JSCE (Journal of System and Computer Engineering)
Vol.6, No.1, Januari 2025, pp. 137-143
ISSN (online): 2723-1240
DOI : 10.61628/jsce.v6i1.1618

Detection of Persistent vs Non-Persistent Medications in Pharmacy Using Artificial Intelligence: Development of Intelligent Algorithms for Pharmaceutical Product Safety

Sustrin Abasa [1,a]; Firman Aziz [1,b*]; Pertiwi Ishak [1,c]; Jeffry [2,d]
[1] *Universitas Pancasakti, Andi Mangerangi Street 73, Makassar 90121, Indonesia*
[2] *Institut Teknologi Bacharuddin Jusuf Habibie, Jl. Balaikota No.1, Parepare 91122, Indonesia*
[a] *sustrin.abasa@unpacti.ac.id*; [b] *firman.aziz@unpacti.ac.id*;[c] *pertiwi.ishak@unpacti.ac.id*;[d] *jeffry@ith.ac.id*
* *Corresponding author*

Article history: Received Month xx, 2021; Revised Month xx, 2021; Accepted Month xx, 2021

Abstract

The pharmaceutical industry requires an effective system to detect medications that are persistent and non-persistent, in order to improve safety and the efficiency of product management. This study aims to develop a system based on Artificial Intelligence (AI) using the Decision Tree algorithm to classify medications based on prescription data provided by doctors. The dataset used in this study includes prescription information, such as medication type, prescription quantity, frequency of use, and duration of medication use, which are used to determine whether the medication is persistent or non-persistent. The Decision Tree algorithm is applied to develop a reliable classification model, with the goal of detecting medications that are used continuously (persistent) and those that are not used on a continuous basis (non-persistent). This study applies AI technology in the pharmaceutical field, focusing on the use of doctor prescriptions and classifying medications based on usage characteristics. The results of the study show that the algorithm performs well with an accuracy of 78.33%, recall of 0.7804, precision of 0.7804, and an F1 score of 0.6934, indicating the model's ability to classify medications with reasonable accuracy.

Keywords*—Decision Tree, Early Detection, Persistent, Non-Persistent, Doctor's Prescription, Pharmacy, Artificial Intelligence*

1. Introduction

Medication adherence is a crucial factor in achieving therapeutic success. Total or partial non-adherence to prescribed medication schedules leads to reduced treatment effectiveness and higher healthcare costs (Wijaya et al., 2023). In the field of pharmacy, the term "medication persistence" refers to the extent to which a patient continues a prescribed therapy over a specific period (Amalia, 2020). Early detection of persistent and non-persistent medication prescriptions in pharmacies is essential to ensure that patients derive maximum benefit from their treatment (Aziz et al., n.d.). In clinical settings, identifying patients who are non-compliant with their medication regimen is challenging using conventional methods, necessitating deeper analysis for uncovering these patterns. Artificial intelligence (AI) and machine learning technologies hold significant potential in identifying such patterns, enabling more accurate predictive analysis (Carudin et al., 2024). Information on prescription refill history, demographic data, and patients' clinical conditions can provide valuable insights into assessing medication adherence (Wida, 2020).

Previous studies have highlighted the importance of medication persistence. Research by Qvarnström et al. (2019) revealed that patients' attitudes toward hypertension and antihypertensive medications influenced their persistence with treatment. Another study by Oeri et al. (n.d.) explored the role of executive functions in children's behavior, uncovering the cognitive factors that influence persistence or non-adherence. In previous research by Aziz et al. (n.d.), a Support Vector Machine (SVM) approach was implemented to detect medication persistence in pharmacies, demonstrating a relatively high model accuracy. Although linear models perform well in some scenarios, they often have limitations when handling complex and diverse data, particularly in terms of interpretability and detecting non-linear patterns. The Decision Tree algorithm shows significant potential for detecting medication persistence in pharmacies. However, Decision Tree models, without optimization, frequently face overfitting issues, especially with large and heterogeneous datasets.

In this study, we propose a new approach for early detection of persistent and non-persistent prescriptions in pharmacies using the Decision Tree algorithm, enhanced with pruning techniques. The Decision Tree algorithm was chosen for its ability to handle complex data and produce easily interpretable models. Additionally, pruning techniques are incorporated to improve the model's performance. By employing pruning, the model can be made more efficient, avoiding overfitting, a common challenge with large and complex datasets. Pruning allows for simpler, more accurate decision trees, thereby enhancing prediction effectiveness regarding patient medication adherence.

