JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 295-303

IMPLEMENTATION OF THE LOCAL OUTLIER FACTOR MODEL FOR ANOMALY DETECTION IN KPU SUKABUMI REGENCY VOTER DATA

Nugraha[1)], Gina Purnama Insany[2)], Roihan Kusuma Wardana*[3)]
1. Informatics Engineering, Faculty of Engineering, Computer Science and Design, Nusa Putra University, Indonesia
2. Informatics Engineering, Faculty of Engineering, Computer Science and Design, Nusa Putra University, Indonesia
3. Informatics Engineering, Faculty of Engineering, Computer Science and Design, Nusa Putra University, Indonesia

Article Info
Keywords: Anomaly Detection; Artificial Intelligence; Local Outlier Factor (LOF); Unsupervised Learning; Voter Data

Article history:
Received 15 October 2024
Revised 14 November 2024
Accepted 9 December 2025
Available online 1 March 2025

DOI :
https://doi.org/10.29100/jipi.v10i1.5795

* Corresponding author.
Corresponding Author
E-mail address:
roihan.kusuma_ti20@nusaputra.ac.id

ABSTRACT
General elections are one of the most significant political activities in the life of a nation, necessitating accurate and reliable voter data. Inaccurate or unreliable voter data can lead to various issues, such as electoral fraud. One primary cause of inaccurate voter data is the presence of anomalies, which are data points that do not match the actual conditions. Anomalies in voter data can arise from several factors, including data entry errors, fraud, or system faults. To detect anomalies in voter data, various methods employing Artificial Intelligence (AI) can be utilized, with the Local Outlier Factor (LOF) method being one notable example. LOF is an unsupervised learning method in machine learning that identifies anomalies by measuring the distance between data points and their nearest neighbors. This study aims to implement the LOF method to detect anomalies in the voter data of the Sukabumi Regency Election Commission. The voter data used in this research was obtained from the Sukabumi Regency Election Commission for the year 2024.

.

I. INTRODUCTION

General elections are a cornerstone of a nation's democratic system, providing the foundation for representative governance. They involve active citizen participation in shaping policy directions and electing representatives to uphold public interests. In the digital age and information technology era, conducting elections not only demands transparent and fair procedures but also hinges on accurate and reliable voter data [1].

The success of an election heavily relies on the accuracy of the voter data, which is used to ensure that the voting rights of every citizen are properly registered [2]. However, threats of fraud, data input errors, and anomalies in the voter data can have a negative impact on the integrity and public trust in the election process [3].

In the analysis of voter data for elections, anomaly detection is crucial to ensuring the integrity and accuracy of the data. Various methods are available for this purpose; however, the Local Outlier Factor (LOF) was chosen due to its several advantages. LOF is an unsupervised learning method that identifies anomalies by measuring the local density of each data point and comparing it with the density of surrounding data points. The advantages of LOF include the detection of local anomalies that may not be visible on a global scale but are significant locally, adherence to the density of data around a specific point, and adaptability in adjusting parameters such as the number of nearest neighbors considered (k). In the context of voter data, LOF can maintain high accuracy under diverse data distribution conditions and handle data with high heterogeneity, making it the ideal choice to ensure that the voter data used is accurate and reliable.

Anomalies in data, deviations from established patterns, present challenges, particularly in maintaining public trust in electoral processes. Voter data integrity is vital, with the Sukabumi Regency Election Commission tasked with ensuring accuracy. To tackle this, the Local Outlier Factor (LOF) offers a robust density-based method for identifying anomalies, separating clean data for prediction while enhancing security and integrity [4].

The LOF is a machine learning algorithm used for anomaly detection in data. LOF falls under the category of unsupervised learning algorithms, designed to identify data points that are unusual or different from the majority of the data. Specifically, LOF measures the degree of deviation or oddness of a data point in the context of its

neighbors [5]. Connecting this research with previous studies enhances the argument for LOF's effectiveness. For instance, studies like Efrem Heri Budiarto's work on detecting anomalies in hospital drug usage data and Ahmad Zulfikar's research on anomaly detection using Isolation Forest in inventory management both demonstrate the practical utility and accuracy of machine learning methods in identifying outliers in various contexts. Such parallels reinforce the potential of LOF in maintaining voter data integrity by highlighting its successful applications in different anomaly detection scenarios.

The LOF method works by calculating the distance between each data point and its nearest neighbors. The local density of each data point is then computed by averaging these distances. LOF is determined for each data point by comparing its local density to that of its neighbors. Points with significantly higher LOF values compared to their neighbors are likely considered anomalies [6].

