ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2239

An Outlier Detection System to Enhance Decision-Making for Akperan Orshi Polytechnic Yandev

Nomishan, Iyorrumun[1]; Wayo, Iorkyaa[2]

[1,2] Computer Science Department, Akperan Orshi Polytechnic, Yandev Gboko Benue State. Nigeria

Publication Date: 2025/06/09

Abstract: In the era of data-driven governance, institutions like Akperan Orshi Polytechnic Yandev are increasingly relying on digital systems to streamline operations and enhance decision-making. However, hidden anomalies or outliers in institutional data—such as irregular staff attendance, student performance, or financial transactions—can distort analyses and lead to suboptimal decisions. This study presents the design and implementation of an Outlier Detection System tailored for Akperan Orshi Polytechnic Yandev. The system employs statistical and machine learning techniques to automatically identify and flag unusual patterns across various administrative and academic datasets. By integrating this system into the Polytechnic's existing data infrastructure, stakeholders can proactively detect inconsistencies, improve data integrity, and make more informed and timely decisions. The results demonstrate the system's effectiveness in revealing hidden anomalies, thereby supporting strategic planning and policy formulation across departments.

Keywords: *Outlier Detection, Decision-Making, Data Integrity, Machine Learning, Anomaly Detection, Polytechnic Administration, Akperan Orshi Polytechnic Yandev, Institutional Data Analysis.*

How to Site: Nomishan, Iyorrumun; Wayo, Iorkyaa; (2025) An Outlier Detection System to Enhance Decision-Making for Akperan Orshi Polytechnic Yandev. *International Journal of Innovative Science and Research Technology,* 10(5), 3996-4002. https://doi.org/10.38124/ijisrt/25may2239

I. INTRODUCTION

Organization and businesses use different technologies to perform analytics, such as simple reporting and dashboarding, which are just reports to inspect and study previous performance (Bumblauskas et al., 2017). However, Organization and businesses certainly use technologies as a means of data analytics in the perspective of digital transformation (OECD-BEIS, 2018). Organization and businesses have a common aim to use digital technologies as an opportunity to improve decision-making competency along the business digital transformation journey (Schwertner, 2017). Advanced data analytics and artificial intelligence are drivers of deep analysis and change the businesses (West and Allen, 2018). Data mining, as a confluence of statistics and machine learning (artificial intelligence), makes businesses predict future trends (Sethi et al., 2016).

Digitalization enables huge developments in after sales services; with prescriptive and predictive analysis opportunities, businesses reduce life-cycle costs and optimize costs (Rudnick et al., 2020). Organization and businesses can identify their customers' behavioral patterns, improve the availability of production facilities, enable evidence-based corporate decision-making, and reduce innovation latency via advanced analytics (Gimpel et al., 2018). Moreover, the customer journeys do not end at the point of sale; digital technologies enable the after sales services strategies of businesses and industries revolutionize (Gimpel et al., 2018).

Business processes or business re-engineering is a process that links to the enhancement of business performance. The aim of business re-engineering is to attain maximum enhancement in the business (Krstic et al., 2016).

In the ever-evolving landscape of institutions, whether educational, governmental, or corporate, the vast amount of data generated daily has the potential to be a goldmine of insights that can revolutionize decision-making processes. Unlocking Institutional Insights through a Data Mining Approach holds the promise of transforming raw data into actionable intelligence, enabling institutions to make informed decisions, optimize operations, and achieve strategic objectives. This methodology involves the systematic application of advanced data mining techniques to extract patterns, relationships, and valuable knowledge from complex datasets, ultimately enhancing the efficiency and effectiveness of decision-making within an institution.

Traditionally, institutional decision-making relied on historical data and human intuition, often constrained by the limitations of manual analysis and the inability to process

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2239

vast datasets. With the advent of data mining techniques, institutions can leverage the power of algorithms and machine learning to uncover hidden patterns and trends within their data. This shift from intuition-driven decisions to evidence-based decision-making marks a paradigm shift in institutional management and strategy.

In an era of information abundance, institutions accumulate vast amounts of data. Leveraging this data through data mining techniques provides an opportunity to uncover hidden patterns, trends, and knowledge that can lead to informed decision-making. This research explores the application of data mining within our institution to harness the full potential of available data resources.
Specifically, the aim of this study was to implement an outlier detection system that could serve as a valuable tool in detecting abnormal values in examination scores and budgetary values. The system can lead to an improved decision-making and increased efficiency in the moderation of examination results and budget estimates for various departments of the Polytechnic.

