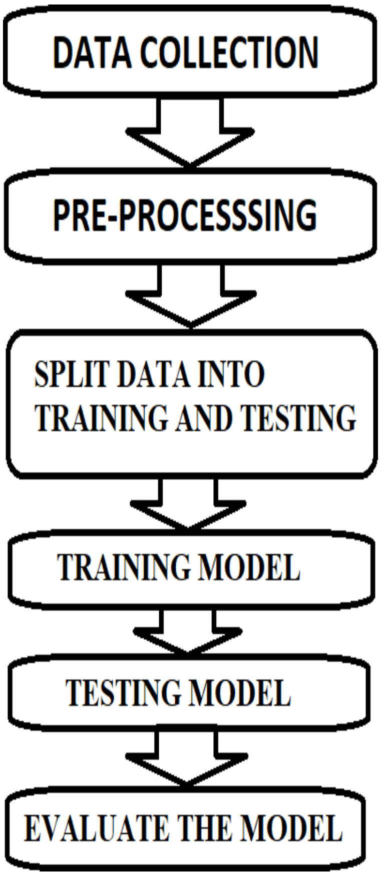
Overall, the objectives of anomaly detection can be summarized as follows:

- Identify anomaly.
- Reduce false statements.
- Improve accuracy.

- Automate the detection process.
- Provide actionable insights.

1.5-ARCHITECTURE:

This is a Supervised learning approach. As the data have categorical values we used classification based machine learning algorithms to predict the condition of crop, which might be normal, suspect or damaged.



We collected the data set firstly and after collecting the data pre-processing is done. Then the data set is get divided into 2 sets (i.e training and testing). Using classification-based machine learning models we trained the model after finding the accuracy on the training data set, we found the accuracy on the testing model. Based on those conclusions we evaluated the model. And this is the architecture followed by us.

CHAPTER-2

LITERATURE SURVEY

Prediction of abnormal behaviors for intelligent video surveillance systems.

Duque, D., Santos, H., & Cortez, P. (2007, March). Prediction of abnormal behaviors for intelligent video surveillance systems. In 2007 IEEE Symposium on Computational Intelligence and Data Mining (pp. 362-367). IEEE.

Early-anomaly prediction in surveillance cameras for security applications

Emad, M., Ishack, M., Ahmed, M., Osama, M., Salah, M., & Khoriba, G. (2021, May). Early-anomaly prediction in surveillance cameras for security applications. In 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 124-128). IEEE.

An efficient dimension reduction based fusion of CNN.

Sharma, R., & Sungheetha, A. (2021). An efficient dimension reduction based fusion of CNN and SVM model for detection of abnormal incident in video surveillance. Journal of Soft Computing Paradigm (JSCP), 3(02), 55-69.

Anomaly detection in traffic surveillance videos.

Nguyen, K. T., Dinh, D. T., Do, M. N., & Tran, M. T. (2020, June). Anomaly detection in traffic surveillance videos with GAN-based future frame prediction. In Proceedings of the 2020 International Conference on Multimedia Retrieval (pp. 457-463).

Real-Time Abnormal Object Detection

Ingle, P. Y., & Kim, Y. G. (2022). Real-Time Abnormal Object Detection for Video Surveillance in Smart Cities. Sensors, 22(10), 3862.

2.1-Analysis of the Survey:

In conclusion, there are several prediction models for abnormal detection that use a variety of approaches, including machine learning, Bayesian modeling , andANN. These models use a combination of environmental and demographic data to predict the potential occurrence of the disease with high accuracy

CHAPTER-3

DATA PRE-PROCESSING

- Image dataset
- It contains 500 images

Python Libraries are also imported in the above dataset and they be as of basic information mentioned:

Python Libraries like NumPy, Pandas, Seabornplots, Matplotlib, Scikit-learn are extensively used in the process of completing the ML Model.

Matplotlib: This library is responsible for plotting numerical data

Pandas: Pandas are an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. It eases data analysis, data manipulation, and cleaning of data