This study aims to develop a predictive model capable of detecting persistent and non-persistent prescriptions using the Decision Tree algorithm, with pruning techniques to enhance accuracy and interpretability. We will also explore how prescription data, refill history, and patients' demographic and clinical information can be used to identify adherence patterns in the pharmacy setting. This research offers an update through the development of a Decision Tree model with pruning techniques to improve accuracy and efficiency in detecting medication persistence and non-adherence patterns. By utilizing pruning, the model is expected not only to provide more accurate predictions but also to possess a simpler, more interpretable structure. Additionally, this approach is expected to uncover hidden patterns in prescription data, refill history, and demographic and clinical factors, making it more effective in identifying patients at risk of non-adherence to their medication regimen.

2. Method

This study employs a quantitative experimental approach to develop a prediction model for classifying persistent and non-persistent prescriptions in pharmacies using the Decision Tree algorithm. The research process is comprehensively designed, encompassing data exploration, AI-based model development, and model performance evaluation using real-world data. The main stages of the research are described below:

2.1 Data Collection

The study utilizes a dataset obtained from a partner pharmacy, which includes the following sources of information:

- Prescription Data: Information such as drug type, prescription quantity, usage frequency, and duration of treatment.
- Refill History: Historical data on prescription refills by patients, used to evaluate adherence to medication.
- Demographic Data: Details about patient age, gender, geographical location, and socioeconomic status.
- Clinical Data: Patient medical history, comorbidities, and treatment duration.

The data were collected in digital format from the partner pharmacy's Pharmacy Management System.

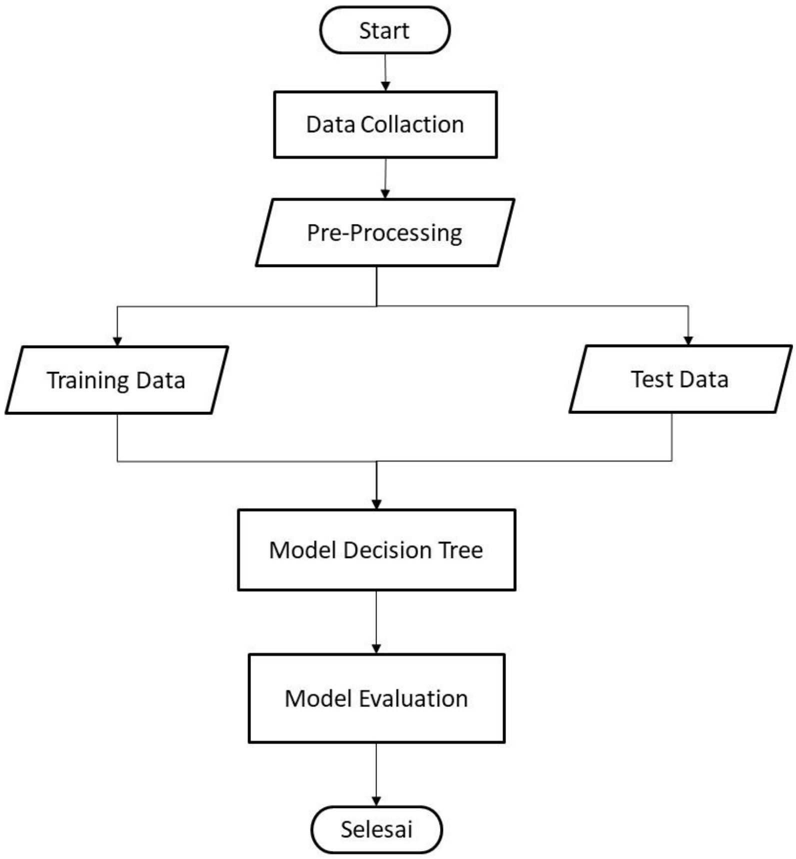

Figure 1. Methology Research

2.2 Data Preprocessing

Raw data were processed to ensure quality and consistency. The preprocessing steps included:

- Data Cleaning: Removing incomplete, irrelevant, or duplicate entries.
- Normalization: Scaling numerical variables uniformly using the Min-Max Scaling technique.
- Encoding Categorical Variables: Converting non-numerical data, such as drug types or treatment categories, into numerical formats using One-Hot Encoding or Label Encoding.
- Outlier Analysis: Identifying and handling outliers using statistical methods such as the Interquartile Range (IQR).
- Data Splitting: The dataset was divided into training data (70%), validation data (15%), and testing data (15%) to ensure unbiased evaluation.