When applying the Local Outlier Factor (LOF) method to voter data, potential challenges include the large scale of data affecting computational performance, data variability complicating parameter determination, and data quality issues. To address these, data sampling techniques and stronger computational infrastructure can be employed, as well as cross-validation to find optimal parameters and exploratory analysis to understand data characteristics. Data quality issues can be managed through stringent data cleaning, including imputation of missing values and data normalization. Logistical challenges such as access to accurate data and limited human resources can be mitigated by collaborating with government agencies and providing training in data science and machine learning. Thus, the application of LOF in voter data anomaly detection can be effective and efficient, ensuring the integrity and accuracy of voter data in elections.

This study provides a conceptual and practical foundation for the Sukabumi Regency Election Commission (KPU) to address potential issues in managing voter data. Determining the voter list is a time-consuming and labor-intensive process in elections. Although seemingly procedural, each step in voter list determination involves various dynamics, engaging multiple stakeholders, and making the process more complex [7].

This research aims to improve local elections by investigating anomalies' causes and impacts on the election process and implementing the LOF method. LOF is utilized to identify emerging anomaly patterns or trends in voter data, enabling targeted preventive measures for data cleansing or correction. Regular LOF application in voter data management facilitates ongoing anomaly monitoring and detection, ensuring the integrity and accuracy of voter data. Focusing on voter data from the Sukabumi Regency Election Commission in 2024, this study provides insights into anomaly identification and LOF's application in addressing these issues. The findings of this research can concretely enhance the election process in Sukabumi Regency by ensuring more accurate and reliable voter data, thereby improving the overall electoral integrity and potentially serving as a model for other regions. This underscores the importance of supporting transparent, fair, and trustworthy elections in Sukabumi.

The study by Efrem Heri Budiarto explores anomaly detection, aiming to improve data quality by identifying anomalies in hospital drug usage data using methods like LOF, effectively detecting outliers [8]. In "Automatic Hyperparameter Tuning Method for Local Outlier Factor, with Applications to Anomaly Detection," Zekun Xu suggests a heuristic approach to adjust LOF parameters. LOF employs two hyperparameters: the neighborhood size (k), determining neighbors for density calculation, and contamination (c), specifying the proportion of points labeled as anomalies. In essence, k ranks training data, while c sets the anomaly cutoff [9].

Based on the research findings by Ahmad Zulfikar (2019) titled "Anomaly Detection Using Isolation Forest in Inventory Consumption Goods Purchasing at the Indonesian National Police Work Unit", it is explained that anomaly detection is crucial for enhancing efficiency and accountability in data management. By detecting anomalies, organizations can identify and address potential issues. Additionally, in the study by Handi Mulyaningsih (2020) titled "Validity of Potential Voter Data (DP4) in the 2019 Simultaneous Elections in Lampung (A Study in Pesawaran Regency)", the research results indicate that the voter data in the 2019 simultaneous elections in Lampung are invalid due to containing elements that do not meet the requirements. Moreover, there is a significant margin or difference between DP4 and the Permanent Voter List (DPT) [10][11].

The Election Commission (KPU) holds a pivotal role in Indonesia's electoral process, ensuring its adherence to principles of directness, transparency, fairness, and honesty as per its slogan. Operating independently, the KPU oversees all aspects of elections, including voter registration, campaign regulations, and vote counting, crucially monitoring the integrity of voter data to prevent electoral violations and disputes [12]. General elections (Pemilu) in Indonesia serve as the cornerstone of democratic governance, allowing the populace to exercise sovereignty in electing representatives. Evolving from colonial times, Indonesia's electoral system has embraced democratic principles since the Reform era, with legal frameworks, including the Constitution and laws like UU No. 7 of 2017, delineating electoral processes, dispute resolution mechanisms, and the roles of institutions such as the KPU,

Bawaslu, and DKPP, addressing issues like proportional representation, electoral phases, participants, and dispute resolution [13].