II. RELATED LITERATURE

A. Outliers and Outlier Detection

In data science, an outlier is a data point that significantly differs from other observations in a dataset (Duraj & Szczepaniak, 2020; Singh & Upadhyaya, 2012; Smiti, 2020). It lies far outside the range of most values and may indicate variability in measurement, experimental errors, or novel insights. Some key characteristics of outliers are as follows: they are unusually high or low compared to the rest of the data, they often represent rare or exceptional events, they can skew statistical analyses and affect the performance of machine learning models (ur Rehman, & Belhaouari, 2021). For example, in a dataset of student test scores: [70, 75, 72, 74, 73, 71, 99], the value 99 could be an outlier as most students scored around 70–75. According to Hodge and Austin (2024), outliers could be caused by several factors including measurement errors, data entry errors, experimental variation, and true variability.

Outlier detection is the process of identifying data points that significantly differ from the majority of a dataset (Smiti, 2020; Won, 2020). These unusual values, known as *outliers*, can indicate variability in measurement, experimental errors, or rare events (e.g., fraud, cyber-attacks, or system failures). Detecting outliers is critical in many domains because they can distort statistical analyses or reveal important, rare phenomena (Mel-Gal, 2005).

Literature evidence has shown that outlier detection methods have been broadly categorized into three groups (Acuna & Rodriguez, 2004; Ben-Gal, 2005; Ski, 2024). These are statistical-based outlier detection methods, outlier detection using clustering, and distance-based outlier detection methods. Statistical-based methods either assume that a set of observations follow a particular distribution or the underlying distribution is assumed (Ben-Gal, 2005). Those observations that deviate from the model assumptions are flagged as outliers. Examples of the techniques in this category include Z-Score/Standard Score, Grubbs' test, and Boxplot method. In clustering-based outlier detection methods, scattered outliers will form a cluster of size 1 and clusters of small sizes can be considered as clustered outliers (Acuna & Rodriguez, 2004). Examples of clustering techniques used for outlier detection include DBSCAN (Density-Based Spatial Clustering of Applications with Noise), K-means clustering, and LOF (Local Outlier Factor). Distance-based outlier detection methods use distance measures to detect outliers where an instance with very few neighbors within a distance λ can be regarded as stronger outlier as a an instance with more neighbors within a distance λ (Acuna & Rodriguez, 2004). An example is k-Nearest Neighbors (k-NN) which measures the distance of a data point to its neighbors; large distances can indicate outliers.

Outlier detection has broad applications across industries. These include fraud detection, cybersecurity (for intrusion detection in networks, where unusual patterns may indicate attacks), healthcare (identifying abnormal patient records or laboratory test results for early disease detection), environment monitoring (detecting abnormal weather conditions, pollution spikes, or seismic activities), finance (unusual stock price movements or trading volumes), and e-commerce and web analytics (identifying fake reviews, bot activity, or unusual user behavior) (Ski, 2024).

B. Application of Outlier Detection for the Moderation of Examination Results and Budget Estimates

- *Outlier Detection for the Moderation of Examination Results*

Outlier detection in the context of examination moderation is a crucial statistical tool used to identify unusually high or low scores that could indicate errors, unfair marking, cheating, or exceptional performance (Consul & Ndiwari, 2018). This helps ensure fairness, reliability, and consistency in the grading process. According to Yu (2009), detecting outliers as a form of exam moderation has several benefits including the identification of marking errors or inconsistencies, ensuring no student is unfairly advantaged or disadvantaged, identifying possible cheating or collusion, ensuring score distributions are within expected norms, identifying if an examiner consistently marks higher or lower than others. Other benefits include spotting items with unusual score patterns (too easy, too hard, or poorly worded), detecting scores that deviate significantly from a student's other performances or from cohort norms, and comparing performance across subjects to ensure consistent difficulty levels (Yu, 2009).