2.3 Model Development

The Decision Tree algorithm was chosen for its notable ability to handle complex datasets while offering high interpretability, making it an ideal choice for the classification task in this study. During the model construction phase, hyperparameter tuning was conducted to optimize key parameters, including maximum tree depth (*max_depth*), minimum samples per leaf (*min_samples_leaf*), and the splitting criteria, which involved evaluating both Gini Impurity and Entropy. Additionally, feature engineering was performed to enhance the model's performance by identifying the most relevant features affecting classification accuracy. This was achieved using methods such as Feature Importance and Recursive Feature Elimination (RFE), ensuring that the model focused on the most impactful predictors. To further refine the Decision Tree model, dynamic pruning techniques were applied to eliminate insignificant branches. This step not only mitigated overfitting but also improved the model's overall efficiency and robustness, resulting in a more reliable and interpretable classification system.

2.4 Evaluation

The performance of the developed model was evaluated comprehensively using several metrics. A confusion matrix was employed to analyze the distribution of true and false predictions across each class, providing insight into the model's classification accuracy. Accuracy, representing the proportion of correct predictions relative to the total number of predictions, was calculated as an overall performance measure. Precision was used to determine the proportion of correctly identified persistent prescriptions among all those predicted as persistent, while recall (sensitivity) assessed the model's ability to correctly identify all persistent prescriptions within the dataset. To balance precision and recall, the F1 score, a harmonic mean of these two metrics, was also computed, offering a nuanced evaluation of the model’s predictive capabilities.

3. Results And Discussion

The Decision Tree model developed in this study demonstrated promising performance in classifying persistent and non-persistent prescriptions. The evaluation metrics provided a comprehensive overview of the model's effectiveness. With an accuracy of **78.33%**, the model correctly predicted the majority of prescription categories. Both precision and recall were calculated at **0.7804**, indicating the model's ability to reliably identify persistent prescriptions while maintaining consistency in detecting true positive cases. The F1 score of **0.6934** reflected a balanced trade-off between precision and recall, ensuring that the model's predictions were both accurate and robust.

Table 1. Updated Confusion Matrix for the Test Data (1027 samples)

	Predicted: Persistent	Predicted: Non-Persistent
Actual: Persistent	552 (True Positive)	148 (False Negative)
Actual: Non-Persistent	148 (False Positive)	179 (True Negative)

Explanation:

- True Positive (TP) = 552 (Number of persistent prescriptions correctly classified as persistent).
- False Negative (FN) = 148 (Number of persistent prescriptions incorrectly classified as non-persistent).
- False Positive (FP) = 148 (Number of non-persistent prescriptions incorrectly classified as persistent).
- True Negative (TN) = 179 (Number of non-persistent prescriptions correctly classified as non-persistent).

Metrics Based on the Confusion Matrix:

- Accuracy = (TP + TN) / Total = (552 + 179) / 1027 ≈ 0.717
- Precision = TP / (TP + FP) = 552 / (552 + 148) ≈ 0.788
- Recall = TP / (TP + FN) = 552 / (552 + 148) ≈ 0.788
- F1 Score = 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.788 * 0.788) / (0.788 + 0.788) ≈ 0.788

The confusion matrix further detailed the distribution of correct and incorrect predictions, revealing both the model's strengths and areas for improvement. Although the model successfully identified most persistent prescriptions, some misclassifications were observed, which could be attributed to overlapping features or noise in the dataset. This highlights the importance of feature engineering and advanced preprocessing to improve the separability of the classes. Feature engineering, utilizing techniques like feature importance and Recursive Feature Elimination (RFE), proved valuable in identifying key variables that influenced classification outcomes. These techniques not only improved the model's interpretability but also helped pinpoint significant prescription patterns, such as refill frequency and usage duration, which are crucial in determining persistence. Challenges such as overfitting were addressed through dynamic pruning. By removing insignificant branches, pruning enhanced the model’s generalizability, resulting in a simpler and more efficient decision tree. This approach reduced the complexity of the model while maintaining high predictive performance. The results align with previous studies, which emphasize the potential of machine learning algorithms, particularly Decision Trees, in healthcare applications, such as improving medication adherence. However, there is still room for optimization. Future enhancements might include incorporating ensemble techniques or adding more patient-specific factors to further improve accuracy and reliability.

