SHM PUBLISHER
SURYA HIJAU MANFAAT

Journal of Student Research Exploration

https://shmpublisher.com/index.php/josre
p-ISSN 2964-8246| e-ISSN 2964-1691

Analyzing reading preferences based on gender and education with decision tree method

Jelita Permata Sari[1]**, Maylinna Rahayu Ningsih**[2]
[1, 2] Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received April 18, 2024
Revised January 19, 2025
Accepted February 8, 2025

Keywords:

Classification
Decision tree
Algorithm
Book
Genre

ABSTRACT

This study aims to analyze the suitability of book genre selection with gender and education level. A classification method using a decision tree algorithm with four different criterion parameters is used to examine reading preferences based on various demographic factors, namely Gain Index, Information Gain, Gini Index, and accuracy. Data was obtained from a dummy dataset involving 120 records with three main attributes. The results show variations in accuracy depending on the criteria selected, with the highest accuracy rate achieved being 78.57%.

1. Introduction

Books are a valuable product of reflection because of their ability to not only provide entertainment, but also play a role in the process of education and increasing our knowledge horizons. Books are windows to a world of knowledge, imagination and understanding. In this era of globalization, where technological advances have developed rapidly, reading interest in the new generation tends to decline and is no better than the previous generation [1]. From childhood to adulthood, the role of books in building our understanding of this complex world

[1] ***Corresponding Author:***
Jelita Permata Sari,
Faculty of Mathematics and Natural Sciences,
Semarang State University.
Sekaran, Kec. Gn. Pati, Kota Semarang, Jawa Tengah 50229, Indonesia
Email: jelitaapermata@students.unnes.ac.id
DOI: https://doi.org/10.52465/josre.v3i1.341

is undeniable. However, in the diversity of genres offered by the world of literature, choosing the right book is key to enriching the reading and learning experience.

In Indonesia, reading is one aspect of language skills contained in the learning curriculum which is always present in every learning theme [2]. Reading skills are important to master because they are related to communication skills in the home, school, and community [3]. In the context of education, the selection of book genres that are appropriate for the level of education is very important. Each stage of child and adolescent development has different learning needs, and books with appropriate genres can be a tool.

Classification is a step in grouping objects based on the similarity of their characteristics into certain classes. Classification is a process of creating a function or model to explain the class of data or concepts in order to predict the class of an object whose label has not been obtained [4]. To provide the best classification performance, it is necessary to use a suitable classifier algorithm [5]. The classification model is a technique of predicting data, making predictions of the value of data whose results have been found to come from different data [6]. The purpose of this model is to predict the value of an unknown variable from other variables that have been given [7]. One method in classification is decision tree.

Decision tree is a classification algorithm that uses a decision tree structure to predict outcomes based on a set of decisions on dataset attributes. Decision trees use a "divide and conquer" technique to divide the problem search space into problem sets [8]. The process of a decision tree is to transform tabular data into a tree model [9]. Although it is easy to understand and can handle complex data, decision tree is prone to overfitting and may not produce an optimal model, it is still often used in various applications such as classification and prediction in various fields.

The use of the Decision Tree algorithm in the classification of book genre preference matches offers several advantages. These advantages include ease of interpretation and the ability to handle both categorical and numerical data without the need for special transformations. Decision tree will find a solution to the problem by making criteria as interconnected nodes forming a tree-like structure [10]. Decision tree is a prediction model for a decision using a hierarchical or tree structure [11]. Each tree has branches, branches represent an attribute that must be met to go to the next branch until it ends at the leaf (no more branches) [12]. Thus, the use of the Decision Tree algorithm is a useful step in the selection of book genres.

The purpose of the research conducted was to analyze the match of book genre selection with gender and education level to understand the different reading preferences among various demographic groups. As such, this study aims to provide greater insight into how reading preferences can be influenced by factors such as gender and education level, which in turn can help in designing more effective and relevant literacy programs.

2. Method

This research focuses on classification methods using various division selection criteria in the decision tree algorithm. Each model will be assessed using several selection criteria such as gain index, information gain, gini index, and accuracy. In the process of decision tree construction, metrics such as Gain Index, Information Gain, and Gini Index play a crucial role in feature selection [13]. These metrics effectively measure the ability of a feature to separate data into different classes, thus helping to optimize the tree structure. On the other hand, accuracy metrics are commonly used to evaluate the overall performance of the model, but their limitations in handling unbalanced datasets need to be noted. Figure 1 illustrates the flow of this research.

