International Journal of
Electrical and Computer System Design
ISSN: 2582-8134 www.ijecsd.com

Anomaly Perception Data and Computation For Further Utilization

Angaleswary R*[1], Logesh T[2]

*[1]Student, Department of Computer Science and Applications, Periyar Maniammai Institute of Science and Technology, Thanjavur, India, angal24032001@gmail.com.

[2]Assistant Professor, Department of Computer Science and Applications, Periyar Maniammai Institute of Science and Technology, Thanjavur, India, logeshwr@gmail.com.

Abstract: The outlier identification method has a lot of applications and has recently attracted a lot of attention. Applications for the usage of outlier identification techniques have included clinical trials, voting irregularity analysis, data purification, network intrusion, severe weather prediction, geographic information systems, athlete performance analysis, and other data-mining activities. Outlier identification is an essential task in many safety-critical settings since an outlier flags aberrant operating conditions that may result in considerable performance loss, such as an aviation engine rotation failure or a pipeline flow issue. An outlier is a strange object in a picture, like a land mine. Early detection is essential because an anomaly may reveal a malicious breach into a system. In order to identify aberrant data and use it efficiently for production, this paper explores research on outliers in the automotive industry. Developers should choose an outlier identification method that is acceptable for their data collection in terms of the correct distribution model, the relevant attribute types, scalability, speed, and any incremental capabilities to enable the saving of more exemplars. With less computational complexity and experiments using data from various reporters as well as a synthesis of processed data from deep analysis and forecasting the acquired data for further acknowledgement, our project proposal can cost-effectively identify outliers in large-scale datasets from a variety of data views.

Keywords: Outlier Identification, Automobile Industry, Outlier Detection, Anomaly Detection, Decision Tree Algorithm.

1. Introduction

A substantial majority of data analysis tasks need the recording or sampling of numerous variables. Finding outlying findings is one of the initial steps in creating a cohesive analysis. Even though outliers are typically rejected as error or noise, they may contain important information. Detected outliers are candidates for aberrant data that could otherwise have a detrimental impact on model design, result in skewed parameter estimation, and yield false conclusions. Prior to approval or payment for application processing, such as loan application processing or social security benefit payments, an outlier detection system can identify any abnormalities in the application. Additionally, to ensure that the payout has not deteriorated into fraud, outlier detection can follow a benefit claimant's circumstances over time. Because irregular entries make it difficult to analyze and provide engine specs to the production team, finding outliers will help to speed up the process. The industry will benefit from its assistance in avoiding irrelevant data, improving the right production system, and leading to development across a variety of disciplines in addition to finding the most significant differences.