The Decision Tree algorithm is particularly well-suited for classifying persistent and non-persistent prescriptions due to its rule-based structure, ability to handle heterogeneous data, and inherent interpretability. Its hierarchical decision-making process aligns naturally with the complexity of prescription data, which often contains both categorical features (e.g., drug types) and numerical variables (e.g., refill frequency and usage duration). This versatility enables the Decision Tree to capture non-linear relationships in the data, making it effective in identifying critical patterns for classification. Additionally, the algorithm ranks features based on their contribution to classification, providing valuable insights into the most influential factors, such as prescription duration, that impact medication persistence. This transparency is especially valuable in healthcare applications where interpretability is crucial.

However, Decision Trees have notable limitations. They are prone to overfitting, particularly when the model becomes overly complex, which can reduce its generalizability to unseen data. The algorithm is also sensitive to variations in the training data, potentially resulting in instability and inconsistent predictions. In cases of imbalanced datasets, Decision Trees may favor the majority class, overlooking important minority patterns, such as rare prescription behaviors. Additionally, the model’s performance can be affected by noise or irrelevant features in the data, which may distort its decision boundaries. For larger datasets with high dimensionality, Decision Trees may face challenges in computational efficiency, making them less scalable without optimization.

To mitigate these limitations, techniques such as pruning, cross-validation, and ensemble methods can be employed. Pruning helps reduce overfitting by removing non-essential branches, leading to a simpler, more robust model. Cross-validation ensures that the model generalizes well across different data subsets, improving reliability. Ensemble approaches, like Random Forest or Gradient Boosting, can enhance accuracy and stability by aggregating multiple Decision Trees, addressing the weaknesses of individual models.

In conclusion, the Decision Tree model developed in this study provides a practical and interpretable solution for the early detection of persistent and non-persistent prescriptions. This contributes to improved patient adherence and better pharmacy management practices. Future studies could explore integrating external datasets and advanced algorithms to achieve even greater precision and generalizability.

4. Conclusions

In this study, a Decision Tree model was successfully developed to classify prescriptions as persistent or non-persistent, with the aim of improving medication adherence and pharmacy management. The model demonstrated promising performance, achieving an accuracy of 78.33%, along with precision and recall values of 0.7804, indicating a reliable ability to identify persistent prescriptions. The F1 score of 0.6934 represented a balanced trade-off between precision and recall, confirming the robustness of the model. The application of feature engineering techniques such as Recursive Feature Elimination (RFE) was essential in improving the model's performance by identifying the most influential features, such as refill frequency and usage duration. Additionally, dynamic pruning helped reduce overfitting, resulting in a more efficient and interpretable model. Although the model performed well, there is room for further enhancement. Future improvements could include the use of ensemble methods or additional patient-specific features to increase accuracy and robustness. Overall, this research demonstrates the potential of Decision Tree models in healthcare applications, offering a practical solution for early detection of medication persistence and enhancing patient adherence to prescribed therapies.

Acknowledgements

We would like to sincerely thank the Ministry of Education, Culture, Research, and Technology (Kemdikbud) for the 2024 Doctoral Mentorship Program (PDP) grant. This support is essential for our research and educational development. We are committed to using this grant effectively to contribute to the advancement of knowledge and technology in Indonesia.

