

Analysis of User Age Predictions in Public Satisfaction Surveys at Public Service Malls Using Decision Tree C4.5


Andysah Putera Utama Siahaan¹, Ami Abdul Jabar², Nelviony Parhusip³, Maida Indrayani⁴, Sipra Barutu⁵

^{1,2,3,4,5}Master of Information Technology, Universitas Pembangunan Panca Budi, Indonesia

ABSTRACT

This research analyzes the prediction of user age in the community satisfaction survey at the Public Service Mall (PSM) in Medan using the C4.5 Decision Tree algorithm. The primary objective of the study is to understand the demographic profile of users so that service managers can tailor their approaches to meet the needs of each age group. The data used includes 14,836 respondents with relevant demographic attributes. The analysis begins with data collection and preprocessing. The modeling results indicate that the Decision Tree model is effective in classifying users into age categories, including Late Senior, Early Senior, Middle Aged Adult, Young Adult, Late Teen, Early Teen, Child, and Toddler. The findings reveal a significant concentration in the Young Adult and Early Senior groups, indicating the need for adjustments in public services. The resulting recommendations aim to enhance service responsiveness to demographic needs and improve user satisfaction as well as the effectiveness of service strategies in the future.

Keyword : Decision Tree, C4.5, Age Prediction, Satisfaction Survey, Public Service Mall

 This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Andysah Putera Utama Siahaan,
Master of Information Technology
Universitas Pembangunan Panca Budi
Jl. Jend. Gatot Subroto Km. 4,5 Sei Sikambing 20122, Medan, Indonesia.
Email : andiesiahaan@gmail.com

Article history:

Received Mei 3, 2024
Revised Mei 10, 2024
Accepted Mei 16, 2024

1. INTRODUCTION

The development of technology and communication has been rapidly advancing over time, driving the use and application of technological advancements across various fields and aspects of life. This has enabled human tasks to be completed more efficiently. (Arisusanto Aditya, Suarna Nana, 2023) (Lubis & Hendrik, 2023). One human task that can be efficiently completed using technology is extracting information through data mining. Data mining is a data processing technique used to uncover valuable insights from collected data (Ananda M. Risqi, Maharani Nurul Sandra, Fadhila Eka, Rahma Alvia, 2023). Within data mining, there is a classification technique used to categorize data, making it easier to classify and analyze (A'yuniyah Qurotul, 2023).

The Decision Tree is one of the most popular classification methods because it is easy for humans to interpret. It can break down complex decision-making processes into simpler, more understandable components (Rahmansyah Nur, Ningsih Sari, Lantana Dhieka Avrilia, Suryaningtyas Adisti, Wirawan Putri, Wijaya Sifonne Adi, 2023). The C4.5 algorithm is one of the algorithms used to create decision trees. The Decision Tree method transforms a large set of events into a tree structure that represents rules. These rules can be easily understood in natural language (Sri Rahayu et al., 2023).

The use of data mining can be applied across various fields, including government and public services, as well as the private sector. (Iftitah Amalia, 2023). Public Service Mall is an innovation designed to connect and bring local government closer to the community in providing services (Nababan & Situmorang, 2023). The Public Service Mall (PSM) of Medan has conducted a community satisfaction survey; however, the PSM faces challenges in predicting the age of users based on the results of this survey.

Understanding the predicted age of users based on community satisfaction surveys provides significant benefits for tailoring services and policies. By comprehending the age profiles of users, service managers can adjust the types of services, communication, and approaches to be more effective according to the needs and preferences of different age groups. Therefore, this study will focus on

predicting the age of users at the Public Service Mall (PSM) in Medan using the Decision Tree C4.5 algorithm, with the aim of providing relevant recommendations for the development of better and more responsive services to meet community needs.

2. RESEARCH METHOD

a. Age Classification, according to the Ministry of Health the age classification is as follows: 1) Toddler: 0–5 Years; 2) Childhood: 5–11 Years; 3) Early Adolescence: 12–16 Years; 4) Late Adolescence: 17–25 Years; 5) Early Adulthood: 26–35 Years; 6) Late Adulthood: 36–45 Years; 7) Early Elderly: 46–55 Years; 8) Late Elderly: 56–65 Years; and 9) Senior Citizens: Over 65 Years (Nul, 2020).