Anomaly in a dataset refers to deviations from expected or normal patterns caused by errors, unusual behaviors, or unidentified phenomena. These deviations can take various forms, such as values significantly different from the average, inconsistent patterns, or unexpected relationships between variables. Detecting anomalies in voter data during elections, such as duplicate entries or attempts at manipulation, is crucial for ensuring the integrity and validity of the electoral process, thus upholding public trust in the election outcomes [14]. Anomaly detection involves identifying data patterns differing from normal or common data patterns, aiming to spot deviations from normal or common activities, which can be errors or other unusual behaviors requiring further detection and analysis. The fundamental principle of anomaly detection involves comparing new data patterns with known normal data patterns, categorizing significant differences as anomalies or deviant behaviors. Methods such as clustering, classification, and statistics are commonly used for anomaly detection, with artificial intelligence-based methods often applied for non-normally distributed data or data with high dimensions. K-nearest neighbors (KNN), isolation forest (IF), and local outlier factor (LOF) are common AI-based methods used for anomaly detection, each with its unique approach and characteristics [15][16][17].

Artificial Intelligence (AI) involves computer systems designed to perform tasks that typically require human intelligence, such as reasoning, learning, perception, natural language understanding, and decision-making. It includes various techniques such as machine learning (ML), deep learning, and search algorithms. ML, a subset of AI, enables computers to learn from data without explicit programming, with approaches including supervised, unsupervised, and reinforcement learning. Unsupervised learning algorithms, like Local Outlier Factor (LOF), are given unlabeled data to find hidden patterns or structures, commonly used for clustering, dimensionality reduction, and anomaly detection. LOF detects outliers based on local density, assigning anomaly scores to data points by calculating k-nearest neighbor distances, local reachability density (LRD), and comparing each point's LRD to its neighbors. LOF's advantages include not requiring label information and scaling to different feature scales, but it also faces challenges like computational complexity and sensitivity to parameters, affecting its performance and interpretation [18].

II. RESEARCH METHODOLOGY

This study employs a qualitative approach, which aims to gain in-depth understanding of a phenomenon within a social context through a more naturalistic method.

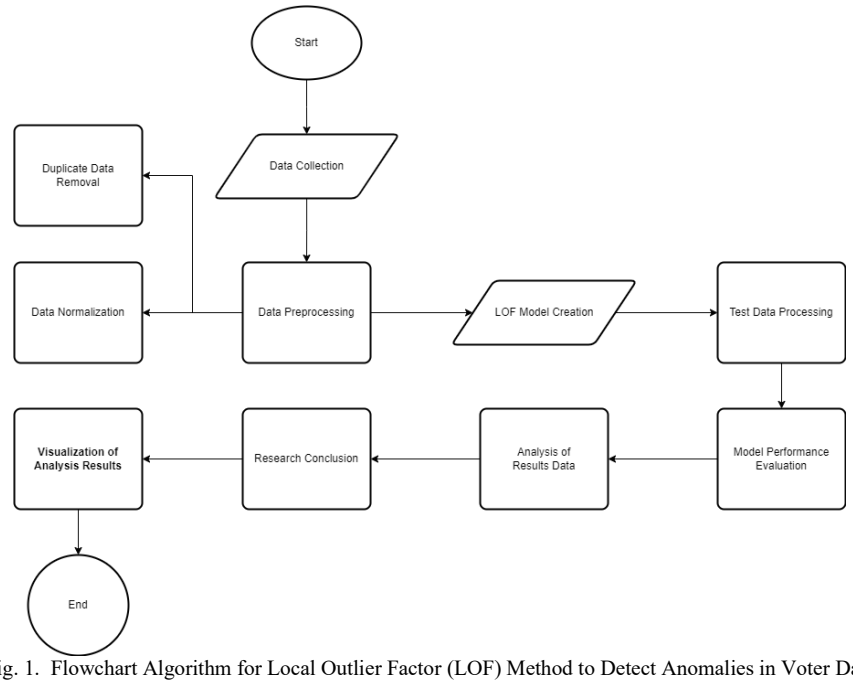

Fig. 1. Flowchart Algorithm for Local Outlier Factor (LOF) Method to Detect Anomalies in Voter Data

A. Data Collection

The research process comprises several stages, starting with data collection from the Sukabumi Regency Election Commission (KPU). Vital attributes such as name, residence, age, and gender were gathered intensively to ensure data credibility and completeness. The collection ensured representativeness across the targeted population, crucial for avoiding analysis bias and ensuring generalizable outcomes. Moreover, strict adherence to privacy regulations, especially when handling personal data provided by the KPU, was maintained.

Following data collection, verification was conducted to ensure accuracy and consistency. The verification process involved checks for duplication, missing values, and reasonable value ranges for each attribute. Errors or inconsistencies identified were then corrected before the data was processed further [19].