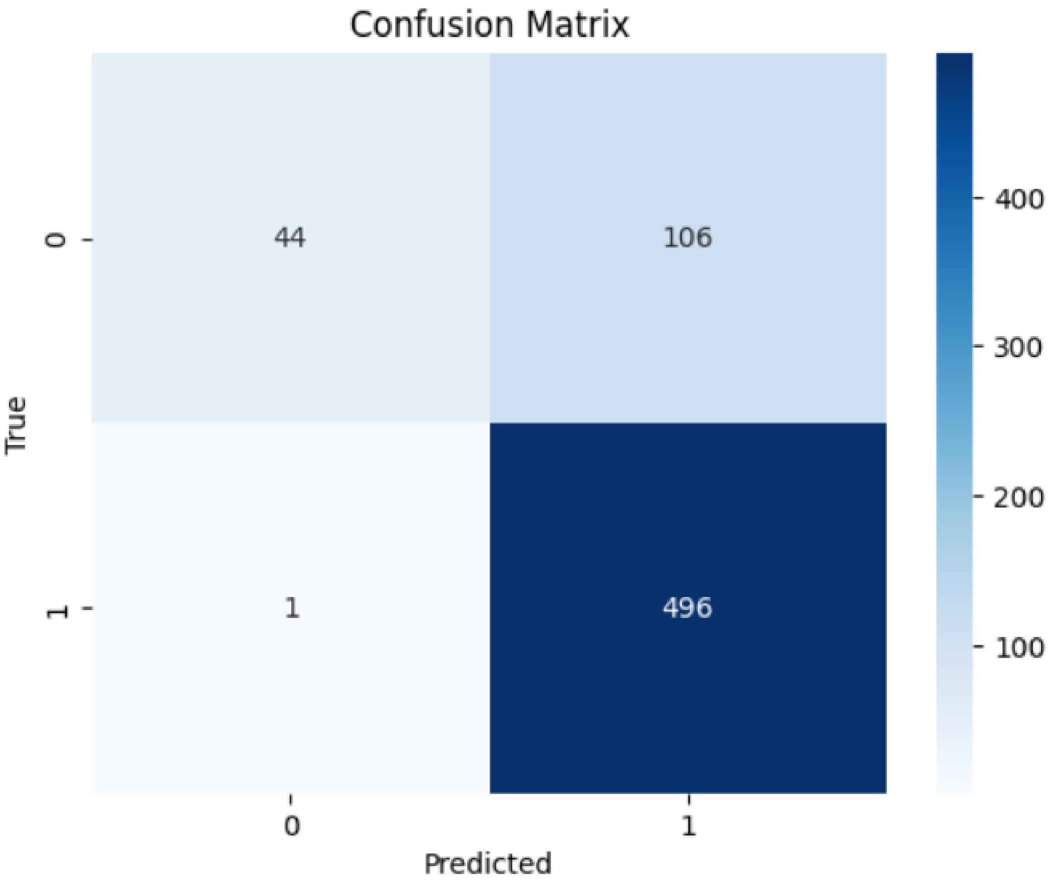
Numpy: The name “Numpy” stands for “Numerical Python”. It is the commonly used library. It is a popular machine learning library that supports large matrices and multi-dimensional data.

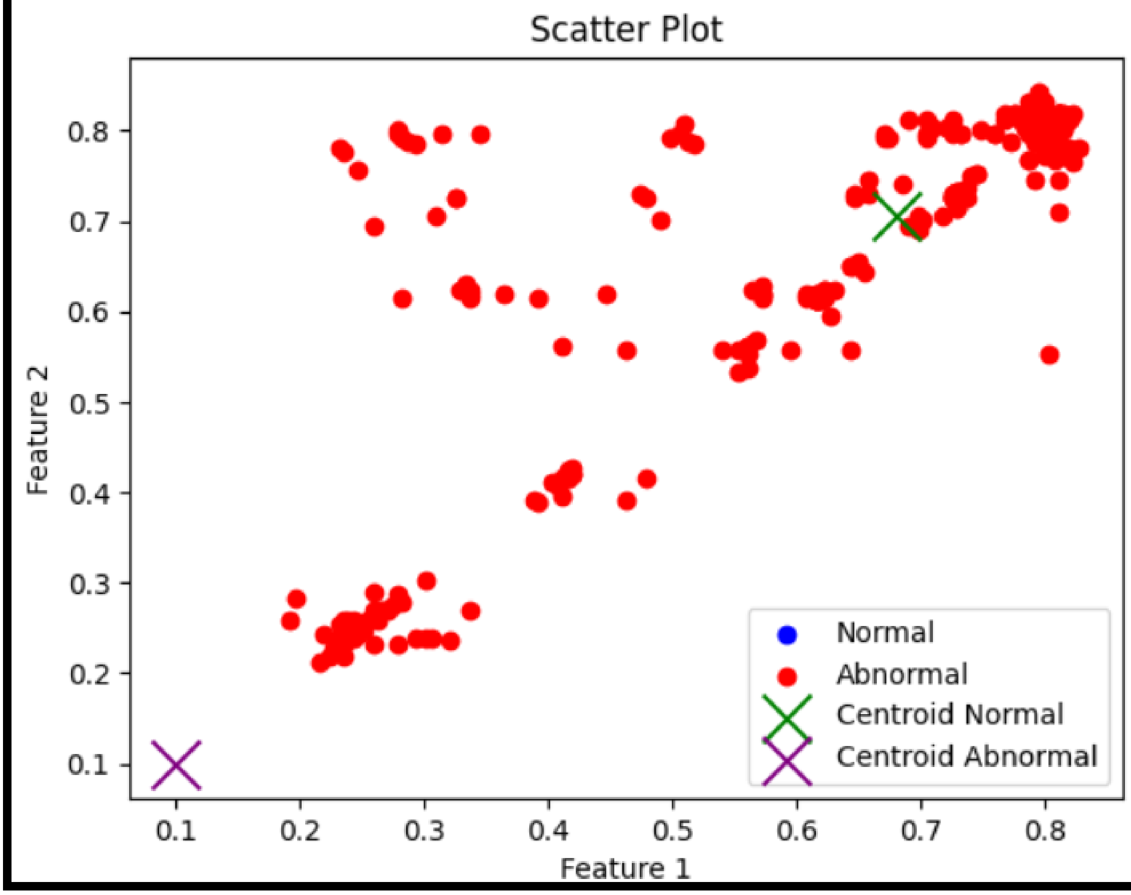
Scikit-learn: It is a famous Python library to work with complex data. Scikit-learn is an open-source library that supports machine learning. It supports variously supervised and unsupervised algorithms like linear regression, classification, clustering, etc.