Following the identification of outliers in examination scores, a number of actions could be carried out to salvage the situation. These include double-checking marked scripts, applying moderation curves or scaling, investigating context (e.g., illness, cheating suspicion), and or the exclusion of extreme values from statistical calculations (Consul & Ndiwari, 2018; Yu, 2009).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2239

➤ *Outlier Detection for the Moderation of Budget Estimates*

Outlier detection for moderation of budget estimates is a critical process to ensure accuracy, consistency, and reasonableness in financial planning. It helps identify budget items that are significantly higher or lower than expected, possibly due to data entry errors, unrealistic assumptions, or intentional manipulation (Adams et al., 2019). Errors or anomalies can be identified in budget submissions by flagging unrealistic estimates for review. This can help to maintain fairness and consistency across departments or projects.

In order to detect outliers in a budget, the following procedure is followed: collect the relevant budget data, clean and normalize the data to ensure consistency, set thresholds based on business rules or past trends, use a layered approach which combines statistical methods with rule-based filters, and then prioritize flagged items by severity (e.g., mild, moderate, extreme outlier) (Adams et al., 2019).

III. MATERIALS AND METHODS

A. Outlier Detection using Z-Score

Outlier detection using the Z-score is a statistical technique based on the standard deviation of the dataset (Chikodili et al., 2021; Yaro et al., 2023). It is commonly used to identify data points that are significantly different from the rest of the data. According to (Chikodili et al., 2021; Yaro et al., 2023), the formular for Z-**score** is given in Equation (1).

$$Z = (x-\mu)/\sigma \tag{1}$$

where,

x = the data point, μ = the mean of the dataset, σ = the standard deviation of the dataset. The Z-score tells you how many standard deviations a data point is from the mean. The commonly used threshold value for outlier detection is ±3 (Yaro et al., 2023). That is, when the Z-score value of an observation is ≥ ±3, it is classified as an outlier.

B. Algorithm Design for Outlier Detection Using Z-Score

In this section, the algorithm to implement the Z-score formula as an outlier detector was designed. The six steps involved in the procedure are listed in Listing 1. In step 1, the dataset from which outliers are to be detected is uploaded to the computer. In step 2, the mean and standard deviation parameters are computed from the dataset. In step 3, the Z-score is computer for each data point while in step 4, the threshold values are determined. In steps 5 and 6, the outliers are flagged and printed out respectively.

➤ *Listing 1: Algorithm to compute Z-score*

- **Input:** A dataset X with numerical values.
- **Compute the mean (μ) and standard deviation (σ) of the dataset.**
- **For each data point x in X, compute the Z-score:**
- $Z = (x-\mu)/\sigma$
- **Set a threshold: 3 or -3.**
- **Flag any data point as an outlier if Z > threshold**
- **Output: Outliers**

C. System Design with Unified Modeling Language

Unified Modeling Language (UML) offers a wide range of tools to enable the design of computer applications (Atsa'am et al., 2024; Bell, 2023; Oyelere et al., 2018). Specifically, the UML class diagram and UML activity diagram were employed for application design in this study.

➤ *System Design using Class Diagram*

The class diagram in Figure 1 uses classes, methods and attributes to describe the static view of the system (Atsa'am et al., 2024). The figure consists of three classes: *Z-Score, UploadData,* and *Outlier*. Each class consists of various attributes and methods which are executed in the process of determining outliers from a dataset.