2. Literature Review

The study of this paper, using longitudinal sMRI slices, we use a GAN-based encoder-decoder framework, where the error between the reconstructed and actual neighbouring three slices stacks is used to identify ASD data as outliers. Reconstruction quality and, thus, the effectiveness of ASD detection were tested for three architectures: UNet, GAN, and SAGAN. The SAGAN with self-attention modules obtains the highest level of reconstruction and detection accuracy, while the UNet trained with the L2 goal performs poorly and creates hazy reconstructions [1]. The study of this paper, monitoring one or more of a DNN's hidden layers using the well-known outlier detection techniques Isolation Forest (IF) and Local Outlier Factor (LOF), and we compare the results to closely related works like the Softmax-based method, the Box abstraction-based method, and the BDD-based method for two well-known datasets like MNIST and GTSRB [2]. The technique suggested in this research solely utilises photos in order to minimise the dimension while maintaining as many of the images' unique qualities as feasible. For classifying objects, the dimension around their appearance was lowered, and the edge detection technique is used first to gather the object's appearance data. This study suggests a dimension-reduction and edge-detection methodology for detecting outliers in chest X-ray images [3]. The study about a strong theoretical foundation was prioritised, which allowed us to put today's two main lines of development—the more traditional kernel world and the more recent world of deep learning and representation learning for AD—into context. We specifically make linkages between traditional "shallow" and cutting-edge "deep" techniques and demonstrate how these connections may cross-fertilize or extend in both ways [4]. The study of the paper is both the plane wave and vibrating plate observations, the proposed model exhibits strong label prediction accuracies (all > 85%) towards the outlier site. With accuracy up to 85:6%/93:1% and 99:9%/99:9%, respectively, the proposed model can find outliers within pre-divided regions of simulated plane wave and vibrating plate observations [5]. In terms of performance, simplicity of implementation, and computational complexity, the newly offered ideas are determined to be effective. In addition, two suggested strategies are described in this study. These techniques are computationally simple and practical regardless of the high dimensions of the data since they transform the data to a one-dimensional distance space before looking for outliers [6]. A fundamental problem in data mining is outlier detection. The discovery of relevant information can result from the detection of outliers, and it has many real-world applications in fields like transaction records, stock market movement, sensor data, weather forecasting, etc. All types of data have outliers, and finding them in data mining is a difficult task. Density Based Algorithm is the most effective one. Compared to other algorithms, it takes a shorter period of time [7]. While some strategies are more general, many outlier detection methods have been developed that are specialised to particular application domains. Research on some application domains, such as research on crime and terrorism, is conducted under stringent confidentiality. The analyst can choose the most pertinent outliers by using a domain-specific threshold using scoring-based outlier detection approaches. Analysts cannot directly make this decision using techniques that assign binary labels to test instances, but they can indirectly do so by selecting different parameters for each technique [8]. The suggested method makes it possible to clean data at the university level quickly and accurately. Noise should be eliminated before outlier detection because noise is a random error or measurable variance. Finding patterns in data that deviate from expected behavior is the goal of outlier detection [9]. A novel method known as outlier detection, in which the Neighbourhood Outlier Factor (NOF) is used to assess the anomalous dataset. When compared to the suggested IDS system, which requires less execution time and storage to test the dataset, machine learning approaches detect the intrusion in the computer network with enormous execution time and storage to predict. In this study, the suggested IDS outperforms other current machine learning techniques and can effectively identify nearly all anomalous data in the computer network [10].

3. Existing System

Since the beginning of time, abnormal occurrences have been detected and, if necessary, eliminated from data using outlier detection. Mechanical problems, alterations in system behavior, dishonest behavior, human error, instrument error, or even variations in the population as a whole can all be causes of outliers. System flaws and fraud can be found and stopped before they get worse and have disastrous effects. We used prediction algorithms to detect the precise kind and novel ways to identify outlier presentations. Isolation Forest has various disadvantages, including the inability to detect local anomalous points, which has an impact on the algorithm's performance. It has been commonly employed in the past following outlier detection for further prediction.

Disadvantages:

- Outliers will collapse the total process due to presence of irrelevant data will affect the whole data.
- Inaccuracy- more probability of solution inaccuracy.
- They can also alter the underlying assumption of regression as well as other statistical models.

- Detecting local anomaly point- leads to affect the accuracy of the algorithm.
- Affect variance- In a data distribution, with extreme outliers, the distribution is skewed in the direction of the outliers which makes it difficult to analyze the data.

4. Proposed System

The simple yet important phase in data analysis is outlier analysis. Your datasets can yield stronger conclusions by removing unusual observations, which are typically misleading or incorrect. Outliers can be caused by human, instrument, and natural population variances, fraud, changes in system behavior, and design errors in systems. In our study, we employed a low data accuracy anomaly detecting technique called LOF (Local Outlier Factor) to find outliers that showed up in the data we had collected. If outliers were found, they would be looked at and eliminated before proceeding to the next stage. One of the most effective clustering algorithms, Decision Tree, also known as a supervised algorithm, works by recursively partitioning the data, which is then used to make predictions after identifying outliers from team data analysis. By providing the range of outliers, this algorithm increases accuracy and helps the industry produce transportation more efficiently.

Advantages:

- **Removing outliers**- outliers increase the variability in your data, which decrease the Statistical power, removing outliers becomes statistically significant
- **Increase accuracy-** removing outliers will increases the data accuracy improves the quality of datasets
- **Time efficient-** the provided result helps solve the problem quick and efficient helps in less time consuming and provide more efficient time
- **Uncomplicated-** Simple to understand and to interpret data
- **Measurements errors-** will help in improve measurement errors, data entry and processing errors.