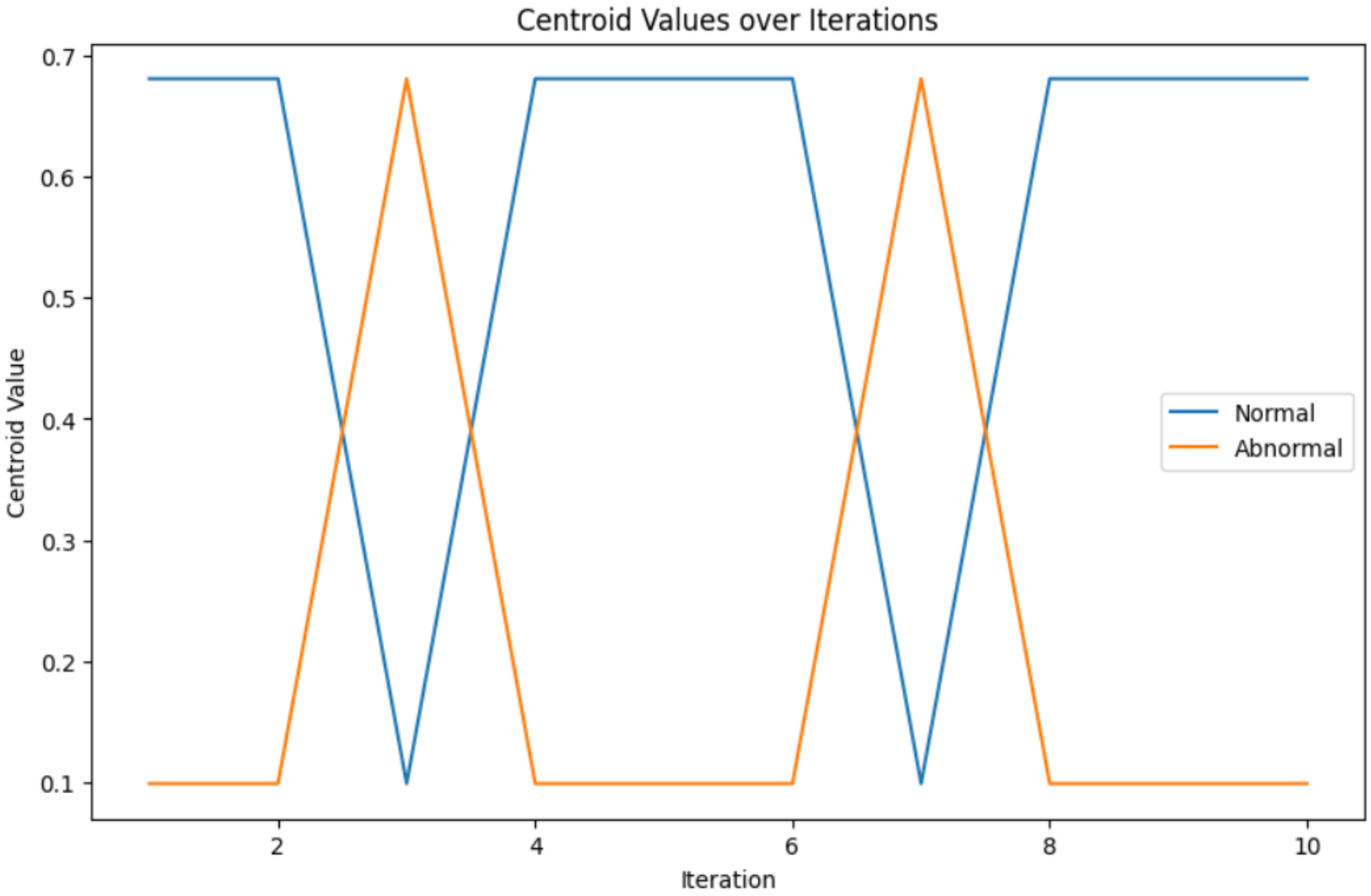
3.1 DATA DESCRIPTION

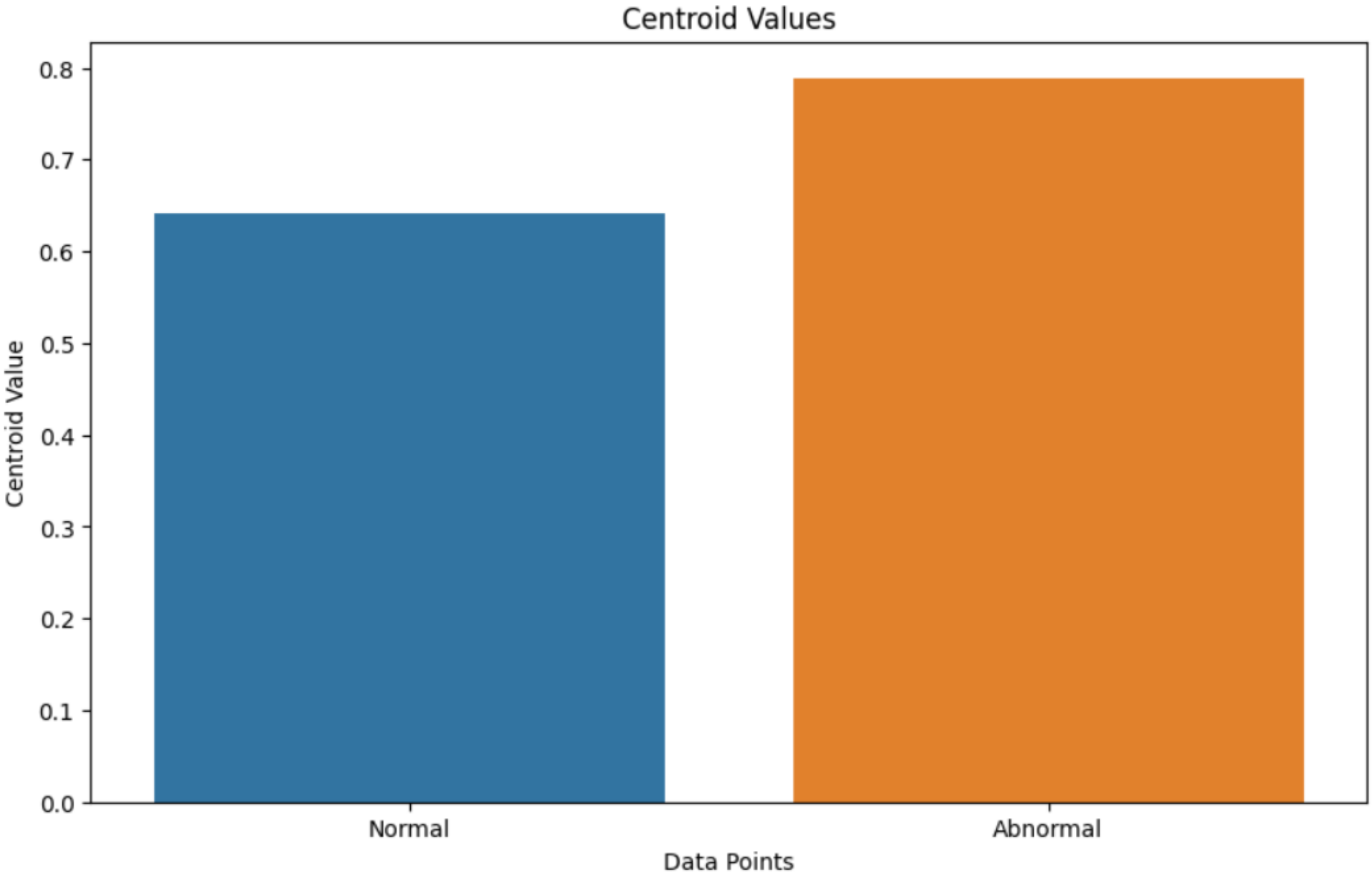
The goal of data preprocessing is to ensure that the data is accurate, complete, and consistent before any analysis is performed.