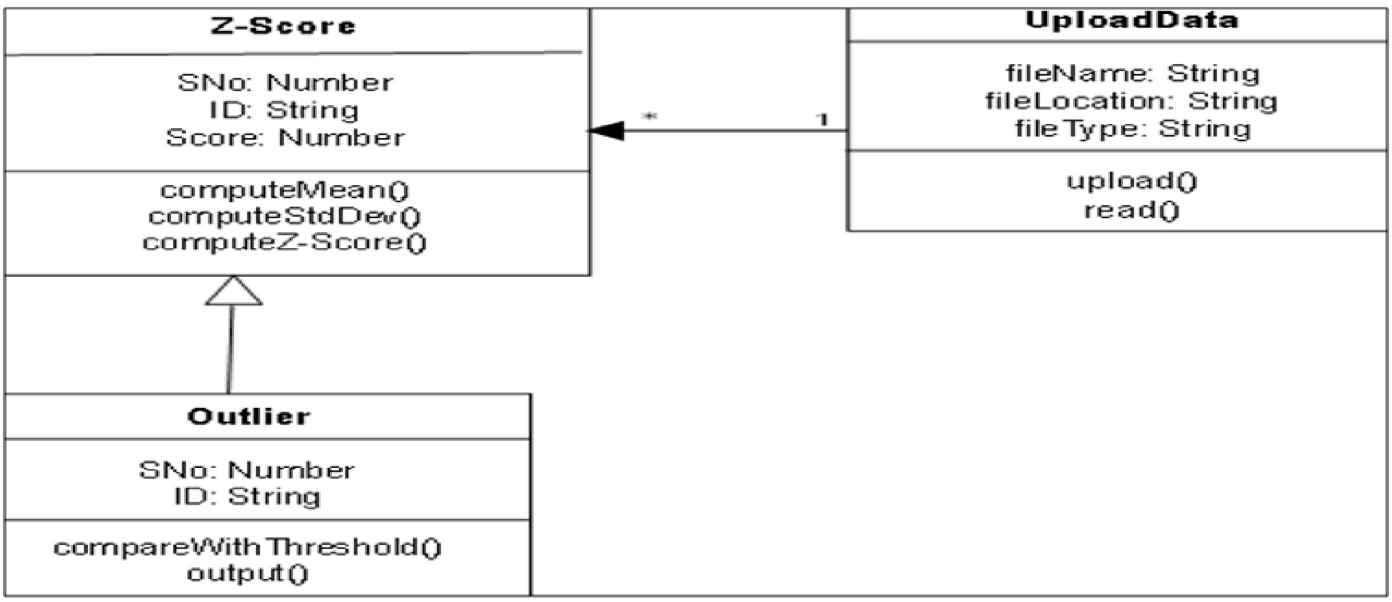

Fig 1 Class Diagram of the System

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2239

➢ *System Design using Activity Diagram*

The activity diagram in Figure 2 represents the flow control and operations of the system (Bell, 2023). The procedures and activities required to calculate Z-score for each observation are illustrated. Each Z-score value is compared with the threshold. -3 and 3, and observations with Z-score values above these thresholds are considered as outliers (Anusha et al., 2019).