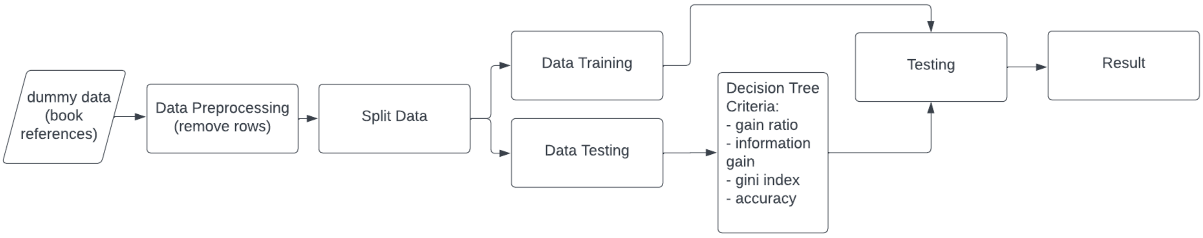

Figure 1. Research flow

Figure 1 shows the research flow with the first stage of splitting the dataset. The data will be divided into training and testing data three times with different ratios to get the best accuracy value. In each model will be set with different parameters, namely with four options namely gain ratio, information gain, gini index and accuracy. Next is to display the accuracy results of each data splitting ratio and criteria. The first step is to create a dataset that will be used. In this research, the dataset is created as dummy data with the help of an AI generator, ChatGPT. This data shows the compatibility of book genre preferences with the current level of education. The dummy data has 150 records with some missing values that have been handled using the row deletion method resulting in 120 records that have 3 attributes namely gender, education and book genre. The label that will be the goal is a preference match with matching and non-matching options. The attribute descriptions and their values are shown in Table 1.

Tabel 1. Atribut

Category	Name	Value
Artibut	Gender	Male, Female
	Education	SD,SMP,SMA,Sarjana,Megister,Doktor

	Book Genres	Fiction, Non-Fiction
Labels	Preference Match	Suitable, Not suitable

The second step is to divide the training and testing data in the ratio of 90%:10%, 80%:20%, and 70%:30%. The third step involves creating decision tree models, where each model uses attribute selection with four different criteria. The fourth step is to display the accuracy of each experiment that has been conducted, and then analyze it to explain the contribution of the research.

3. Results and Discussion

The result of this research is the accuracy value of testing using a decision tree model with four different criteria parameters, as well as testing techniques using split data. The data splitting process involves dividing the data into 90:10, 80:20, and 70:30. The process of creating a decision tree model with split data is shown in Figure 2.

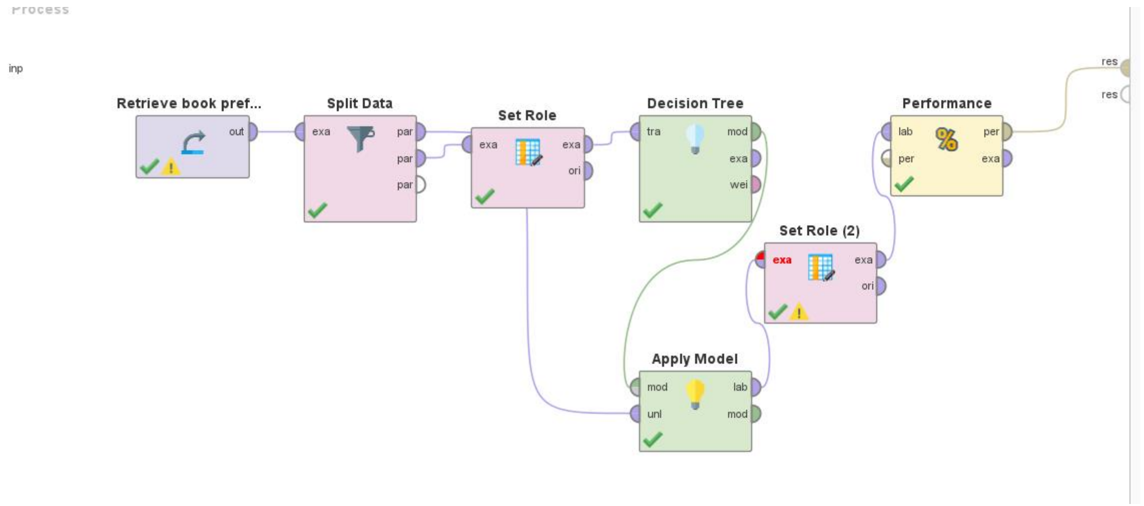

Figure 2. Decision tree process

. The steps in creating a decision tree model and split data start with separating the data into training and testing sections, then a decision tree model is created and attributes are selected for testing. After the model is applied to the testing data, the last step is to display the accuracy.