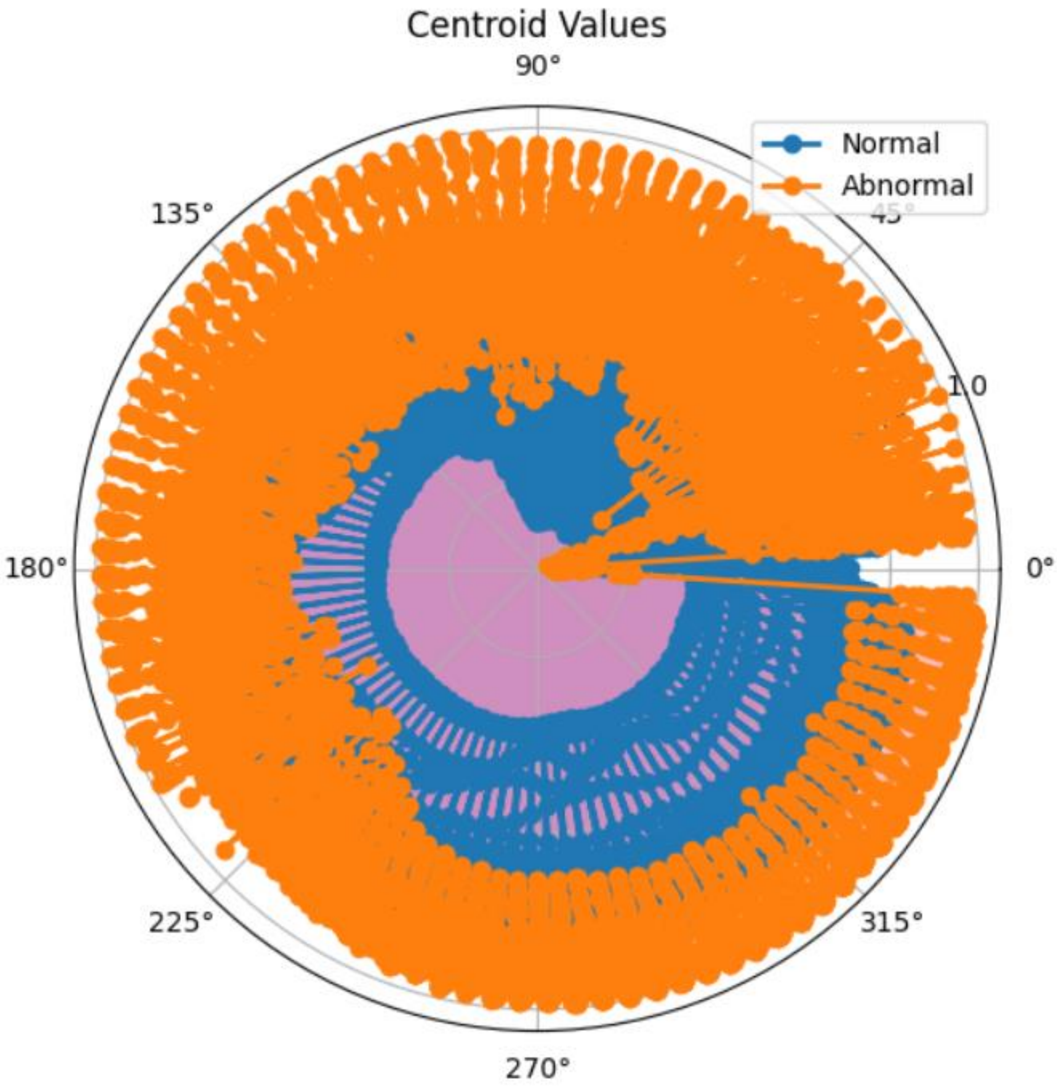
3.2 DATA VISUALIZATION






Centroid Values over Iterations
Centroid Value
0.7
0.6
0.5
0.4
0.3
0.2
0.1
2
4
6
8
10
Iteration
Normal
Abnormal


Centroid Values
Centroid Value
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Normal
Abnormal
Data Points


Centroid Values
90°
45°
135°
180°
0°
225°
270°
315°
1.0
Normal
Abnormal

CHAPTER 4

METHODOLOGY

4.1.1 PROCEDURE TO SLOVE THE GIVEN PROBLEM:

Clustering Techniques:

Rawat and many more found that Clustering techniques work by grouping the observed data into clusters, according to a given similarity or distance measure. There exist at least two approaches to clustering based anomaly detection. In the first approach, the anomaly detection model is trained using unlabeled data that consist of both normal as well as attack traffic. In the second approach, the model is trained using only normal data and a profile of normal activity is created. The idea behind the first approach is that anomalous or attack data forms a small percentage of the total data. If this assumption holds, anomalies and attacks can be detected based on cluster sizes large clusters correspond to normal data, and the rest of the data points, which are outliers, correspond to attacks.