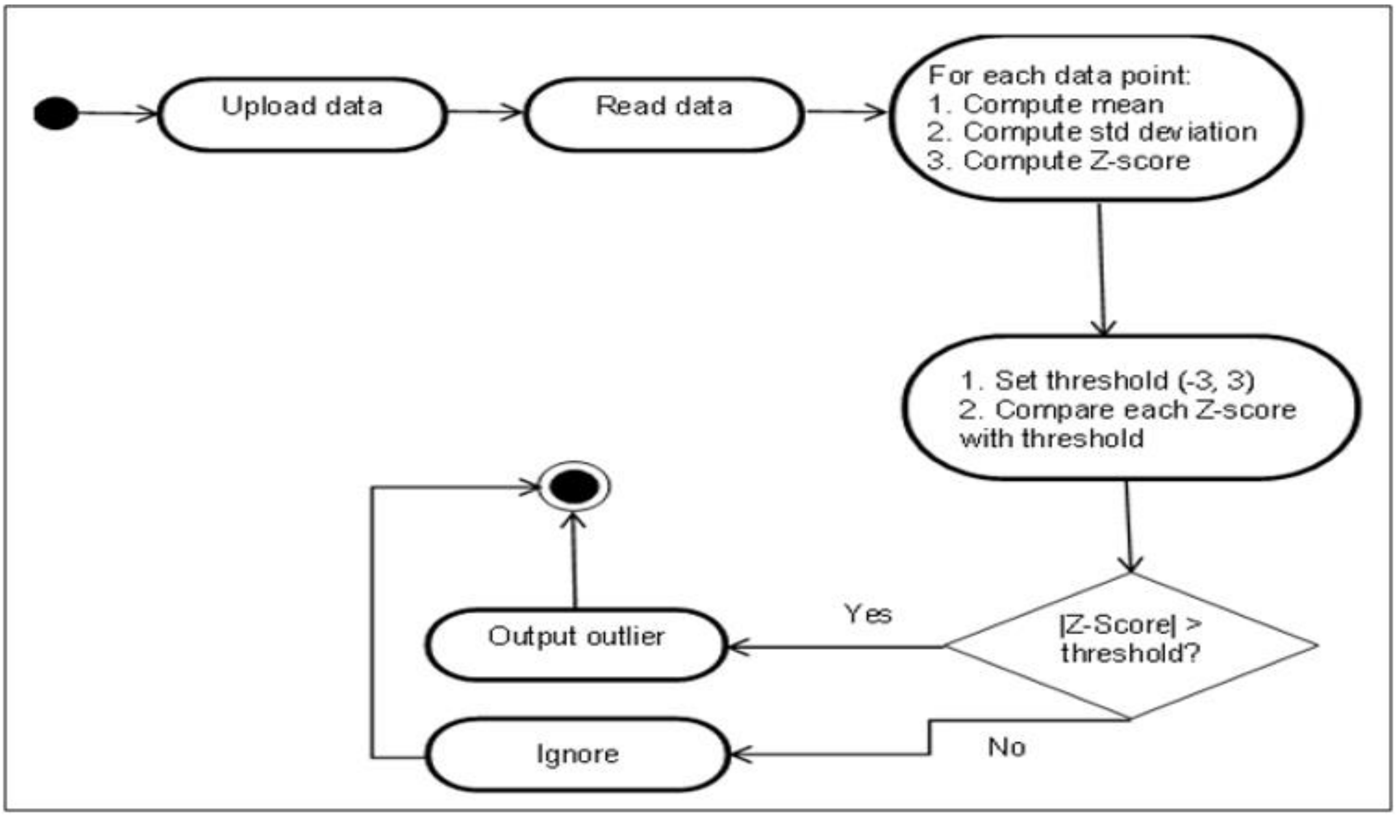

Fig 2 Activity diagram of the system

D. Conceptual Input and Output Design

Figure 3 illustrates the conceptual input design of the system. An input form is provided where users make use of input boxes for data entry to the application. The information gives a description of the dataset from which outliers are to be detected.

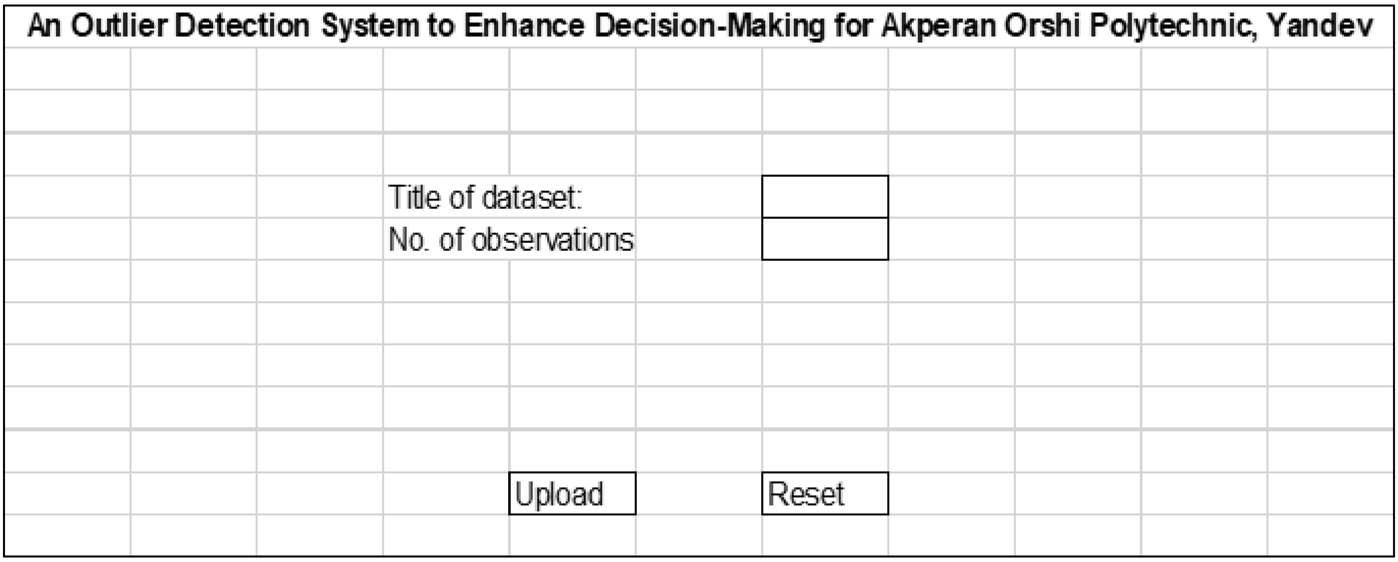

Fig 3 Conceptual input design

Figure 4 illustrates the conceptual output design of the system. Conceptually, the interface consists of information about the dataset and a list of the outliers detected from the dataset.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2239