In the gain ratio, information gain and gini index criteria, the highest accuracy is 78.57% with 70% training data and 30% testing data, while the lowest accuracy is 63.89% obtained from 90% training data and 10% testing data. While the accuracy criteria produce different accuracy compared to the other three criteria, the test results are shown in table 2.

Table 2. Decision tree model accuracy test results

Criterion	Training	Testing	Accuracy
gain ratio	90	10	63.89%
	80	20	70.83%
	70	30	78.57%
information_gain	90	10	63.89%
	80	20	70.83%
	70	30	78.57%
gini_index	90	10	63.89%
	80	20	70.83%
	70	30	78.57%
accuracy	90	10	63.89%
	80	20	70.83%
	70	30	63.10%

From the table above, the best accuracy is generated in the data division with a ratio of 70:30. The resulting model achieved an accuracy of 78.57% which can be categorized as a fairly good result.

4. Conclusion

In conclusion, this research emphasizes the importance of selecting book genres that are appropriate to the educational context as well as the success of applying classification methods using decision tree algorithms with various data splitting selection criteria. Using a dummy dataset, this research illustrates the process of data splitting, decision tree modeling, and accuracy testing by varying the ratio of training and testing data. The results show variations in accuracy depending on the criteria selected, with the gain ratio, information gain, and gini index criteria achieving the highest accuracy at 70% training data and 30% testing data, reaching an accuracy of 78.57%. However, the accuracy criteria also gave different results, showing the importance of considering various factors in the selection of classification methods. Overall, this research makes a significant contribution in understanding reading preferences influenced by gender and education level, and offers valuable insights in the development of a more accurate and efficient classification system in book genre selection. For future research, it is recommended to explore alternative criteria combinations and data splitting techniques to broaden the understanding in a wider context [14][12], [15], [24]–[27], [16]–[23].

REFERENCES

[1] R. D. Utami, D. C. Wibowo, dan Y. Susanti, "Analisis Minat Membaca Siswa pada Kelas Tinggi di Sekolah Dasar Negeri 01 Belitang," *J. Pendidik. Dasar Perkhasa*, vol. 4, no. 1, hal. 179–188, 2018, doi: 10.31932/jpdp.v4i1.22.

[2] P. Handayani dan H. D. Koeswanti, "Pengembangan Media Komik untuk Meningkatkan

Minat Membaca Siswa Sekolah Dasar," *J. Basicedu*, vol. 4, no. 2, hal. 396–401, 2020, doi: 10.31004/basicedu.v4i2.365.

[3] A. R. Putri dan E. Purbaningrum, "Penggunaan Metode Mind Mapping terhadap Kemampuan Membaca Pemahaman Siswa Tunarungu Kelas 5 di SLB Diajukan Kepada Universitas Negeri Surabaya Penggunaan Metode Mind Mapping terhadap Kemampuan Membaca Pemahaman Siswa Tunarungu Kelas 5 di SLB," *Pendidik Khusus*, vol. 8, no. 2, hal. 1–10, 2016.

[4] P. P. Putra dan A. S. Chan, "Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4.5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor)," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, hal. 24, Jun 2018, doi: 10.35314/isi.v3i1.296.

[5] F. Baharuddin dan A. Tjahyanto, "Peningkatan Performa Klasifikasi Machine Learning Melalui Perbandingan Metode Machine Learning dan Peningkatan Dataset," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 1, hal. 25–31, Mar 2022, doi: 10.32736/sisfokom.v11i1.1337.

[6] M. Maulidah, Windu Gata, Rizki Aulianita, dan Cucu Ika Agustyaningrum, "ALGORITMA KLASIFIKASI DECISION TREE UNTUK REKOMENDASI BUKU BERDASARKAN KATEGORI BUKU," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, hal. 89–96, Des 2020, doi: 10.51903/e-bisnis.v13i2.251.

[7] A. Franseda, W. Kurniawan, S. Anggraeni, dan W. Gata, "Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, hal. 282, Jul 2020, doi: 10.26418/justin.v8i3.40982.