TABLE I
DATA STRUCTURE

NO	NAMA	JENIS KELAMIN	USIA	DESA/KELURAHAN	RT	RW	KET.
1	XXXXX	L/P	123	XXXXX	00	00	
2	XXXXX	L/P	123	XXXXX	00	00	
3	XXXXX	L/P	123	XXXXX	00	00	
4	XXXXX	L/P	123	XXXXX	00	00	
5	XXXXX	L/P	123	XXXXX	00	00	

The data obtained from the Sukabumi Regency Election Commission encompasses several key attributes, which are the primary focus of this analysis. These attributes include: Name, Residence, Age, and description. Each attribute plays a different role and contributes to the anomaly detection process using the Local Outlier Factor (LOF) model.

B. Data Preprocessing

Data preprocessing involves several stages, including data cleaning and categorical variable encoding. During this phase, efforts are made to remove incomplete or irrelevant data, ensuring a refined dataset suitable for subsequent analysis. Additionally, attributes initially in string format are transformed into numerical format using techniques such as label encoding [20]. The goal of data preprocessing is to ensure a refined dataset suitable for subsequent analysis.

a) The initial step in data preprocessing is data cleaning, which involves meticulous identification and rectification of errors such as spelling mistakes, missing values, and inconsistencies. Techniques such as mean or median imputation are employed to address missing values, chosen based on the data's characteristics. This step is crucial to ensure that the data meets the requirements for processing with machine learning algorithms, which typically demand numeric input. Verifying data accuracy and consistency is vital to enhance confidence in the validity of the data used for research purposes.
b) Following data cleaning is data normalization, aimed at standardizing numeric values in the dataset to a consistent scale. This ensures that all features contribute equally during analysis. Common normalization techniques include Min-Max Scaling, which confines data within the range [0, 1], and Z-score Standardization, which adjusts data to have a mean of 0 and a standard deviation of 1 [21]. The choice of normalization technique is guided by the data distribution and specific analytical needs.

The first step in data normalization involves selecting the appropriate technique. For example, Min-Max Scaling transforms data values to fit within the [0, 1] range, while Z-score Standardization adjusts data values to ensure a mean of 0 and standard deviation of 1. The selection depends on the specific characteristics of the dataset and the objectives of the analysis.

Once the normalization technique is chosen, it is applied uniformly across the dataset, ensuring consistency in data representation. This systematic approach results in a normalized dataset that is prepared for use in developing LOF models, enhancing the reliability and validity of the data used in research.

C. LOF Model Creation

In this phase, the Local Outlier Factor (LOF) model is developed to identify anomalies within the dataset. LOF operates by comparing the local density of each data point with that of its neighbors. Anomalies are identified as data points with significantly lower density compared to their neighbors. The key parameter in LOF is the number of neighbors (k), which plays a crucial role in determining the accuracy of anomaly detection.

The initial step in constructing the LOF model involves optimizing the parameter k. This parameter specifies the

number of neighbors used for calculating local density. Finding the optimal value of k is typically achieved through experimentation or cross-validation techniques. Experimentation involves testing different values of k and evaluating their impact on the model's performance. Cross-validation verifies the robustness of the chosen k value by splitting the dataset into training and validation sets, ensuring that the model generalizes well to unseen data.

Once the optimal value of k is determined, the LOF model proceeds to training using normalized datasets. The LOF algorithm computes the LOF value for each data point based on its local density and that of its neighbors. A higher LOF value indicates a greater likelihood of the data point being an anomaly. This methodical approach ensures that the LOF model is accurate and reliable in identifying anomalies within the dataset.

III. RESULT AND DISCUSSION

This section delves into the implementation of Local Outlier Factor (LOF) for anomaly prediction using the provided data. The discussion will be divided into two parts: testing results and analysis of testing results.

A. Data Testing Process

Once the LOF model is established, the next step is to test the model using test data. Test data is a subset of data not used during model training, utilized to measure the model's performance in detecting anomalies. This process involves applying the LOF model to the test data and evaluating the results.

The first step in the data testing process is preparing the test data using the same preprocessing steps as before. This includes cleaning the data by removing invalid values in the gender column (only 'M' for male and 'F' for female are included). Sklearn's LabelEncoder is used to convert categorical gender data into numeric format, enabling further processing by the LOF algorithm. Features used for anomaly detection are age and gender, which are then processed to ensure no empty values using the numpy nan_to_num function. The LOF model is then initialized with parameters: k_neighbors set to 20 and contamination level set to 0.1, to detect 10% of the most suspicious data as anomalies.