K_MEANS:

K-means is a clustering algorithm that is commonly used in unsupervised learning for grouping similar data points together. While it is not a dedicated abnormal detection algorithm, it can be used as part of an anomaly detection pipeline to identify clusters of data points that are dissimilar from the rest of the data.

The basic idea of using k-means for abnormal detection is to first group the data points into clusters using the k-means algorithm, and then identify the clusters that contain a small number of data points or have a large variance. These clusters may contain abnormal data points that are dissimilar from the rest of the data and can be further investigated to identify abnormal events.

However, using k-means for abnormal detection has some limitations. For example, it assumes that the data can be grouped into well-defined clusters, which may not be the case for complex data distributions. Also, k-means is sensitive to the choice of the number of clusters (k), and choosing the wrong value of k can result in incorrect clustering results.

In summary, while k-means is not a dedicated abnormal detection algorithm, it can be used as part of an anomaly detection pipeline to identify clusters of data points that are dissimilar from the rest of the data. However, it has limitations and may not be suitable for all types of data distributions.

ISOLATION FOREST:

Isolation forest is machine learning algorithm that detects anomalies by partitioning data recursively using random splits. Anomalies have low isolation scores, useful for rare and unusual event detection in large datasets.

Isolation Forests(IF), similar to Random Forests, are build based on decision trees. And since there are no pre-defined labels here, it is an unsupervised model.

In an Isolation Forest, randomly sub-sampled data is processed in a tree structure based on randomly selected features. The samples that travel deeper into the tree are less likely to be anomalies as they required more cuts to isolate them. Similarly, the samples which end up in shorter branches indicate anomalies as it was easier for the tree to separate them from other observations.

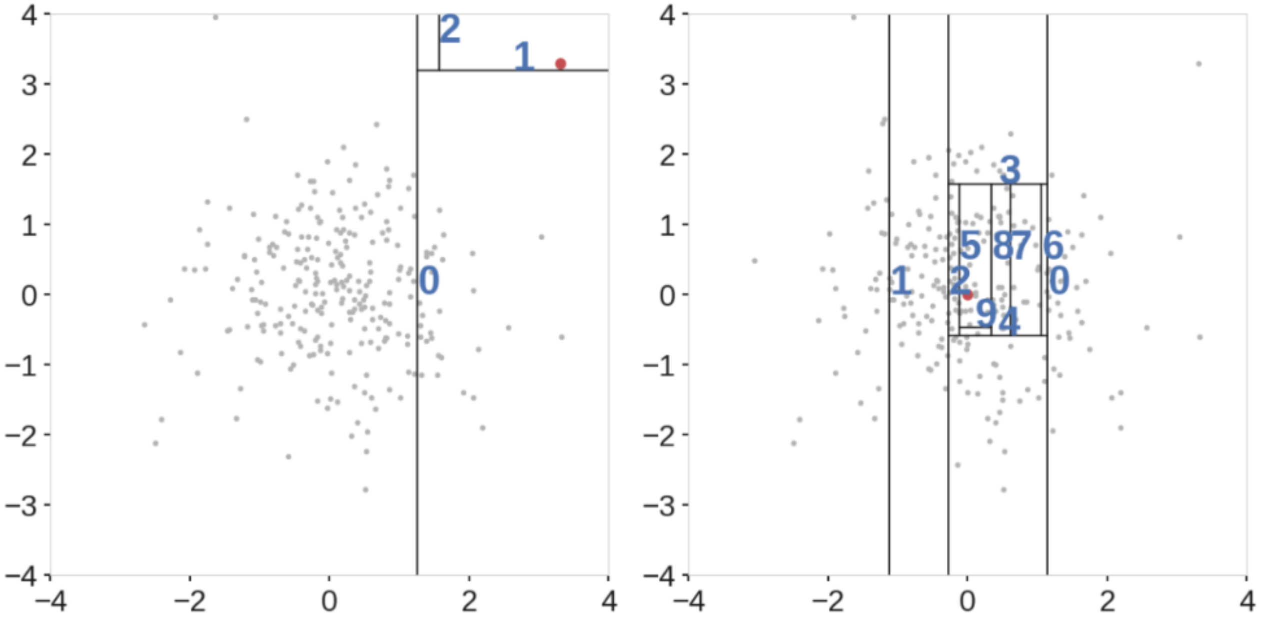
WORKING OF ISOLATION FOREST:

As mentioned earlier, Isolation Forests outlier detection are nothing but an ensemble of binary decision trees. And each tree in an Isolation Forest is called an Isolation Tree(iTree). The algorithm starts with the training of the data, by generating Isolation Trees.