5. System Architecture

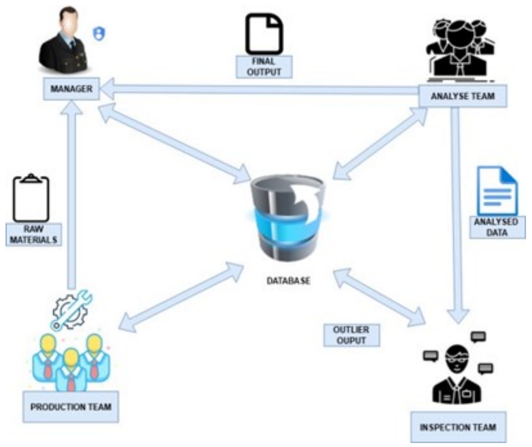


Fig. No 5.1

6. Module Description

1) Production Team

2) Manager

3) Analyse Team

4) Inspection Team

Production: The development of an automotive industry requires the constant involvement of a production team, which is made up of highly trained and skilled individuals. The production team in the automotive industry is essential to increasing industry productivity. In our project, the production team register with details and redirect to the application which the production team can send raw material details which consist of process like (Stamping, Welding, Assembly, and painting) team confirms the details of the raw materials and the production team details that's include (Production manager name, production team email, number of team fields and number of employers currently working in respective fields and the information will be forwarded to the manager, then receives an data of the analysis of the production report from the manager, from the use of that report begins the next process of production.

Manager: The manager has a crucial role to perform in an organisation managing the process inside the firm, or organisation. , Initially manager logs into the application by entering the specified user name and password then initially receive the raw material information from production team and then verify the raw material details that was uploaded by production team that was the preliminary role of manager, and then manager also check the production team details for smooth purposes and then a manager also receive an inspection team details to continue the process that totally depend upon the detail report uploaded by inspection team, and then receives the final production detail from the analysis team, verifies it, and sends it to the production team. According to inspection team report the manager decided to assign work to inspection team.

Analyse Team: The role of analyse team in industry to collect the dataset from various tier cities and by using the collected data the analyse team will using analyse method to provide the recommended solution to production team, after the manager receives the product information from production team, the analyse team from industry collect data set from the users for the further process, following completion of the analysis process by the analysis team, the data consist of (data's from various people), various location ,and various condition , which was collected by analyse team and then it will be forwarded to the inspection team for finding and reducing outliers that is presented in collected data. Once the analyse team's work has been reviewed by the Inspection team, the data will then continue through the analysis process before being forwarded to the manager and then manager verifies further it move onto to the production team.

Inspection Team: The role of Inspection team in the industry was to provide the accurate data , which the data was received from the analyse team the role of inspection team check whether the data was accurate or else the data should be processed correctly and then move on to the further process in that process the inspection team will find the range of outliers that are presented in the collected data, In the beginning information data received from the manager, The inspection team will start their process and , receive data from the analysis team which was collected, and evaluate the collected data using the one of the supervised learning algorithm decision tree algorithm to identify outliers and its range. If outliers are identified in the data, the inspection team will identify the outlier range to classify them, and the data will accurate data will be sent to the analysis team for the data should be in accurate so, it will easy to process for the analyse team.

7. Methodology

In huge datasets with many different data views, a project proposal can efficiently and with less computing complexity identify the outliers. Even though outliers are typically rejected as error or noise, they may contain important information. Because irregular entries make it difficult to analyze and provide engine specs to the production team, finding outliers will help to speed up the process. The industry will benefit from its assistance in avoiding irrelevant data, improving the right production system, and leading to development across a variety of disciplines in addition to finding the most significant differences. In this research, employed a low data accuracy anomaly detecting technique called LOF (Local Outlier Factor) to find outliers that showed up in the data we had collected. If outliers were found, they would be looked at and eliminated before proceeding to the next stage. One of the most effective clustering algorithms, Decision Tree, also known as a supervised algorithm, works by recursively partitioning the data, which is then used to make predictions after identifying outliers from data analysis. By providing the range of outliers, this algorithm increases accuracy and helps the industry produce cars more efficiently.