Following the application of the LOF model, the anomaly detection results are incorporated into the original dataframe by marking rows identified as anomalies. Data detected as anomalies are exported back to an Excel file for further analysis. Visualization of the results using matplotlib shows the differences between the original and anomaly data in a scatter plot. Anomalies are displayed in red, while original data is displayed in blue. This plot provides a clear overview of the distribution of anomalies based on age and gender, enabling deeper visual analysis of deviations in the dataset. This implementation demonstrates the effectiveness of the LOF model in identifying suspicious data and provides a foundation for further action in voter data management.

However, it is important to recognize and address the limitations present in this research. One notable limitation is the quality and comprehensiveness of the data used. The dataset may not fully represent the population, potentially leading to biased results. Additionally, the LOF model's performance can be influenced by the choice of parameters, such as the number of neighbors (k) and the contamination level. These parameters require careful tuning to avoid overfitting or underfitting the model.

Another limitation is the simplicity of the features used for anomaly detection, which are limited to age and gender. This may not capture the full complexity of the data, and incorporating additional relevant features could improve the model's accuracy. Furthermore, the model's ability to detect anomalies may be constrained by the inherent noise and inaccuracies in the data, which could affect the reliability of the results.

To mitigate these limitations in future research, more comprehensive datasets should be used, encompassing a wider range of features and a larger sample size to better represent the population. Advanced techniques such as feature engineering and more sophisticated anomaly detection models could also be explored. Additionally, incorporating cross-validation and other robust validation techniques can help ensure the model's generalizability and accuracy.

In conclusion, while the LOF model demonstrates effectiveness in identifying anomalies in voter data, the results are subject to limitations related to data quality, feature selection, and model parameters. Addressing these challenges in future work will be crucial for improving the robustness and reliability of anomaly detection in voter data management.

B. Model Accuracy Performance Evaluation

Assessing the accuracy performance of the model is crucial to ensure that the LOF model functions effectively in detecting anomalies. It involves measuring the proportion of true anomalies detected and determining the proportion of anomalies correctly identified by the model. The F1-score, a harmonic mean of precision and recall,

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 295-303

provides an overall insight into the model's performance.

The following formulas calculate the accuracy rate, precision, recall, and F1 score based on the provided code:

a) Accuracy

Accuracy measures the proportion of correct predictions (both positive and negative) out of the total predictions made.

$$Accuracy = \frac{TP+TN}{TP + TN + FP + FN} \quad (1)$$

b) Precision

Precision measures the proportion of true positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

c) Recall

Recall measures the proportion of positives that are correctly detected.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

d) F1 Score

F1 Score represents the harmonic mean of precision and recall.

$$F1\ Score = \frac{TP}{TP + FN} \quad (4)$$

In the context of anomaly detection, TP (True Positive) represents the count of correctly detected anomalies, while TN (True Negative) indicates the count of normal data correctly identified as such. FP (False Positive) refers to the count of normal data incorrectly flagged as anomalies, and FN (False Negative) denotes the count of anomalies mistakenly classified as normal data.

The accuracy of LOF detection varies significantly based on two datasets, reflecting differences in model performance. Below is an explanation regarding these findings.

```
Akurasi (berdasarkan Anomali): 0.9998061096099503
Precision (berdasarkan Anomali): 1.0
Recall (berdasarkan Anomali): 0.6356589147286822
F1 Score (berdasarkan Anomali): 0.7772511848341233
```

Fig. 2. Model Performance Evaluation Results

Based on the following results, it can be concluded that:

TABLE 2
MODEL PERFORMANCE EVALUATION RESULTS

Evaluation	Score	Percentage	Explanation
Accuracy	0.999806109609950	99.98%	Extremely high, indicating the model is almost always correct in its predictions.
Precision	1.0	100%	Indicates that everything detected as an anomaly is indeed an anomaly.
Recall	0.6356589147286822	63.57%	Only about 63.57% of the actual anomalies are detected.
F1 Score	0.7772511848341233	77.73%	A good balance between precision and recall.

The accuracy assessment of the LOF detection results indicates that the model exhibits very high accuracy and precision, with a precision value reaching 1.0, meaning every anomaly prediction made by the model is correct. However, the lower recall of 63.57% suggests that the model can only detect a small portion of all actual anomalies present in the data. Therefore, to enhance the model's performance, further efforts are needed to improve recall. These efforts might include parameter tuning, adding relevant features, or ensuring that the training data covers various anomaly patterns. Thus, although the current LOF model demonstrates high reliability in identifying anomalies, there is room for improvement to detect more existing anomalies.