[8] M. H. Dunham, *Data Mining: Introductory and Advanced Topics*. Pearson Education India, 2003.

[9] S. Bahri dan A. Lubis, "METODE KLASIFIKASI DECISION TREE UNTUK MEMPREDIKSI JUARA ENGLISH PREMIER LEAGUE," *J. SIntaksis*, vol. 2, no. 1, hal. 63–70, 2020.

[10] S. H. Babic, P. Kokol, V. Podgorelec, M. Zorman, M. Sprogar, dan M. M. Stiglic, "The art of building decision trees," *J. Med. Syst.*, vol. 24, no. 1, hal. 43–52, Feb 2000, doi: 10.1023/a:1005437213215.

[11] N. Jayanti, S. Puspitodjati, dan T. Elida, "TEKNIK KLASIFIKASI POHON KEPUTUSAN UNTUK MEMPREDIKSI KEBANGKRUTAN BANK BERDASARKAN RASIO KEUANGAN BANK," 2008.

[12] D. Sartika dan D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *Jatisi*, vol. 1, no. 2, hal. 151–161, 2017.

[13] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, hal. 81–106, 1986, doi: 10.1007/BF00116251.

[14] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, "ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI," *J. Teknoinfo*, vol. 14, no. 2, hal. 115, Jul 2020, doi: 10.33365/jti.v14i2.679.

[15] A. Riski, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, hal. 22–28, 2019.

[16] Ferry Irawan, "Sistem Prediksi Penyakit Jantung Menggunakan Perbandingan Teknik Klasifikasi Data Mining," *J. Ilm. Akunt.*, vol. 1, no. 1, hal. 41–65, 2024.

[17] D. Fatmawati, W. Trisnawati, Y. Jumaryadi, dan G. Triyono, "Klasifikasi Tingkat Kepuasan Penggunaan Layanan Teknologi Informasi Menggunakan Decision Tree," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, hal. 1056–1062, 2023, doi: 10.30865/klik.v3i6.803.

[18] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, Okt 2018, doi: 10.30998/faktorexacta.v11i3.2777.

[19] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmaddeni, dan D. Wulandari, "Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, hal. 570–579, Mar 2024, doi: 10.57152/malcom.v4i2.1255.

[20] A. Nugroho, "Analisa Splitting Criteria Pada Decision Tree dan Random Forest untuk Klasifikasi Evaluasi Kendaraan," *JSITIK J. Sist. Inf. dan Teknol. Inf. Komput.*, vol. 1, no. 1, hal.

41–49, Des 2022, doi: 10.53624/jsitik.v1i1.154.

[21] D. Marutho, "Perbandingan Metode Naïve Bayes, KNN, Decision Tree Pada Laporan Water Level Jakarta," *J. Ilm. Infokam*, vol. 15, no. 2, hal. 90–97, 2019.

[22] S. A. Pratiwi, A. Fauzi, S. Arum, P. Lestari, dan Y. Cahyana, "KLIK: Kajian Ilmiah Informatika dan Komputer Prediksi Persediaan Obat Pada Apotek Menggunakan Algoritma Decision Tree," *Media Online*, vol. 4, no. 4, hal. 2381–2388, 2024, doi: 10.30865/klik.v4i4.1681.

[23] A. Y. Rahman, "Klasifikasi Citra Burung Lovebird Menggunakan Decision Tree dengan Empat Jenis Evaluasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, hal. 688–696, Agu 2021, doi: 10.29207/resti.v5i4.3210.

[24] D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, hal. 15–19, Mei 2023, doi: 10.57152/malcom.v3i1.591.

[25] Dewi Eka Putri dan Eka Praja Wiyata Mandala, "Hybrid Data Mining berdasarkan Klasterisasi Produk untuk Klasifikasi Penjualan," *J. KomtekInfo*, hal. 68–73, Jun 2022, doi: 10.35134/komtekinfo.v9i2.279.

[26] M. Firmansyah dan R. Aufany, "Implementasi Metode Decision Tree Dan Algoritma C4.5 Untuk Klasifikasi Data Nasabah Bank," *Infokam*, vol. XII, no. 1, hal. 1–12, 2016.

[27] A. Irma Purnamasari dan A. Rinaldi Dikananda, "Klasifikasi Kualitas Berita Pada Majalah Menggunakan Metode Decision Tree," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, hal. 48–54, 2023, doi: 10.56854/jtik.v1i2.52.