Proceedings of the Multidisciplinary Research Community
MultiVerse 2024
ISSN: XXX-XXX-XXXXX-X-X
Desember 2024
Hal: 434 - 438

Detecting Unusual Student Academic Grades Using the Isolation Forest Method

Hendra Nusa Putra [1*], Nanda Tommy Wirawan [2]
STIKES Dharma Landbouw Padang, Indonesia [1,2]
Coressponding author: hendranusaputra@journals.ai-mrc.com

Abstract

This study aims to detect anomalies in student final grades in the Computer Networks course using the Isolation Forest algorithm. The dataset comprises student grades for the current semester, encompassing variables such as Discipline, Practice, Final Semester Exam (UAS), and Final Grades. The Isolation Forest algorithm was employed to identify students whose grades are inconsistent compared to their peers. The analysis revealed that 10 students were identified as having anomalous final grades. These anomalies may result from various factors, including personal circumstances, variations in study effort, external influences, assessment methodologies, and potential data inaccuracies. Grade distribution was analyzed using histograms and boxplots, indicating the presence of outliers in several variables. Correlations between variables were examined through a heatmap, which showed that the UAS has the strongest correlation with Final Grades. The identification of anomalies was visualized through scatter plots to facilitate understanding of the distribution and detection of outliers in student final grades. The conclusions underscore the need for further investigation to comprehend the root causes of these anomalies and to implement suitable interventions. This study also highlights the advantages of employing data mining techniques in education to enhance the quality of assessments and the monitoring of student academic performance.

Keywords: Anomalies; Detection; Isolation Forest; Student Grades; Education

INTRODUCTION

In the context of education, assessing students' academic performance is fundamental to both tracking their progress and enhancing learning outcomes (Xu, & Zhou, 2022). This process is crucial for understanding how well students grasp course materials and for pinpointing areas where improvements are needed. One common challenge in this area is the presence of anomalies in academic data, such as grades, which may indicate a variety of underlying issues (Marx, Krötzsch, & Thost, 2017). Such anomalies can be signs of inconsistencies in how assessments are conducted, variations in the amount of effort put forth by students, or external factors that affect performance.

This study specifically examines anomalies in the final grades of a Computer Networks course. The anomalies can provide critical insights into potential discrepancies or irregularities in academic performance that might not be apparent at first glance (Hariri, Kind, & Brunner, 2019). Detecting and understanding these anomalies is essential for educators and administrators to ensure that the grading process is fair and that any issues affecting students' performance are properly addressed.

To achieve this, the research utilizes the Isolation Forest algorithm, a machine learning-based anomaly detection technique. The algorithm is particularly effective at identifying data points that significantly deviate from the expected pattern or distribution, which in this case are the students' grades. By analyzing these deviations, the research aims to identify students whose academic performance is significantly different from their peers, providing a basis for further investigation.

Through the application of this algorithm, the study seeks to ensure that the assessment methodologies employed in the course are both accurate and equitable. If the algorithm detects

ISSN: XXX-XXX-XXXXX-X-X
Desember 2024
Hal: 434 - 438

anomalies that are linked to issues such as assessment bias, inconsistent grading practices, or particular student challenges, these can be addressed to improve the overall educational process. Identifying these anomalies also allows for interventions to be put in place, supporting students who may need additional resources or assistance.

Furthermore, the findings of this study are crucial not only for individual student support but also for refining the overall grading system and academic policies. The ability to detect and analyze anomalies can enhance the quality of education by ensuring that student evaluations are comprehensive and reflect their true academic abilities. This ultimately supports the goal of providing a fair and supportive learning environment.

In conclusion, by leveraging advanced data mining techniques like the Isolation Forest algorithm, the study contributes significantly to the field of education. It demonstrates how such techniques can be employed effectively to monitor academic performance, improve grading fairness, and offer targeted support to students who might be struggling. This approach underscores the importance of using data-driven strategies in education to foster better learning outcomes and more reliable assessment practices.

METHODS

This study employs the Isolation Forest algorithm, a machine learning technique widely recognized for its effectiveness in anomaly detection. The data collected for the analysis originated from student grades in the current semester, focusing specifically on the Computer Networks course (Feng, 2019). The primary variables under consideration include "Discipline," "Practice," "Final Semester Exam (UAS)," and "Final Grades." These variables are critical in assessing the academic performance of students and understanding how different aspects of their coursework contribute to their final evaluation. To detect any anomalies in the data, the Isolation Forest algorithm was applied to pinpoint grades that appeared inconsistent when compared to the broader set of student results.