b. Data Collection, the data collection used in this research is secondary data. Secondary data refers to existing data, whether from literature or sources published through reading, studying, quoting, and summarizing data related to the issues being discussed (Rahmawati Wahyu Eka, 2023). The secondary data is sourced from users of the Community Satisfaction Survey at the Public Service Mall in Medan, conducted from February to May 2024, with a total of 14,836 respondents.

c. Preprocessing, the obtained data undergoes preprocessing. Preprocessing is an essential step in the data mining process (Purbolaksono et al., 2021). Preprocessing is a step in preparing textual data to be used so that it can be processed in the next step (Permataning Tyas et al., 2022). The preprocessing stages involve removing data that contains null or empty values and restructuring the data through data cleaning and transformation. This step aims to ensure that the data to be processed is more organized and facilitates the modeling process (Septhya Dhini, Rahayu Khairisma, Rabbani Salsabila, Fitriana Vindi, Rahmaddeni, Irawan Yuda, 2022). To ensure a high-quality dataset that adheres to prediction application rules, we remove the following fields: name, tenant, department, gender, job name, education, date, respondent ID, U1, U2, U3, U4, U5, U6, U7, U8, U9. Thus, the only attributes used will be age and classification.

d. Modeling and Analysis, the processed data is modeled using RapidMiner, and the results of the modeling are then analyzed. RapidMiner is an application or software that serves as a learning tool in the field of data mining. The platform is developed by a company dedicated to all steps involving large amounts of data in commercial business, research, education, training, and learning (Yuanti, 2024). Rapidminer is the modeling tool used in this research (Nazar Yuniar, 2023).

3. RESULTS AND DISCUSSION

A. Data Collection

no	name	tenant	service	age	gender	job name	education	date	id responden	U1	U2	U3	U4	U5	U6	U7	U8	U9
1	Masryan	DINAS KEPENDUDUKAN DAN PENGEMBANGAN DOKUMEN	PENGAMBILAN DOKUMEN	35	Laki-laki	Pegawai SwS1	S1	22/01/2024	45_AB-1	4	4	4	4	4	4	4	4	4
2	Arto	DINAS KEPENDUDUKAN DAN PENDAFTARAN PELAYANA	PENDAFTARAN PELAYANA	35	Laki-laki	PNS/TNI/PGSD Keabawal	S1	21/01/2024	46_AA-2	4	4	4	4	4	4	4	4	4
3	Karno	DPMPSTSP KOTA MEDAN	LAYANAN PERBANTUAN PI	35	Laki-laki	Pegawai SwS1	S1	22/01/2024	86_BB0-1	4	4	3	2	2	4	4	4	4
4	Kevin	DPMPSTSP KOTA MEDAN	LAYANAN PERBANTUAN PI	33	Laki-laki	Pegawai SwS1	S1	22/01/2024	87_BB-2	4	4	4	4	4	4	3	4	3
5	Wina Afriani Purba	DPMPSTSP KOTA MEDAN	LAYANAN PERBANTUAN PI	34	Laki-laki	PNS/TNI/PGS1	S1	22/01/2024	88_BB-3	4	4	4	4	4	4	4	4	4
6	Ari	BADAN PENDAPATAN DAERAH	PAJAK REKLAME, AIR TANAH	22	Perempuan	PNS/TNI/PGSD Keabawal	S1	22/01/2024	91_AD-1	4	4	4	4	4	4	4	4	4
7	Yani	BADAN PENDAPATAN DAERAH	PAJAK HOTEL, RESTORAN	28	Perempuan	Lainnya	SMA	22/01/2024	90_AE-1	4	4	4	4	4	4	4	4	4
8	Juandi	PERUMDA TIRTANADI WILAYAH	PEMBAYARAN ONLINE	54	Laki-laki	Pegawai SwS1	S1	22/01/2024	109_BH-1	4	4	4	4	4	4	4	4	4
9	Ririn	PERUMDA TIRTANADI WILAYAH	PEMBAYARAN ONLINE	54	Perempuan	Pegawai SwS1	S1	22/01/2024	110_BH-2	4	4	4	4	4	4	4	4	4
10	Ahmad	BPJS KESEHATAN	LAYANAN INFORMASI	23	Laki-laki	Pegawai SwS1	S1	22/01/2024	119_CJ-1	4	4	4	4	4	4	4	4	4
11	Kurnia	BPJS KETENAGAKERJAAN	KLAIM JHT	33	Laki-laki	PNS/TNI/PGS1	S1	22/01/2024	121_CN-1	4	4	3	4	4	3	4	4	4
12	ELLA	BPJS KESEHATAN	LAYANAN ADMINISTRASI	17	Perempuan	Wiraswasta	S1	22/01/2024	124_CI-1	4	4	4	4	4	4	4	4	4
13	Prayatno	PLN UP3 MEDAN DAN UP3	PENGADUAN GANGGUAN	23	Laki-laki	PNS/TNI/PGS2 Keatas	S1	22/01/2024	136_BJ-1	4	4	3	3	4	4	4	4	3
14	Abi	BPJS KETENAGAKERJAAN	KLAIM JHT	23	Laki-laki	Lainnya	SMA	22/01/2024	123_CN-2	4	4	3	3	4	3	4	4	4
15	Indi	BPJS KETENAGAKERJAAN	KLAIM JHT	35	Laki-laki	Lainnya	S1	22/01/2024	133_CN-3	4	4	3	4	4	4	4	4	4
16	Gladius Aulia Barus	DINAS KETENAGAKERJAAN	CARTU PENCARI KERJA AK	19	Perempuan	Wiraswasta	SMA	22/01/2024	130_BI-3	4	3	4	4	4	4	4	4	4
17	Riri	DINAS KOPERASI, UMKM,	LAYANAN UMKM	27	Perempuan	PNS/TNI/PGS1	S1	23/01/2024	150_CQ-1	3	3	4	4	4	4	4	4	4
18	Juan	DINAS KEPENDUDUKAN DAN PENDAFTARAN PELAYANA	PENDAFTARAN PELAYANA	18	Laki-laki	Lainnya	SMA	23/01/2024	144_AA-1	4	4	4	4	4	4	4	4	4
19	Nanda	DINAS KEPENDUDUKAN DAN PENDAFTARAN PELAYANA	PENDAFTARAN PELAYANA	23	Laki-laki	Lainnya	SMA	23/01/2024	147_AA-2	4	4	4	4	4	3	3	4	4
20	Jeli	DINAS KEPENDUDUKAN DAN PENDAFTARAN PELAYANA	PENDAFTARAN PELAYANA	24	Laki-laki	Wiraswasta	S1	23/01/2024	148_AA-3	4	4	4	4	4	4	4	4	4
21	Aulia	DINAS KEPENDUDUKAN DAN PENDAFTARAN PELAYANA	PENDAFTARAN PELAYANA	35	Laki-laki	Lainnya	IS1	23/01/2024	151_AA-4	4	4	4	4	4	4	4	4	4

