Journal of Information Technology, computer science and Electrical Engineering (JITCSE)
Vol. 1, No. 2, June 2024, pp. 149~152
ISSN: 3046-4900, DOI: 10.61306/jitcse.v1i2

Analysis of Age and Gender Classification Using Decision Tree Model in the Context of Nursing Homes

Andysah Putera Utama Siahaan[1], Muhammad Indra[2]
[1,2]Master of Information Technology, Universitas Pembangunan Panca Budi, Indonesia

ABSTRACT

By leveraging Data Mining technology, specifically the Decision Tree algorithm, this study focuses on clustering data based on age and gender to enhance the efficiency and personalization of services in nursing homes. The data used spans from January 2024 to April 2024, encompassing 333 rows that have been processed for classification purposes. The developed Decision Tree model accurately separates the data based on age, with results showing the gender distribution within each age group. These findings indicate that the Decision Tree algorithm is effective in identifying gender based on specific age boundaries, which can be applied to improve the quality and effectiveness of nursing home services. The analysis provides valuable insights for better planning and management of social services, making this approach relevant for demographic data management in nursing homes.

Keyword : Decision Tree, Age Classification, Gender, Data Mining, Nursing Homes

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 May 14, 2024
Revised May 17, 2024
Accepted May 23, 2024

1. INTRODUCTION

Technological advancement is essentially a testament to the evolution of human thought (Aminullah & Ali, 2020). Fundamentally, technology is designed to make human tasks easier (Putri & Sutabri, 2023). Data Mining technology can be used to make predictions by leveraging historical data to infer future events (Putri & Sutabri, 2023). Data Mining is a common method for extracting information from big data (Alwendi et al., 2023). It can also be defined as a process that employs statistical, mathematical, artificial intelligence, and machine learning techniques to identify and extract useful information and knowledge from large databases (Yolanda Paramitha et al., 2023).

The Decision Tree algorithm is a frequently used Data Mining algorithm for solving classification problems (Kurnia Illahi et al., 2023). Classification involves grouping data or studying something based on its characteristics (Widiastuti et al., 2023). RapidMiner is one of the software tools for data mining processing (Indranu et al., 2024).

In a social context, nursing homes provide essential services for the elderly, including basic needs such as food and clothing, physical and spiritual care, social guidance, and appropriate religious counseling (Kumalasari et al., 2023). By applying age and gender classification using Data Mining technology and Decision Tree algorithms, this information can assist in tailoring services and meeting the needs of nursing home residents. This technology enables more efficient data management and service personalization, thereby enhancing the effectiveness and quality of the assistance provided.

2. RESEARCH METHOD

a. Data Classification: Data is classified based on age and gender.
b. Data Collection: This is a crucial step in obtaining the necessary information (Saefuddin et al., 2023). The data collected includes the "DAFTAR WARGA BINAAN SOSIAL UPTD PELAYANAN SOSIAL LANJUT USIA KISARAN, JL. PERINTIS KM.8 SIMPANG TIGA LEMANG, KEC. SIMPANG EMPAT, KAB. ASAHAN" from January 2024 to April 2024.

c. Data Preprocessing: This is an important step before starting the data mining process, as it ensures the quality of the dataset before modeling (Khairunnas et al., 2023).
d. Model and Analysis: After the data preprocessing phase, the next step is to generate the Decision Tree model and then analyze the results of this model.

3. RESULTS AND DISCUSSION

A. Data Collection

	Name WBS	Wife/Husband's Name	Number of Cl	Age	Gender	Date of Entry	Room	No. KTP	No. KK	Origin Location
313	Toga Sinambela	-	-	60	Male	08/04/2024	Jeruk	'402171058(	(402021012(	Kisaran
314	Misnan	-	-	63	Male	08/04/2024	Jeruk	-	-	Tanjungbalai
315	Sudirman	-	-	72	Male	08/04/2024	Jeruk	-	-	Tanjungbalai
316	Irianto	-	-	85	Male	08/04/2024	Jeruk	-	-	Pinangsori
317	Albiner Simamora	-	-	60	Male	08/04/2024	Jeruk	-	-	Asahan
318	Rahmawati	-	-	75	Male	15/04/2024	Langsat	-	-	Kisaran Barat
319	Roliyah	-	2	80	Male	15/04/2024	Langsat	'402711241(	(402121107(	Kisaran
320	Naffiah	-	-	74	Female	15/04/2024	Langsat		-	Tanjung Balai
321	Gustina Wati Sitorus	-	-	64	Female	15/04/2024	Langsat	'830521159(	-	Tanjung Balai
322	Tomoh Sihombing	-	-	64	Female	17/04/2024	Mangga	-	-	Medan
323	Ayub Silalahi	-	-	60	Female	17/04/2024	Mangga	-	-	Tanjung Balai
324	Saman Ilyas	Darmawati	1	74	Male	17/04/2024	Mangga	'4012811480002		Rantau Prapat
325	Baharuddin	-	-	68	Male	17/04/2024	Mangga	'4051412070004		Pinangsori
326	Siti Fatimah	Alm.Amin Panjaitan	-	61	Male	25/04/2024	Nenas	'920450859(	(920051107(	Tanjungbalai
327	Rubiah	Alm.Sukarno	3	66	Male	25/04/2024	Nenas	-	-	Tanjungbalai
328	Mariati Purba	-	-	60	Female	25/04/2024	Nenas	-	-	Kisaran
329	Ade Rospita Nasution	-	-	60	Female	25/04/2024	Nenas	-	-	Tanjungbalai
330	Mahfuri Siregar	-	-	87	Female	25/04/2024	Nenas	-	-	Siantar
331	Suwarni	-	-	64	Female	25/04/2024	Pepaya	'922711258(	(922020720(	Medan
332	Meri Sianturi	-	-	65	Female	25/04/2024	Pepaya	'107500358(	-	Tanjungbalai
333	Ernawati Lingga	-	-	60	Female	25/04/2024	Pepaya	-	-	Siantar