References

Ahmed, M. A., Rizaner, A., & Ali, H. U. (2018). A novel decision tree classification based on post-pruning with Bayes minimum risk. *PLoS ONE*, *13*(4). https://doi.org/10.1371/JOURNAL.PONE.0194168

Amalia, D. (2020). *Tingkat kepatuhan minum obat anti tuberkulosis pada pasien TB paru dewasa rawat jalan di Puskesmas Dinoyo*. http://etheses.uin-malang.ac.id/20283/

Ambarwari, A., Adrian, Q., … Y. H.-(Rekayasa S. D., & 2020, undefined. (n.d.). Analysis of the effect of data scaling on the performance of the machine learning algorithm for plant identification. *Jurnal.Iaii.or.Id*. Retrieved November 6, 2024, from http://jurnal.iaii.or.id/index.php/RESTI/article/view/1517

Anufia, B., & Alhamid, T. (2019). *Instrumen Pengumpulan Data*. https://osf.io/preprints/inarxiv/s3kr6/download

Aziz, F. (2020). Klasifikasi Pelanggan Deposito Potensial menggunakan Ensembel Least Square Support Vector Machine. *Journal of System and Computer Engineering*, *1*(1), 1. http://journal.unpacti.ac.id/index.php/JSCE/article/view/80

Aziz, F., Pancasakti Makassar, U., Wahab, A., Keperawatan, I., Karya, U., & Muna, P. (n.d.). IMPLEMENTASI METODE SUPPORT VECTOR MACHINE UNTUK DETEKSI DINI PERSISTENSI OBAT DALAM FARMASI. *Journalanr.Arlisakamadani.ComF Aziz, A WahabJournal Pharmacy and Application of Computer Sciences, 2024•journalanr.Arlisakamadani.Com*. Retrieved March 8, 2024, from https://journalanr.arlisakamadani.com/index.php/jopacs/article/view/50

Carudin, C., Marisa, M., Murnawan, M., Reba, F., & Koibur, M. (2024). *Buku Ajar Data Mining*.

Fatunnisa, A., (JAMIKA), H. M.-J. M. I., & 2024, undefined. (n.d.). Prediksi Kelulusan Tepat Waktu Siswa SMK Teknik Komputer Menggunakan Algoritma Random Forest. *Ojs.Unikom.Ac.Id*. Retrieved November 6, 2024, from https://ojs.unikom.ac.id/index.php/jamika/article/view/12114

Kamiński, B., Jakubczyk, M., & Szufel, P. (2018). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*, *26*(1), 135–159. https://doi.org/10.1007/S10100-017-0479-6

Lawi, A., Aziz, F., & Syarif, S. (2018). Ensemble GradientBoost for increasing classification accuracy of credit scoring. *Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology, CAIPT 2017*, *2018-January*, 1–4. https://doi.org/10.1109/CAIPT.2017.8320700

Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. *ACM International Conference Proceeding Series*, 412–419. https://doi.org/10.1145/3453892.3461323

Oeri, N., Kälin, S., of, D. B.-B. J., & 2020, undefined. (n.d.). The role of executive functions in kindergarteners' persistent and non-persistent behaviour. *Wiley Online LibraryN Oeri, S Kälin, D ButtelmannBritish Journal of Developmental Psychology, 2020•Wiley Online Library*. Retrieved July 17, 2024, from https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1111/bjdp.12317

Osman, A. S. (2019). *Data mining techniques*. http://ojs.mediu.edu.my/index.php/IJDSR/article/view/1841

Pires, I. M., Hussain, F., Garcia, N. M., Lameski, P., & Zdravevski, E. (2020). Homogeneous data normalization and deep learning: A case study in human activity classification. *Mdpi.Com*. https://doi.org/10.3390/fi12110194

Qvarnström, M., Kahan, T., Kieler, H., Brandt, L., Hasselström, J., & Wettermark, B. (2019). Medication persistence to antihypertensive drug treatment–a cross-sectional study of attitudes towards hypertension and medication in persistent and non-persistent. *Taylor & FrancisM Qvarnström, T Kahan, H Kieler, L Brandt, J Hasselström, B WettermarkBlood Pressure, 2019•Taylor & Francis*, *28*(5), 309–316. https://doi.org/10.1080/08037051.2019.1627858

Wida, K. (2020). *Evaluasi ketepatan penggunaan obat pada Pasien Skizofrenia di Instalasi Rawat Inap Rumah Sakit Jiwa Mutiara Sukma Tahun 2019*. http://etheses.uin-malang.ac.id/id/eprint/25265

Wijaya, H., Rosyadi, A., Kurniawan, K., & Putri, W. (2023). *KOMUNIKASI FARMASI: Komuniasi Yang Efektif Dalam Praktik Farmasi*.