To facilitate a comprehensive analysis, several visualization methods were used. Histograms, boxplots, and scatter plots were generated to examine the distribution and spread of the grades, helping to identify any outliers or unusual patterns within the data set. These visual tools provide a clear view of how the grades are distributed and highlight any instances where a student's performance deviates significantly from their peers. Additionally, a heatmap was created to explore the correlations between the different variables, aiming to understand how each factor, such as "Discipline" or "UAS," may influence the "Final Grades." This correlation analysis allows for a deeper understanding of the relationships between different components of the course and their combined effect on student outcomes.

RESULTS

The use of the Isolation Forest algorithm in this study successfully identified 10 students whose final grades were considered anomalous. These anomalies were detected based on significant deviations in their grades when compared to the overall performance of the cohort. A variety of factors could potentially explain these anomalies, including personal issues affecting student performance, varying levels of effort and study habits, external influences such as life events or

ISSN: XXX-XXX-XXXXX-X-X
Desember 2024
Hal: 434 - 438

distractions, and inconsistencies in assessment methodologies. Additionally, the possibility of data errors was taken into account to ensure the accuracy and validity of the results.

To further explore the nature of these anomalies, the study conducted a detailed distribution analysis using histograms and boxplots. These visualizations revealed the presence of outliers, especially in variables like "Practice" and "Final Semester Exam (UAS)." Such outliers indicate that certain students exhibited performance patterns that were significantly different from the majority, warranting closer examination. Furthermore, a correlation analysis was carried out to understand the relationships between different assessment components. The results indicated that the "Final Semester Exam (UAS)" held the strongest correlation with the "Final Grades," suggesting that the exam's results played a crucial role in determining the overall performance of students in the course.

The study also utilized scatter plots to visualize the distribution of grades across different variables and to facilitate a deeper understanding of the detected anomalies. The scatter plots provided a clear visual representation of how the final grades were spread across various components of the course, aiding in the identification of patterns or clusters of anomalies. This visualization proved to be an effective tool for not only detecting outliers but also for interpreting the impact of different factors on student performance, thereby providing a comprehensive view of how these anomalies emerged within the academic data.

DISCUSSION

Identifying anomalies in student grades is only the first step; it requires a thorough investigation into the underlying causes that contribute to such irregularities. These anomalies could arise from various personal factors, including a student's motivation level, their study habits, and any external influences, such as family circumstances or mental health issues, that may affect their academic performance. It is essential to acknowledge that these factors can greatly vary from one student to another, making it necessary to adopt a personalized approach to understand each case. Additionally, the presence of anomalies could be related to variations in assessment methods used by instructors, or even errors in the data collection process, which can impact the accuracy of the anomalies detected. Therefore, it is crucial to scrutinize both the students' circumstances and the assessment mechanisms to fully understand the root causes of these outliers.

This research emphasizes that a comprehensive understanding of the anomalies in academic performance can significantly aid in providing targeted interventions for students who may need additional support. By identifying specific issues affecting students' grades, educators and administrators can tailor resources and assistance to address their needs. This may involve improving teaching strategies, offering academic counseling, or providing additional learning materials. Understanding the sources of irregularities not only helps individual students improve but also contributes to enhancing the overall educational experience, ensuring that grading is fair and accurately reflects each student's abilities and efforts.

Moreover, this study illustrates the effectiveness of data mining techniques, such as the Isolation Forest algorithm, within an educational context. These advanced analytical tools allow for the automatic detection of anomalies and can be particularly useful in large datasets where manual examination would be time-consuming and less efficient. The ability to quickly identify students whose performance deviates from the expected pattern enables educators to act promptly, ensuring timely interventions that can significantly impact students' academic trajectories.

The application of such data mining techniques enhances the quality and fairness of assessments. Traditional grading systems may overlook subtle patterns or inconsistencies in student performance,

ISSN: XXX-XXX-XXXXX-X-X
Desember 2024
Hal: 434 - 438

but algorithms like Isolation Forest are capable of uncovering complex relationships and irregularities in data. By implementing these techniques, educational institutions can adopt a more data-driven approach to grading and evaluation, ultimately fostering a learning environment that is more responsive to students' needs.

Furthermore, the use of anomaly detection not only benefits the students but also provides educators with valuable insights into their teaching methods. If anomalies are found to be consistently related to certain aspects of the course structure, such as specific assignments or exams, it may indicate areas where the instructional design needs to be revisited or improved. This feedback loop allows educators to continuously refine their assessment strategies, ensuring that they effectively measure student learning outcomes and provide equitable opportunities for all students to succeed.