Figure 1: Data Collection

In Figure 1, it can be seen that the data from the Community Satisfaction Survey at the Public Service Mall in Medan consists of 14,836 respondents, which includes the columns for name, tenant, department, age, gender, job name, education, date, respondent ID, and U1 through U5.

B. Preprocessing

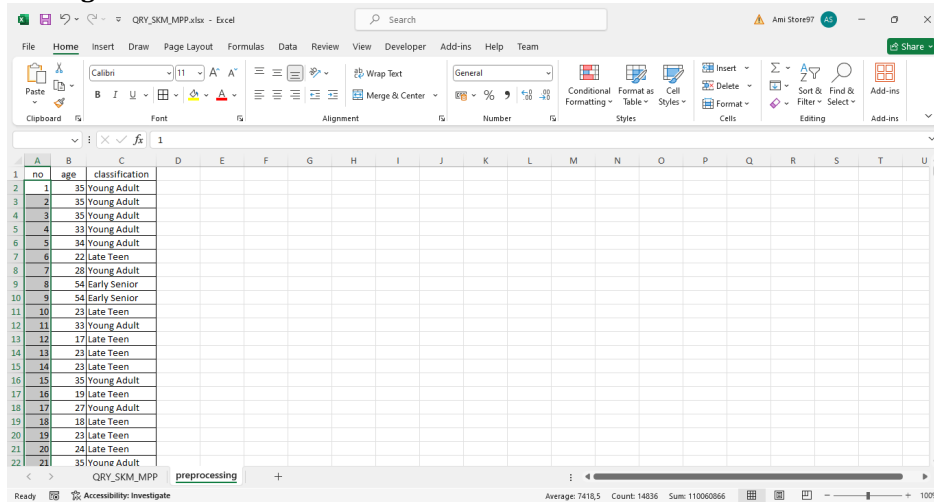


Figure 2: Preprocessing

In Figure 2, the data has undergone preprocessing, resulting in the attributes for age and classification.