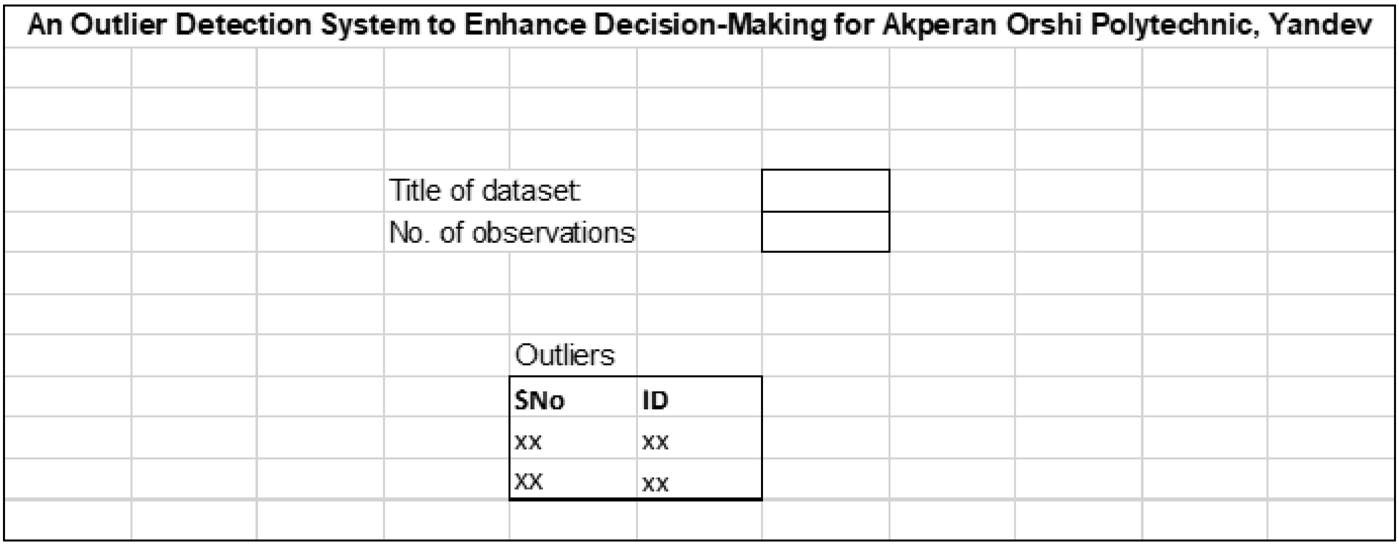

Fig 4 Conceptual output design

IV. RESULTS AND DISCUSSION

A. System Implementation and Graphical User Interfaces

The Z-score algorithm and other design aspects of the system were implemented into functional components to form a computer application. With the applications, a user can determine outliers from a dataset with the use of key press and mouse-click events.

Figures 5 and 6 are the two main graphical user interfaces (GUI) of the application. The GUI in Figure 5 enables the user to upload a dataset to the application for the detection of outliers. A list of outliers (if any) detected from the dataset is presented to the user via the GUI in Figure 6.

<Figure 5 here>

<Figure 6 here>

B. Model Simulation Cases

Various simulations were conducted to evaluate the performance of the system. The simulation cases including their objectives and expected outcomes are shown in Table 1.

Table 1 Simulation cases

S/No	Simulation Case	Simulation Objective	Expected Outcome(s)
1	SC-1	To confirm that the application can successfully launch to the main menu	The application should be able to launch and display the main menu.
2	SC-2	To confirm that the input tools function correctly and the application can detect outliers from uploaded datasets.	i. The input tools and the upload function of the application should function correctly. ii. The application should compute outliers from the uploaded dataset, detect and output outliers if they are present.

C. Simulation Results and Discussion

The various simulation cases produced results which were compared against the expected outcomes outlined in Table 1. In this section, the simulation results are presented and discussed.