In comparison to baseline or other methods, such as isolation forest, the results provide insightful contrasts. For instance, Ahmad Zulfikar, Farhan Ariq Rahmani, and Nurul Azizah's research on anomaly detection using isolation forest on inventory consumption data within the Indonesian police force showed that with an optimal contamination parameter of 0.3%, the model effectively identified anomalies [10]. This indicates that isolation forest can be highly efficient in certain contexts, highlighting the necessity to compare and possibly integrate different approaches for improved results.

Moreover, Zekun Xu et al. proposed a heuristic methodology for tuning LOF hyperparameters, achieving high F1 scores and AUC on small datasets. On larger datasets, the performance of the tuned LOF was comparable to the best results from one-class SVM on datasets like Http and Smtp, and it outperformed other methods on Credit and Mnist datasets [9]. This suggests that fine-tuning hyperparameters is crucial for optimizing LOF performance, especially in varied data environments.

Efrem Heri Budiarto et al. compared K-Means, LOF, and One Class SVM, finding that while OC-SVM showed better results on certain datasets with an anomaly threshold of 0.05, LOF excelled in performance on others [8]. This comparison underlines the importance of selecting the right algorithm and thresholds based on specific dataset characteristics.

Following the model performance evaluation, a thorough analysis is conducted to understand the results and identify areas for improvement. The evaluation metrics, particularly the accuracy, precision, recall, and F1-score, provide a comprehensive view of the model's performance. These metrics guide the refinement of the model, helping in the selection of better parameters to ensure optimal operation of the LOF model in detecting anomalies in voter data. This iterative process of evaluation and refinement aims to enhance the model's ability to identify all anomalies accurately, thereby improving the overall quality and reliability of the data used in subsequent analyses.

C. Data Analysis Results

The data analysis process involves interpreting and understanding the outcomes obtained from the LOF model. During this phase, the author examines the data points identified as anomalies and seeks patterns or characteristics that may explain why these data points are considered anomalies. This analysis can provide valuable insights into voter data and aid in identifying potential issues such as inaccurate data or duplicate voters.

The first step in the data analysis process is grouping anomalies based on their characteristics. For instance, anomalies can be categorized based on attributes like age, address, and others. This grouping helps in understanding if there are any specific patterns emerging among the anomalies.

Subsequently, a thorough analysis of each anomaly group is conducted to identify potential causes. For example, anomalies within a specific age group may indicate issues with data recording or input errors. The results of this analysis are used to provide recommendations for improvement and to enhance the quality of voter data in the future.

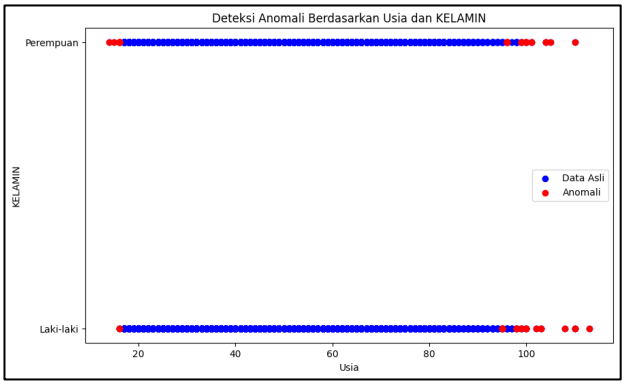

Fig. 3. Age & Gender Anomaly Results

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 1, Maret 2025, Pp. 295-303

The first displayed graph depicts anomaly detection categorized by age and gender. Blue points represent original data classified as normal, while red points indicate anomalies. In the male group, anomalies are scattered primarily among very young individuals (below 20 years old) and very elderly individuals (above 100 years old). Anomalies in younger ages may arise from erroneous data entries or exceedingly rare occurrences within the population. Similarly, anomalies in older ages suggest unusual data or potential input errors, given that ages above 100 are exceptionally uncommon in the general population and warrant closer examination. The female group exhibits a comparable pattern, with anomalies detected at very young and very old ages. This indicates a consistent trend across genders in detecting extreme age anomalies. Notably, this initial graph's anomaly detection offers a comprehensive view of how anomaly distribution varies between male and female groups, facilitating a detailed analysis of data integrity within each subgroup.