Proposed Work

1. An Analysis of different methods used to detects the outliers based on the methods accuracy score.

Table 1: Accuracy Level

Algorithms	**Accuracy**
Auto Encoder	65%
Decision Tree	91%
LOF	62%
Isolation Forest	35%

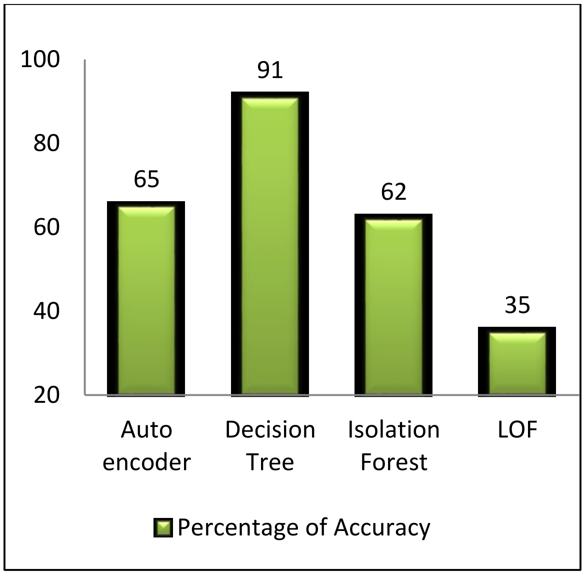


Fig. No 7.1

Auto Encoder: As a neural network learns efficient data representations (encoding), an auto encoder trains the network to ignore signal "noise" through unsupervised learning. With the help of auto encoders, it is possible to perform image denoising, image compression, and in some cases, even the creation of picture data. Auto Encoder offers an average level of accuracy when compared to the other two methods.

Decision Tree: The decision tree classifier had a 91% accuracy rate. The fact that 6 observations have been determined to be erroneous suggests that. Let's begin by showing the model's predicted results. Because this method uses a portion of the training data for the test, it is more accurate. If you offer the decision tree the same data to forecast with, it will provide an exact match because it learned about the data during training. Because of this, decision trees consistently arrive to the correct results.

Isolation Forest: Fei Tony Liu and Zhi-Hua Zhou created the Isolation Forest technique for data anomaly identification in 2008. Binary trees are used by Isolation Forest to find abnormalities. The algorithm's minimal memory need and linear time complexity make it effective for handling large amounts of data. Isolation Forest Algorithm, however, provided medium accuracy.

LOF: The local density deviation of a specific data point in relation to its neighbors is determined by the Local Outlier Factor (LOF) method. It is an unsupervised strategy for identifying anomalies. Outliers are samples that are significantly less dense than their surrounding samples. This method's accuracy rating is incredibly poor when compared to other approaches.