C. Modeling and Analysis

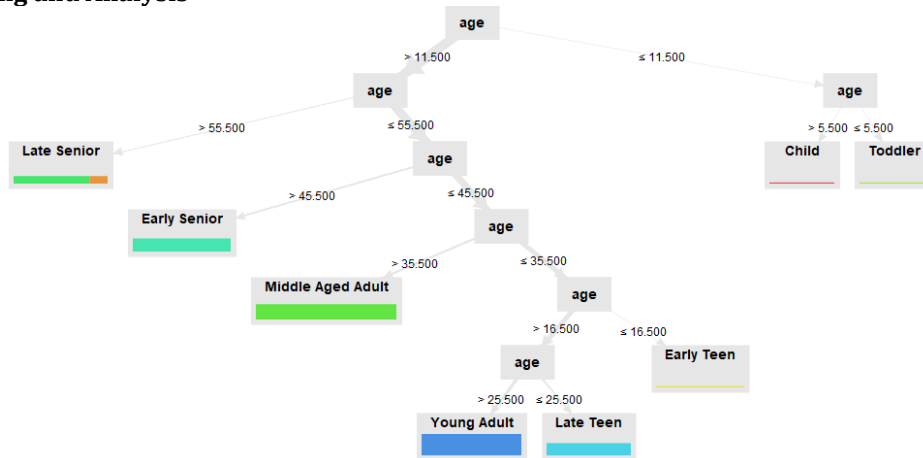


Figure 3: Modeling

In Figure 3, after preprocessing the data, the modeling is performed using the C4.5 decision tree. The Decision Tree model for predicting the age of users in the community satisfaction survey at the Public Service Mall in Medan consists of several levels, starting with a root node that determines the initial split based on age, such as age > 11,500. From here, the first branch identifies categories like Late Senior (age > 55,500) with the number of users in that category. The second branch separates users in a lower age range, including Early Senior (ages 45,500 - 55,500) and Middle Aged Adult (ages 35,500 - 45,500), while the third branch groups younger users, such as Young Adult (ages 16,500 - 35,500) and Late Teen (ages 16,500 - 25,500). The leaf nodes in this model indicate the number of users in each category, providing a clear picture of the demographic distribution. The results of this analysis not only enhance the understanding of user characteristics but also provide a foundation for formulating more responsive and effective public service strategies tailored to the needs of each age group.



Figure: 4 Description Tree

In Figure 4, it can be seen that this decision tree categorizes users by age into several groups, including Late Senior, Early Senior, Middle Aged Adult, Young Adult, Late Teen, Early Teen, Child, and Toddler. For users older than 11,500, there are two main branches: Late Senior (age > 55,500) with 1,206 users and Elderly with 286 users, as well as Early Senior (age ≤ 55,500, > 45,500) comprising 2,765 users. Under the Middle Aged Adult category (ages 35,500 - 45,500), there are 3,258 users, while younger users are grouped into Young Adult (4,638 users, ages 16,500 - 25,500) and Late Teen (2,566 users). Users under 11,500 consist of 33 Children and 42 Toddlers. This analysis highlights a significant concentration among Young Adults and Early Seniors, indicating the need for tailored public services to meet the specific needs of these two age groups, which could enhance user satisfaction and improve marketing strategies.

4. CONCLUSION

The analysis using the C4.5 Decision Tree algorithm demonstrates its effectiveness in predicting the age of users in the community satisfaction survey at the Public Service Mall (PSM) in Medan, based on data from 14,836 respondents. This model successfully classifies users into age categories, including Late Senior, Early Senior, Middle Aged Adult, Young Adult, Late Teen, Early Teen, Child, and Toddler. The results indicate a significant concentration in the Young Adult and Early Senior groups, highlighting the need for tailored public services to meet their specific needs. Understanding user demographics allows service managers to formulate more effective communication strategies and policies, enhancing community satisfaction and service effectiveness. Recommendations from this analysis support the development of more responsive public services that align with age-based preferences, potentially improving service quality in the future.