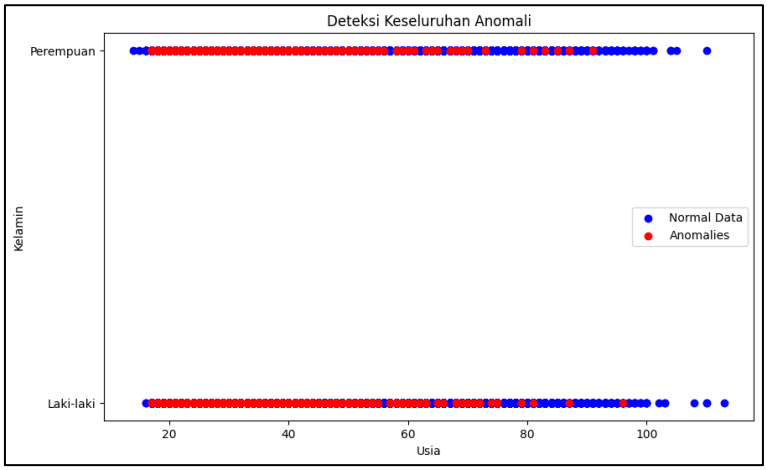

Fig. 4. Overall Anomaly Results

The second graph provides a broader perspective on anomaly detection without gender-based segregation. It illustrates anomalies marked with red dots scattered across various age ranges. This graph indicates anomalies detected primarily in data that should not be included in the final dataset used by the Election Commission. This data might be of people who have already moved domicile or passed away and should have been removed from the main dataset. Clearly, this demonstrates the most obvious anomaly in the group of data being studied, because these categories of data should not be in the dataset.

In the male subgroup, there is a wider spread of anomalies compared to the first graph, indicating more deviations from the typical data pattern. This suggests that anomaly detection encompasses a broader range of data points that may necessitate closer scrutiny to ensure accuracy. Similarly, the female subgroup in the second graph displays a distribution pattern of anomalies similar to that of the male subgroup, with anomalous points scattered across different age brackets. Anomalies observed at specific age ranges or potentially problematic. Such anomalies could stem from data input errors or negligence in data management.

IV. CONCLUSION

The results from both graphs provide an overview of anomaly distribution in the dataset based on age and gender. The first graph aids in identifying differences and similarities in anomaly detection between male and female groups, while the second graph offers a more comprehensive view of anomaly distribution in the data without separation based on labels in the existing data. The detection outcomes indicate that very young and elderly individuals tend to have more anomalies, suggesting potential issues with data entries or specific characteristics of the population within those age ranges. However, it should be noted that according to Indonesian law Article 38 paragraph (4) of Law Number 7 of 2017 concerning General Elections, individuals who are married and have children are allowed to vote in elections even if they are under 17 years old [22].

Nevertheless, it's crucial to understand that this article has several conditions that must be met:

a) The individual must be legally married (registered at the Office of Religious Affairs or civil registry).
b) The individual must have children born from a legal marriage.
c) Marital status and the presence of children must be proven with valid documents.

The LOF method proves effective in identifying anomalies in the dataset based on age and gender. This method

can detect outliers not only at the tail end of the data distribution but also at various points within the age range. Thus, the LOF method aids in identifying data that may require further review or special attention to ensure data quality. This analysis is crucial to ensure that the data used in further research is accurate and reliable.