8. Results

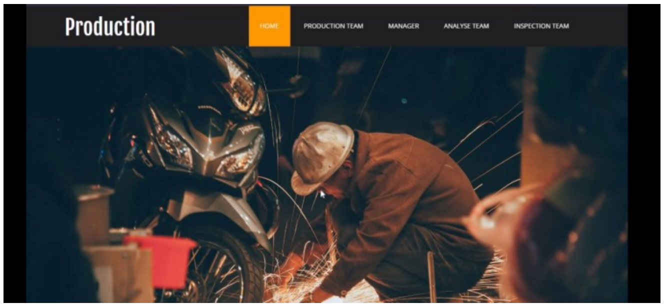


Fig. No 8.1

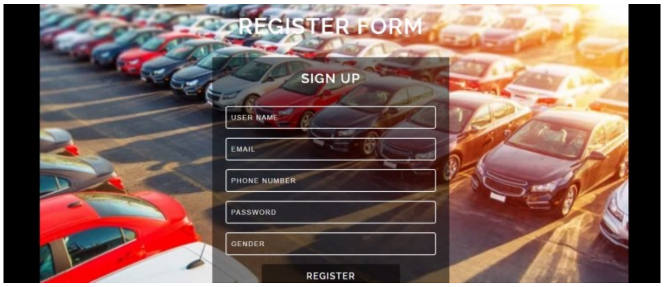


Fig. No 8.2

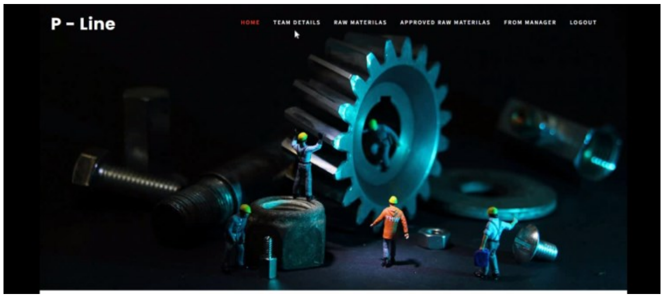
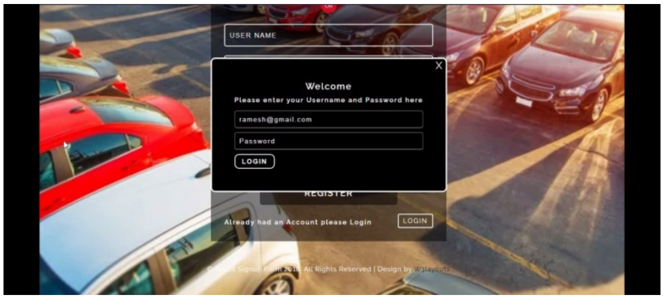


Fig. No 8.3


USER NAME
X
LOGIN
REGISTER
LOGIN
© Signup Form 2018. All Rights Reserved | Design by W3layouts

Welcome

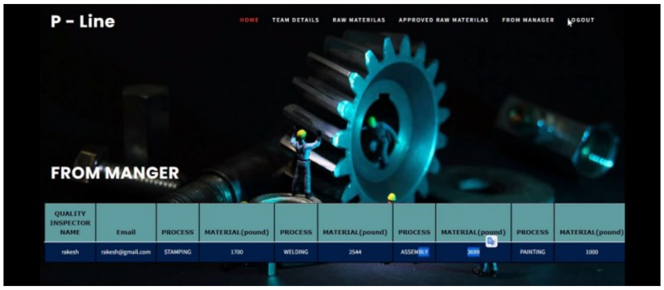
Please enter your Username and Password here

ramesh@gmail.com

Password

Already had an Account please Login

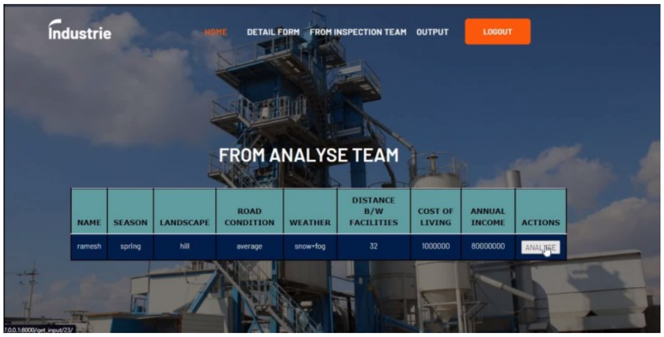
Fig. No 8.4



FROM MANGER

QUALITY INSPECTOR NAME	Email	PROCESS	MATERIAL(pound)	PROCESS	MATERIAL(pound)	PROCESS	MATERIAL(pound)	PROCESS	MATERIAL(pound)
rakesh	rakesh@gmail.com	STAMPING	1700	WELDING	2544	ASSEMBLY	3699	PAINTING	1000

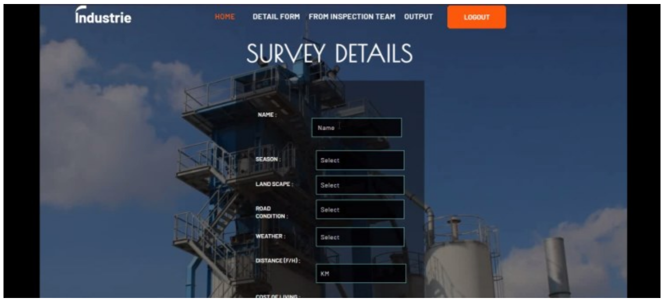
Fig. No 8.5



FROM ANALYSE TEAM

NAME	SEASON	LANDSCAPE	ROAD CONDITION	WEATHER	DISTANCE B/W FACILITIES	COST OF LIVING	ANNUAL INCOME	ACTIONS
ramesh	spring	hill	average	snow+fog	32	1000000	80000000	ANALYSE