Let us look at the complete algorithm step by step:

1. When given a dataset, a random sub-sample of the data is selected and assigned to a binary tree.
2. Branching of the tree starts by selecting a random feature (from the set of all N features) first. And then branching is done on a random threshold (any value in the range of minimum and maximum values of the selected feature).
3. If the value of a data point is less than the selected threshold, it goes to the left branch else to the right. And thus a node is split into left and right branches.
4. This process from step 2 is continued recursively till each data point is completely isolated or till max depth(if defined) is reached.
5. The above steps are repeated to construct random binary trees.

After an ensemble of iTrees(Isolation Forest) is created, model training is complete. During scoring, a data point is traversed through all the trees which were trained earlier. Now, an ‘anomaly score’ is assigned to each of the data points based on the depth of the tree required to arrive at that point. This score is an aggregation of the depth obtained from each of the iTrees. An anomaly score of -1 is assigned to anomalies and 1 to normal points based on the contamination(percentage of anomalies present in the data) parameter provided.



(a) Anomaly point

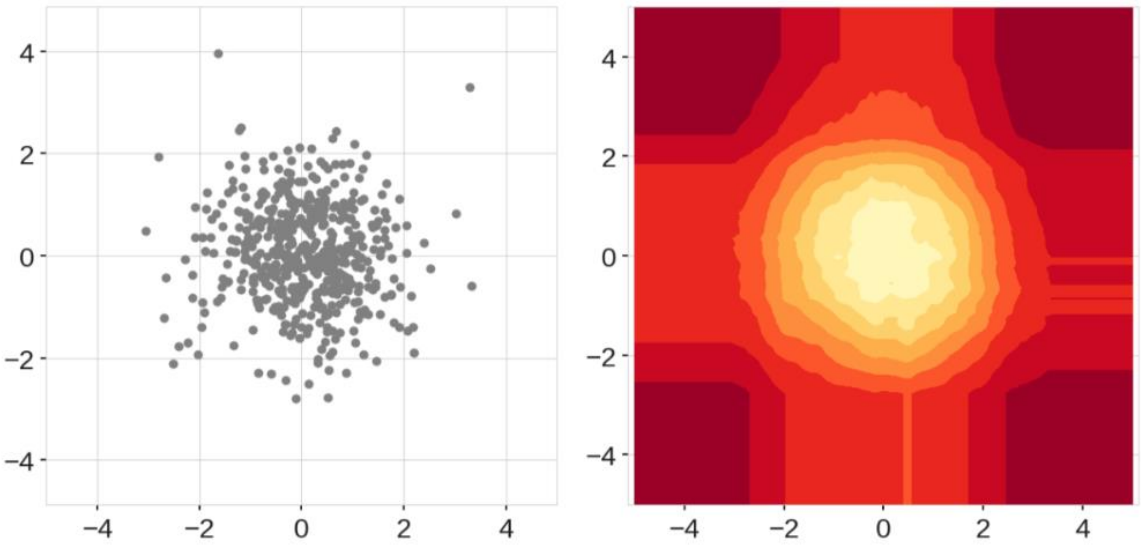
(b) Nominal point

Limitations of Isolation Forest:

Isolation Forests are computationally efficient and have been proven to be very effective in Anomaly detection. Despite its advantages, there are a few limitations as mentioned below.

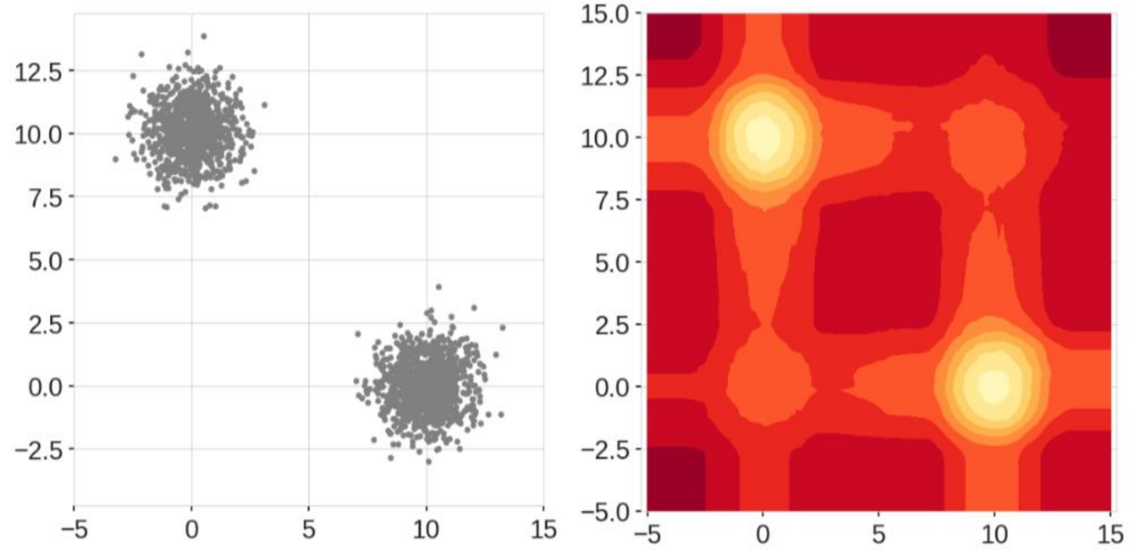
1. The final anomaly score depends on the contamination parameter, provided while training the model. This implies that we should have an idea of what percentage of the data is anomalous beforehand to get a better prediction.