REFERENCES

[1] A. Habibi, "Upaya Menyelamatkan Pemilihan Umum Di Tahun 2020," vol. 4, pp. 167–172, 2020.
[2] P. Finance, *Funding of Political Parties and Election Campaigns*.
[3] P. Gleko, A. Suprojo, A. W. Lestari, U. Tribhuwana, and T. Malang, "Strategi komisi pemilihan umum dalam upaya meningkatkan partisipasi politik masyarakat pada pemilihan umum kepala daerah," vol. 6, no. 1, pp. 38–47, 2017.
[4] E. H. BUDIARTO, "Pendeteksian Anomali Menggunakan Local Outlier Factor Pada Data Untuk Meningkatkan Performa Prediksi Jumlah Obat," pp. 2–3, 2020, [Online]. Available: https://etd.repository.ugm.ac.id/penelitian/detail/183405
[5] M. M. Breunig, H. Kriegel, R. T. Ng, and J. Sander, "LOF : Identifying Density-Based Local Outliers," pp. 93–104, 2000, doi: 10.1145/342009.335388.
[6] D. Version, "Outlier Selection and One-Class Classi cation," 2024.
[7] M. H. Prof. Dr. H. Nandang Alamsah Deliarnoor, S.H., M. S. Dr. Hj. Ratnia Solihah, S.IP., M. Mustabsyirotul Ummah Mustofa, S.IP., and M. K. Tripanji Aryawardhana, S.H., "Riset Daftar Pemilih Provinsi Jawa Barat," pp. 1–98, 2019.
[8] E. H. Budiarto, A. Erna Permanasari, and S. Fauziati, "Unsupervised anomaly detection using K-Means, local outlier factor and one class SVM," *Proc. - 2019 5th Int. Conf. Sci. Technol. ICST 2019*, 2019, doi: 10.1109/ICST47872.2019.9166366.
[9] Z. Xu, D. Kakde, and A. Chaudhuri, "Automatic Hyperparameter Tuning Method for Local Outlier Factor, with Applications to Anomaly Detection," *Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019*, no. February, pp. 4201–4207, 2019, doi: 10.1109/BigData47090.2019.9006151.
[10] A. Zulfikar, F. A. Rahmani, N. Azizah, D. J. Perbendaharaan, K. Keuangan, and P. Pinang, "Deteksi Anomali Menggunakan Isolation Forest Belanja Barang Persediaan Konsumsi Pada Satuan Kerja Kepolisian Republik Indonesia," *J. Manaj. Perbendaharaan*, vol. 4, no. 1, pp. 1–15, 2023, doi: 10.33105/jmp.v4i1.435.
[11] H. Mulyaningsih, H. Hertanto, and D. Wibisono, "VALIDITAS DATA PEMILIH POTENSIAL PEMILU (DP4) PADA PEMILU SERENTAK 2019 DI LAMPUNG (Studi Di Kabupaten Pesawaran)," *Sosiol. J. Ilm. Kaji. Ilmu Sos. dan Budaya*, vol. 22, no. 1, pp. 64–78, 2020, doi: 10.23960/sosiologi.v22i1.48.
[12] D. A. Nugroho and R. M. Sukmariningsih, "Peranan Komisi Pemilihan Umum Dalam Mewujudkan Pemilu Yang Demokratis," *J. JURISTIC*, vol. 1, no. 01, p. 22, 2020, doi: 10.35973/jrs.v1i01.1449.
[13] N. K. Arniti, "Partisipasi Politik Masyarakat Dalam Pemilihan Umum Legislatif Di Kota Denpasar," *J. Ilm. Din. Sos.*, vol. 4, no. 2, p. 329, 2020, doi: 10.38043/jids.v4i2.2496.
[14] S. Thudumu, P. Branch, J. Jin, and J. (Jack) Singh, "A comprehensive survey of anomaly detection techniques for high dimensional big data," *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00320-x.
[15] S. Situmorang, "Analisis Kinerja Algoritma Machine Learning Dalam Deteksi Anomali Jaringan (LAZY LEARNING," *J. Mat. dan Ilmu Pengetah. Alam*, vol. 1, no. 4, pp. 259–269, 2023, [Online]. Available: https://doi.org/10.59581/konstanta.v1i4.1722
[16] J. Auskalnis, N. Paulauskas, A. Baskys, C. Technologies, and V. G. Technical, "Application of Local Outlier Factor Algorithm to Detect Anomalies in Computer Network," pp. 96–99, 2018.
[17] S. Sugidamayatno and D. Lelono, "Outlier Detection Credit Card Transactions Using Local Outlier Factor Algorithm (LOF)," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 4, p. 409, 2019, doi: 10.22146/ijccs.46561.
[18] A. Desmet and M. Delore, "Leak detection in compressed air systems using unsupervised anomaly detection techniques," *Proc. Annu. Conf. Progn. Heal. Manag. Soc. PHM*, pp. 211–220, 2017.
[19] M. A. Kusnaldi, N. F. Syani, and Y. Afifah, "Perlindungan Data Pribadi dalam Penyelenggaraan Pemilu : Tantangan dan Tawaran," vol. 7, no. 4, pp. 710–725, 2024.
[20] R. Sistem, F. Teknik, and U. P. Bangsa, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan," vol. 1, no. 10, pp. 504–510, 2021.
[21] C. N. Nasution and Y. Widyaningsih, "Klasifikasi Pemilih dalam Pemilu 2019 di Indonesia Menggunakan Regresi Logistik Multinomial dan Chi-Square Automatic Decision Tree (CHAID)," vol. 6, no. 2, 2022.
[22] S. Waworuntu, "Tinjauan Yuridis Mengenai Hak Pilih Masyarakat dalam Pemilihan Umum di Indonesia yang Belum 17 Tahun Tetapi Sudah Menikah," *Lex Adm.*, vol. 10, no. 5, 2022.