➢ *Simulation Case 1 (SC-1): Launching the Application*

Launching of the application was simulated in SC-1. The output shown in Figure 7 confirms that the application can successfully launch to the main menu.

<Figure 7 here>

➢ *Simulation Case 2 (SC-2): Simulation of Data Input and Outlier Detection*

This simulation confirmed that the input tools of the application including the upload function function as required. Furthermore, it was confirmed that the application can successfully detect outliers from a dataset. Several sample datasets were employed to conduct this simulation case, and the output are shown in Figures 8 and 9.

<Figure 8 here>

<Figure 9 here>

D. Practical Applications for Akperan Orshi Polytechnic

The outlier detection application developed in this study is a user-friendly tool for administrative decision-making. The application is intended for use by the Polytechnic for the detection of unusual examination scores and budget estimates across departments.

When outliers are identified in examination scores, a number of follow-up actions could be taken by the administration. These include double-checking marked

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2239

scripts, applying moderation strategies, or investigating possible incidents of cheating or favouritism (Consul & Ndiwari, 2018; Yu, 2009). Such follow-up actions build stakeholder confidence and help ensure fairness, reliability, and consistency in the grading process. Furthermore, it should be a concern to the administration to ensure that students are not unfairly advantaged or disadvantaged.

In addition to outlier detection in examination scores, this application is a valuable tool in the detection of outliers in budget estimates submitted by the departments. The application can help in identifying budget items that are unusually higher or lower than expected. Several factors could be responsible for such anomalies. These include data entry errors, unrealistic assumptions, or intentional manipulation (Adams et al., 2019). The application can flag such abnormal estimates for administrative review. The outcome could maintain fairness and consistency across departments or projects.

V. CONCLUSION

Examination and budget moderation are an integral part of the administrative functions of a Polytechnic. To enhance a guided decision-making process and increase efficiency, this study leveraged on the data mining concept of outlier detection to develop a moderation system that utilizes Z-score. This technique identifies data points that are significantly different from the rest of the data. Flowing from this, the application has the potential to detect unusual examination scores in various courses/modules in a class. In addition, the application can isolate budget estimates that are unusually higher or lower than other departments or items. The outcome produced by the application can aid administrative decision-making that ensures no student is victimized or favoured in an examination. Further, abnormal budget items can be isolated for administrative review to maintain fairness, consistency and prudence in financial management in the Polytechnic. There are several techniques for outlier detection, However, this study employed the Z-score technique. In a future study, other methods should be implemented and the results compared with this study for confirmation.

I wish to sincerely appreciate the Tertiary Education Trust Fund (TETFund) for sponsoring me to undertake this Institutional Based Research (IBR). This opportunity has significantly contributed to the advancement of my academic and professional development. The support provided by TETFund not only enabled the successful execution of this research but also reinforced the importance of research in solving real-world challenges within our institutions and communities.

Thank you, TETFund, for your unwavering commitment to the growth of research and development in Nigeria's tertiary institutions.