Fig. No 8.6



SURVEY DETAILS

NAME : Name

SEASON : Select

LAND SCAPE : Select

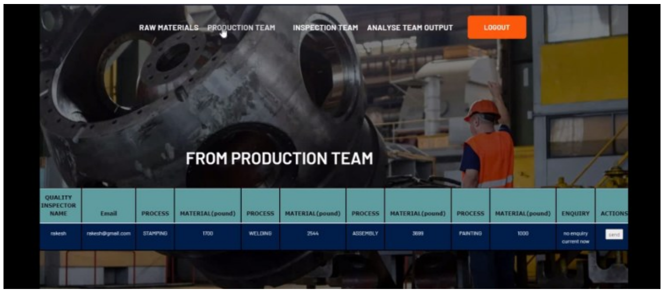
ROAD CONDITION : Select

WEATHER : Select

DISTANCE (F/H) : KM

COST OF LIVING :

Fig. No 8.7



FROM PRODUCTION TEAM

QUALITY INSPECTOR NAME	Email	PROCESS	MATERIAL(pound)	PROCESS	MATERIAL(pound)	PROCESS	MATERIAL(pound)	PROCESS	MATERIAL(pound)	ENQUIRY	ACTIONS
rakesh	rakesh@gmail.com	STAMPING	1700	WELDING	2544	ASSEMBLY	3699	PAINTING	1000	no enquiry current now	send

Fig. No 8.8

9. Conclusion

As a consequence, we proposed a statistical analysis method for detecting outliers from the retrieved data in the same way. This project is a general study of the performance of facing outliers in a clustering algorithm, so many know substantial algorithms like Z-core method, isolation forest, and graphical approach method are some popular method for doing so. In order to find the outlier items in a big information space, a variety of alternative algorithms have been developed. Removing outliers is not our primary goal because the data will be more accurate and accumulate after they have been removed. Instead, we use the decision tree algorithm to find the specifications and provide solutions. However, because data inaccuracies can lead to a variety of results, it is crucial to improve data analysis and the outlier finding system by utilizing various algorithms, methods, and techniques.

References:

[1]. K. Devika, Dwarikanath Mahapatra, Ramanathan Subramanian, and Venkata Ramana Murthy Oruganti, “Outlier-Based Autism Detection Using Longitudinal Structural MRI”, Vol 10, 2022.

[2]. Siyu Luan, Zonghua Gu, Leonid B. Freidovich Lili Jiang, And Qingling Zhao, “Out-of-Distribution Detection for DNN With Isolation Forest and Local Outlier Factor”, Vol. 9, 2021.

[3]. Chang-Min Kim, Ellen J. Hong, and Roy C. Park, “Chest X-Ray Outlier Detection Model Using Dimension Reduction and Edge Detection” Vol 9, P 780, 2021.

[4]. Lukas Ruff, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft Thomas G. Dietterícha, “Unifying Review of Deep and Shallow Anomaly Detection”, Vol. 109, No. 5, 2021.

[5]. Ruiheng Zhang, Quan Zhou, Lulu Tian, Libing Bai and Jie Zhang, “A Novel Outlier Detection Model for Vibration Signals Using Transformer Networks”, Vol.10, 2022.

[6]. Atiq ur Rehman and Samir Brahim Belhaouari, “Unsupervised outlier detection in multidimensional data”, Vol. 8, 2021.

[7]. Dipannita Kar, Mr. Haresh Chande, and Mr. Rajendra Gaikwad, “A Study Paper on Outlier Detection on Time Series Data”, Vol. 5, 2017

[8]. Karanjit Singh and Dr. Shuchita Upadhyaya, “Outlier Detection: Applications and Techniques”, Vol.9, 2012.

[9]. Deepika Pahuja and Romika Yadav “Outlier

Detection for Different Applications: Review”, Vol. 2, 2013.

[10]. Jabez J and Dr.B.Muthukumar “Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach”, 2015.