REFERENCES

- Ananda M. Risqi, Maharani Nurul Sandra, Fadhila Eka, Rahma Alvia, N. (2023). Data Mining Dalam Perusahaan PT Indofood Lubuk Pakam. *Jurnal Ekonomi Manajemen Dan Bisnis (JEMB)*, 02(1), 97–102. <https://doi.org/10.47233/jemb.v2i1.1009>
- Arisusanto Aditya, Suarna Nana, D. G. (2023). Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma Random Forest Dengan Optimize Parameter Grid. *Jurnal Teknologi Ilmu Komputer*, 1(2), 43–47. <https://doi.org/10.56854/jtik.v1i2.51>
- A'yuniyah Qurotul, R. M. (2023). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 3(1), 39–45. <https://doi.org/10.57152/ijirse.v3i1.484>
- Iftitah Amalia, S. R. (2023). Penerapan Algoritma C.45 Untuk Analisis Pengadaan Peralatan dan Mesin Kantor. *Journal of Information System Research (JOSH)*, 4(2), 434–442. <https://doi.org/10.47065/josh.v4i2.2673>
- Lubis, S. S., & Hendrik, B. (2023). Implementasi Data Mining Pengelompokan Data Penjualan Berdasarkan Pembelian Dengan Menggunakan Algoritma K-Means Pada UD.Martua. *Journal of Information System*

- and Education Development*, 1(3), 36–41. <https://journal.widyakarya.ac.id/index.php/jusiik-widyakarya/article/view/1531%0Ahttps://journal.widyakarya.ac.id/index.php/jusiik-widyakarya/article/download/1531/1563>
- Nababan, W., & Situmorang, E. G. V. (2023). STRATEGI PENINGKATAN KEEFEKTIFAN MAL PELAYANAN PUBLIK PADA PENYELENGGARAAN PELAYANAN ADMINISTRASI KEPENDUDUKAN DI KABUPATEN SUMEDANG MENGGUNAKAN ANALISIS SOAR DAN MATRIKS QSPM. *Jurnal Registratie*, 5(1), 1–19. <https://doi.org/10.33701/jurnalregistratie.v5i1.3343>
- Nazar Yuniar, M. (2023). Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve bayes. *Jurnal Sains Dan Teknologi*, 5(1), 243–246. <https://doi.org/10.55338/saintek.v5i1.1383>
- Nul, H. L. (2020). Urgensi Revisi Undang-Undang tentang Kesejahteraan Lanjut Usia. *Aspirasi: Jurnal Masalah-Masalah Sosial*, 11(1), 43–55. <https://doi.org/10.22212/aspirasi.v11i1.1589>
- Permataning Tyas, S. M., Rintyarna, B. S., & Suharso, W. (2022). The Impact of Feature Extraction to Naïve Bayes Based Sentiment Analysis on Review Dataset of Indihome Services. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 13(1), 1–10. <https://doi.org/10.31849/digitalzone.v13i1.9158>
- Purbolaksono, M. D., Irvan Tantowi, M., Imam Hidayat, A., & Adiwijaya, A. (2021). Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 393–399. <https://doi.org/10.29207/resti.v5i2.3008>
- Rahmansyah Nur, Ningsih Sari, Lantana Dhieka Avrilia, Suryaningtyas Adisti, Wirawan Putri, Wijaya Sifonne Adi, P. D. N. (2023). Komparasi Metode Knn, Naïve Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *INNOVATIVE: Journal Of Social Science Research Volume*, 3(2), 13880–13892.
- Rahmawati Wahyu Eka, S. (2023). Analisis Inflasi-Kurs dan BI Rate Terhadap Indeks Harga Saham Gabungan di Bursa Efek Indonesia Tahun 2015-2019. *Jurnal Akuntansi Dan Teknologi Keuangan*, 1(2), 64–82. <https://doi.org/10.56854/atk.v1i2.164>
- Septhya Dhini, Rahayu Khairisma, Rabbani Salsabila, Fitria Vindi, Rahmaddeni, Irawan Yuda, H. R. (2022). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19.
- Sri Rahayu, D., Afifah, J., & Intan, S. (2023). Klasifikasi Penyakit Diabetes Melitus Menggunakan Algoritma C4.5, Support Vector Machine (SVM) dan Regresi Linear. *Institut Riset Dan Publikasi Indonesia (IRPI) SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 56–63. <https://journal.irpi.or.id/index.php/sentimas>
- Yuanti, A. H. (2024). Analisis Pengaruh Covid-19 Terhadap Kesehatan Mental dengan Visualisasi Data Rapidminer. *Gudang Jurnal Multidisiplin Ilmu*, 2(1), 183–187. <https://doi.org/10.59435/gjmi.v2i1.225>
-