REFERENCES

[1]. Acuna, E., & Rodriguez, C. (2004). A meta analysis study of outlier detection methods in classification. Technical paper, Department of Mathematics, University of Puerto Rico at Mayaguez, 1(25), 14.
[2]. Adams, J., Hayunga, D., Mansi, S., Reeb, D., Verardi, V. (2019). Identifying and treating outliers in finance. Financial Management, 48, 345–384. doi:10.1111/fima.12269
[3]. Anusha, P. V., Anuradha, C., Murty, P. C., & Kiran, C.S. (2019). Detecting outliers in high dimensional data sets using z-score methodology. International Journal of Innovative Technology and Exploring Engineering, 9(1), 48-53. doi: 10.3594 0/ijitee.A3910.119119
[4]. Atsa'am, D.D., Gbaden, T., & Wario, R. (2024). DrugApp: A simulation of drug suspects and offenders classification. International Journal of Simulation and Process Modelling, 21(3), 147-154. doi: 10.1504/IJSPM.2024.10066647
[5]. Bell, D. (2023, June 25). An introduction to the unified modeling language. https://developer.ib m.com/articles/an-introduction-to-uml/
[6]. Ben-Gal, I. (2005). Outlier Detection. In: Maimon, O., Rokach, L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA. https://doi.org/10.1007/0-387-25465-X_7
[7]. Bumblauskas, D., Nold, H., Bumblauskas, P., and Igou, A. (2017). Big data analytics: transforming data to action. Business Process Management Journal, 703-720.
[8]. Chikodili, N. B., Abdulmalik, M. D., Abisoye, O. A., Bashir, S. A. (2021). Outlier detection in multivariate time series data using a fusion of k-medoid, standardized euclidean distance and z-score. In: Misra, S., Muhammad-Bello, B. (eds) Information and Communication Technology and Applications (ICTA 2020). Communications in Computer and Information Science, 1350. Springer, Cham. doi:10.1007/978-3-030-69143-1_21
[9]. Consul, J. I., & Ndiwari, G. (2018). The use of outliers in the detection of suspicious examination malpractices in the scores of students in a Nigerian university. International Journal of Education and Research, 6(3), 65-74.
[10]. Duraj, A., Szczepaniak, P. S. (2020). Outlier detection in data streams — A comparative study of selected methods. Procedia Computer Science, 192, 2769-2778. doi:10.1016/j.procs.2021.09.047.
[11]. Gimpel, H., Hosseini, S., Huber, R. X. R., Probst, L., Röglinger, M. and Faisst, U. (2018). Structuring Digital Transformation: A Framework of Action Fields and its Application at ZEISS. Journal of Information Technology, Theory and Application, 1(19).
[12]. Hodge, V. J. & Austin, J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22 (2), 85-126.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25may2239

[13]. Krstic, J., Jovanov, G., Radovanovic, R., Ljusic, M. and Nikolic, M. (2016). Process of Business Reengineering from the Aspect of E-Business. Journal of Textile Science & Engineering, 272(6).
[14]. OECD and the United Kingdom Department for Business, Energy and Industrial Strategy (BEIS). (2018). Implications of the Digital Transformation for the Business. http://www.oecd.org/sti/ind/digital-transformation-business-sector-summary.pdf.
[15]. Oyelere, S. S., Atsa'am, D. D., Ayuba, H. M., Olawumi, O. Suhonen, J., & Joy, M. (2018). TerrorWatch: A prototype mobile app to combat terror in terror-prone nations. In F.J. Mtenzi (Ed) Mobile technologies and socio-economic development in emerging nations (pp 203-233) Hershey, PA: IGI Global. doi:10.4018/978-1-5225-4029-8.ch010
[16]. Rudnick, M., Riezebos, J., Powell, D. J. and Hauptvogel, A. (2020). Effective after-sales services through the lean servitization canvas. International Journal of Lean Six Sigma, 5(11), 943-956.
[17]. Schwertner, K. (2017). Digital transformation of business. Trakia Journal of Sciences, 15(1), 388-393.
[18]. Sethi, S., Malhotra, D., and Verma, N. (2016). Data mining: current applications & trends. International Journal of Innovations in Engineering and Technology, 6(4), 586- 589.
[19]. Singh, K., & Upadhyaya, K. (2012). Outlier Detection: Applications and techniques. International Journal of Computer Science Issues, 9(1),307-323.
[20]. Ski, M. C. (2024). An overview of outlier detection methods. Lond Journal of Engineering Research, 24(2), 37-79.
[21]. Smiti, A. (2020). A critical overview of outlier detection methods. Computer Science Review, 38, 100306. doi: 10.1016/j.cosrev.2020.100306
[22]. ur Rehman, A., & Belhaouari, S. B. (2021). Unsupervised outlier detection in multidimensional data. Journal of Big Data, 8(80), 1-27. (2021). doi:10.1186/s40537-021-00469-z
[23]. West, D. M. and Allen, J. R. (2018). How artificial intelligence is transforming the world. Report, Brookings Institution.
[24]. Won, M. (2020). Outlier analysis to improve the performance of an incident duration estimation and incident management system. Transportation Research Record, 2674(5), 486-497. doi: 10.1177/0361198120916472
[25]. Yaro, A. S., Maly, F,. & Prazak, P. (2023). Outlier detection in time-series receive signal strength observation using z-score method with S_n scale estimator for indoor localization. Applied Sciences, 13, 3900. doi:10.3390/app13063900
[26]. Yu, T. W. (2009). The moderation of liberal studies school based assessment scores: How to ensure fairness and reliability? International Education Studies, 2(4), 91-98.