In conclusion, the integration of machine learning techniques like the Isolation Forest algorithm in educational settings holds significant promise for improving the quality of assessments and the monitoring of academic performance. By systematically identifying and analyzing anomalies, educators can enhance the reliability of grading processes, support students more effectively, and ultimately contribute to better educational outcomes. The use of such techniques demonstrates how data-driven strategies can play a pivotal role in both improving student learning and advancing educational methodologies.

CONCLUSION

This study effectively identifies anomalies in the final grades of students enrolled in a Computer Networks course through the application of the Isolation Forest algorithm. The ability to detect these anomalies opens the door for more in-depth analysis of various factors that influence student performance. It highlights the potential for educators to implement targeted interventions aimed at supporting students who may be struggling or who demonstrate irregularities in their academic records. By isolating outlier grades, the study not only sheds light on potential issues within the student population but also allows for the refinement of the grading process, ensuring a fair and balanced assessment approach.

The research emphasizes the advantages of utilizing data mining techniques in the educational field, particularly for improving the accuracy and quality of academic assessments. The use of the Isolation Forest algorithm demonstrates how advanced analytical tools can be effectively employed to monitor student outcomes, identify patterns, and ensure that grading practices are consistent and reflective of true student abilities. Additionally, the study suggests the need for future research to delve deeper into the specific causes behind detected anomalies, as understanding these underlying factors is crucial for developing comprehensive strategies to address them. This could involve examining the impact of personal issues, teaching methodologies, and assessment structures on student performance, thereby contributing to a more equitable and supportive educational environment.

REFERENCES

Anandakrishnan, A., Kumar, S., & Statnikov, A. (2018). Anomaly detection in finance: Editors' introduction. KDD Workshop on Anomaly Detection in Finance, 1-7. Retrieved from Google Scholar.

ISSN: XXX-XXX-XXXXX-X-X
Desember 2024
Hal: 434 - 438

Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 93-104. Retrieved from Google Scholar.

Campos, G. O., et al. (2016). On the evaluation of unsupervised outlier detection: Measures, datasets, and an empirical study. Data Mining and Knowledge Discovery, 30(4), 891–927. https://doi.org/10.1007/s10618-015-0444-8.

Feng, Y., et al. (2019). Anti-money laundering (AML) research: A system for identification and multi-classification. In WISA 2019, Lecture Notes in Computer Science, 11817, 169–175. https://doi.org/10.1007/978-3-030-30952-7_19.

Hariri, S., Kind, M. C., & Brunner, R. J. (2019). Extended isolation forest. IEEE Transactions on Knowledge and Data Engineering, 33(4), 1479–1489. https://doi.org/10.1109/TKDE.2019.2947676.

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, 33(2), 494–514. https://doi.org/10.1109/TNNLS.2020.3015724.

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets, and challenges. Cybersecurity, 2(1), 1–22. https://doi.org/10.1186/s42400-019-0038-7.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In 2008 Eighth IEEE International Conference on Data Mining, 413–422. IEEE. Retrieved from Google Scholar.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2010). On detecting clustered anomalies using SCiForest. In ECML PKDD 2010, Lecture Notes in Artificial Intelligence, 6322, 274–290. https://doi.org/10.1007/978-3-642-15883-4_18.

Marx, M., Krötzsch, M., & Thost, V. (2017). Logic on mars: Ontologies for generalized property graphs. In IJCAI 2017, 1188–1194. Retrieved from Google Scholar.

Ruff, L., Kauffmann, J. R., et al. (2021). A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE, 109(5), 756–795. Retrieved from Google Scholar.

Xu, J., & Zhou, J. (2022). A Fine-Grained Anomaly Detection Method Fusing Isolation Forest and Knowledge Graph Reasoning. In Web Information Systems and Applications, Lecture Notes in Computer Science, 13579, 126-141. Springer, Cham. https://doi.org/10.1007/978-3-031-20309-1_12.

Xu, H., Pang, G., Wang, Y., & Wang, Y. (2022). Deep Isolation Forest for Anomaly Detection. Retrieved from arXiv: https://doi.org/10.48550/arXiv.2206.06602.

Zhang, Y., Zha, D., You, S., Wang, S., Hu, X., & Guo, M. (2023). OptIForest: Optimal Isolation Forest for Anomaly Detection. IJCAI, 2380-2387. Retrieved from https://www.ijcai.org/proceedings/2023/326.

Zimek, A., Schubert, E., & Kriegel, H. P. (2013). A survey on unsupervised outlier detection in high-dimensional numerical data. Statistical Analysis and Data Mining, 5(5), 363–387. https://doi.org/10.1002/sam.11280.