Fig 1. Data Collection

In Figure 1, the data collection consists of 333 rows, including the header, and includes the following columns: Name WBS, Wife/Husband's Name, Number of Children, Age, Gender, Date of Entry, Room, ID Number, Family Card Number, Origin Location, and Information.

B. Data Processing

	Age	Gender
313	60	Male
314	63	Male
315	72	Male
316	85	Male
317	60	Male
318	75	Male
319	80	Male
320	74	Female
321	64	Female
322	64	Female
323	60	Female
324	74	Male
325	68	Male
326	61	Male
327	66	Male
328	60	Female
329	60	Female
330	87	Female
331	64	Female
332	65	Female
333	60	Female

Fig 2. Data Preprocessing

In Figure 2, after preprocessing, the dataset remains at 333 rows, including the header, but the number of columns is reduced to only Age and Gender.

B. Model and Analysis

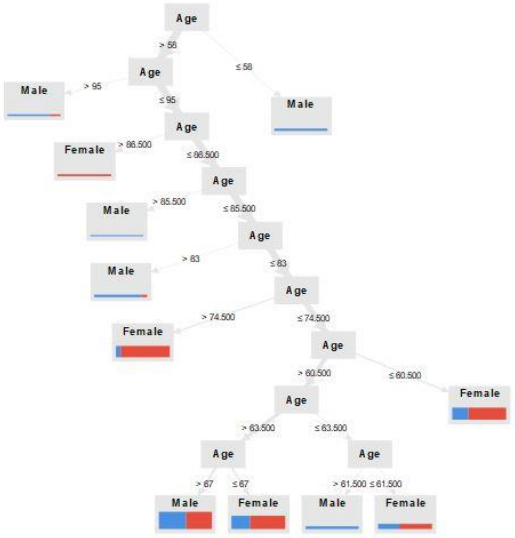

Fig 3. Model Decision Tree

In Figure 3, the Decision Tree model is used to classify data based on age. The process begins with the root node that splits the data at the age of 58 years. If the age is greater than 58, the tree continues to the next branch with the criterion of age greater than 95, which identifies the result as "Male." For ages less than or equal to 95, the tree further divides based on age greater than 86.500, which identifies the result as "Female." If the age is less than or equal to 86.500, the next split is at age greater than 85.500, resulting in "Male." Ages less than or equal to 85.500 are split further at age greater than 83, which identifies the result as "Male." Ages less than or equal to 83 are divided at age greater than 74.500, resulting in "Female." If the age is less than or equal to 74.500, the next division is at age greater than 60.500. For ages greater than 60.500, the tree continues to the next branch with age greater than 63.500, splitting further at age greater than 67, resulting in "Male," and ages less than or equal to 67, resulting in "Female." If the age is less than or equal to 63.500, the next split is at age greater than 61.500, which identifies the result as "Male," while ages less than or equal to 61.500 are identified as "Female." Finally, for ages less than or equal to 60.500, and ages less than or equal to 58, the result is "Male." Each branch of this Decision Tree is designed to separate the data based on age, identifying the gender result at each leaf node.

Tree

```
Age > 58
|   Age > 95: Male {Male=4, Female=1}
|   Age ≤ 95
|   |   Age > 86.500: Female {Male=0, Female=6}
|   |   Age ≤ 86.500
|   |   |   Age > 85.500: Male {Male=2, Female=0}
|   |   |   Age ≤ 85.500
|   |   |   |   Age > 83: Male {Male=8, Female=1}
|   |   |   |   Age ≤ 83
|   |   |   |   |   Age > 74.500: Female {Male=5, Female=51}
|   |   |   |   |   Age ≤ 74.500
|   |   |   |   |   |   Age > 60.500
|   |   |   |   |   |   |   Age > 63.500
|   |   |   |   |   |   |   |   Age > 67: Male {Male=45, Female=42}
|   |   |   |   |   |   |   |   Age ≤ 67: Female {Male=23, Female=43}
|   |   |   |   |   |   |   Age ≤ 63.500
|   |   |   |   |   |   |   |   Age > 61.500: Male {Male=10, Female=0}
|   |   |   |   |   |   |   |   Age ≤ 61.500: Female {Male=9, Female=14}
|   |   |   |   |   |   Age ≤ 60.500: Female {Male=18, Female=42}
Age ≤ 58: Male {Male=8, Female=0}
```