2. Also, the model suffers from a bias due to the way the branching takes place.



(a) Normally Distributed Data

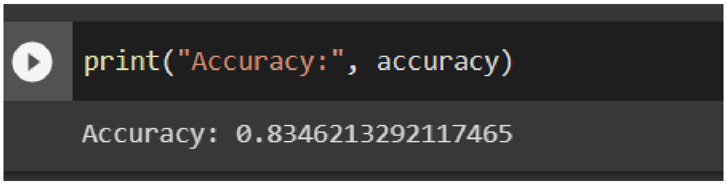
(b) Anomaly Score Map



(a) Two normally distributed clusters

(b) Anomaly Score Map

5.RESULTS:



```
print("Accuracy:", accuracy)

Accuracy: 0.8346213292117465
```

The result of abnormal detection using Isolation Forest can vary depending on the dataset and the parameters used for the algorithm. However, in general, Isolation Forest is known to perform well for detecting anomalies in high-dimensional datasets with a large number of features. It is also typically faster and more scalable than other outlier detection algorithms, such as clustering-based methods.

6.CONCLUSION:

In conclusion, abnormal detection is a crucial task in many applications, such as fraud detection, network intrusion detection, and anomaly detection in sensor data. Traditional statistical methods can be useful for detecting anomalies in low-dimensional datasets with simple distributions, but they often struggle with high-dimensional datasets or datasets with complex distributions.

Machine learning-based approaches, such as Isolation Forest, have shown promising results for abnormal detection in these more challenging scenarios. Isolation Forest is a fast and scalable algorithm that can efficiently detect anomalies in high-dimensional datasets with complex distributions. However, it is still important to carefully evaluate the performance of any abnormal detection algorithm on a specific dataset to ensure that it is suitable for the application at hand.

Overall, abnormal detection is a challenging and important task, and it requires careful consideration of both traditional statistical methods and machine learning-based approaches to achieve the best results.

7.FUTURE SCOPE

The future scope of anomaly detection is vast and includes several areas of research and development. Here are some potential future directions for anomaly detection:

1. Improving the accuracy of anomaly detection: There is always a need to improve the accuracy of anomaly detection algorithms. This can be achieved through better feature selection, improved algorithms, and the development of new techniques that can better handle complex data.

2. Real-time anomaly detection: Real-time anomaly detection is an area of growing importance, especially in applications such as cybersecurity, where timely detection of anomalies is critical. Future research will focus on developing algorithms that can detect anomalies in real-time, even for large datasets.

3. Combining multiple data sources: In many applications, data is collected from multiple sources, such as sensors, logs, and social media. Combining these different data sources can provide a more complete picture of the system being monitored, but it also presents new challenges for anomaly detection.

4. Explaining detected anomalies: One limitation of anomaly detection algorithms is that they often provide little or no explanation of why a particular data point was identified as an anomaly. Future research will focus on developing algorithms that can explain the detected anomalies, providing more transparency and enabling users to understand the underlying causes of the anomalies.

5. Applications in new areas: Anomaly detection is currently used in a wide range of applications, but there is potential for its use in new areas, such as healthcare, transportation, and energy.

Overall, the future scope of anomaly detection is vast and presents many exciting opportunities for research and development. As the use of data continues to grow, the need for accurate and efficient anomaly detection algorithms will only become more critical.

8.REFERENCES

[1] https://ieeexplore.ieee.org/document/9439459https

[2] https://ieeexplore.ieee.org/document/9033532

[3] https://www.researchgate.net/publication/351830421

[4] https://www.researchgate.net/publication/351830421

[5] https://www.sciencedirect.com