Fig 4. Decision Tree Model Description

In Figure 4, the Decision Tree Model Description divides data based on age to identify gender. The process begins with ages greater than 58 years. The first split at ages greater than 95 identifies as "Male," while ages less than or equal to 95 are further divided at ages greater than 86.500, identifying as "Female." If the age is less than or equal to 86.500, the next split occurs at ages greater than 85.500, identifying as "Male." Ages less than or equal to 85.500 are further split at ages greater than 83, resulting in "Male." Ages less than or equal to 83 are divided at ages greater than 74.500, identifying as "Female." For ages less than or equal to 74.500, the data is split at ages greater than 60.500, with further division at ages greater than 63.500. This split shows a balanced result between males and females for ages greater than 67, with differing distributions below this age. Ages less than or equal to 63.500 are split at ages greater than 61.500, resulting in "Male" on one branch and "Female" on the other. Finally, for ages less than or equal to 60.500 and 58, the results are "Female" and "Male," respectively. This Decision Tree illustrates how gender distribution is influenced by age within the dataset.

4. CONCLUSION

The analysis of age and gender classification using the Decision Tree model demonstrates that this algorithm is effective in segmenting data by age to identify gender in a nursing home setting. The model partitions the data by grouping individuals based on specific age thresholds, resulting in accurate classifications of males and females. This approach enables the obtained information to enhance personalization and efficiency of services within the nursing home, ensuring that needs and services are more appropriately tailored to each individual. Utilizing the Decision Tree in this context allows for a deeper understanding of age and gender distribution, contributing to improved planning and management in social services.

REFERENCES

Alwendi, Mandopa, A. S., & Hasibuan, E. A. (2023). Aplikasi Data Mining Untuk Menentukan Masa Studi Mahasiswa Menggunakan Metode Association Rule Data Mining Application to Determine Student Study Period Using Association Rule Method. JURNAL PENDIDIKAN DEWANTARA, 2(1), 1–6. https://doi.org/10.58222/dewantara.v2i1.24

Aminullah, M., & Ali, M. (2020). KONSEP PENGEMBANGAN DIRI DALAM MENGHADAPI PERKEMBANGAN TEKNOLOGI KOMUNIKASI ERA 4.0. Komunike, XII(1), 1–23.

Indranu, F., Kesuma, M., & Kalifia, A. D. (2024). PENGARUH MEDIA SOSIAL TERHADAP TINGKAT ANXIETY PADA REMAJA : SEBUAH ANALISIS DENGAN RAPIDMINER. Jurnal Ilmiah Multidisiplin, 1(3), 177–181. https://doi.org/10.62017/merdeka

Khairunnas, M. A., Jamaludin, A., & Adam, R. I. (2023). Pengaruh Pendapatan Orang Tua terhadap Hasil Belajar Siswa Menggunakan Algoritma K-Means Clustering. Jurnal Pendidikan Tambusai, 31434–31444.

Kumalasari, A., Agustang, A., & Syukur, M. (2023). FENOMENA PENITIPAN ORANG TUA LANSIA DI PANTI SOSIAL TRESNA WERDHA GAU MABAJI, KABUPATEN GOWA. Phinisi Integration Review, 6(1), 111–117.

Kurnia Illahi, P., Rina Viana, A., Fitria, N., Permata, M., & Pratama, M. Y. (2023). SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Application of Decision Tree Algorithm and Linear Regression for Breast Cancer Classification Penerapan Algoritma Decision Tree aan Regresi Linear untuk Klasifikasi Kanker Payudara. SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat, 86–92. https://journal.irpi.or.id/index.php/sentimas

Putri, G. B., & Sutabri, T. (2023). Analisis Manajemen Layanan Teknologi Informasi Menggunakan ITIL V3 Domain Service Operation Pada Perusahaan CV. Cemerlang Komputer Palembang. Indonesian Journal of Multidisciplinary on Social and Technology, 1(2), 162–167. https://doi.org/10.31004/ijmst.v1i2.144

Saefuddin, M. T., Wulan, T. N., Savira, & Juansah, D. E. (2023). TEKNIK PENGUMPULAN DATA KUANTITATIF DAN KUALITATIF PADA METODE PENELITIAN. Pendas : Jurnal Ilmiah Pendidikan Dasar, 08(03), 5962–5974.

Widiastuti, N., Hermawan, A., & Avianto, D. (2023). KOMPARASI ALGORITMA KLASIFIKASI DATAMINING UNTUK PREDIKSI MINAT PENCARI KERJA. JURNAL TEKNOINFO, 17(1), 219–227. https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index

Yolanda Paramitha, N., Nuryaman, A., Faisol, A., Setiawan, E., Dina Eka Nurvazly, dan, Matematika, J., Mipa, F., Lampung Jl Soemantri Brojonegoro No, U., Lampung, B., Kunci, K., & Bayes, N. (2023). Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes. In Jurnal Siger Matematika (Vol. 04, Issue 01). https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-