Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12247
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Application Of The C4.5 Algorithm to Determine Security Guard Work Schedules

Ita Dewi Sintawati[1)*], Widiarina[2)], Kartika Mariskhana[3)]
[1,2,3)]Universitas Bina Sarana Informatika, Indonesia
[1)] ita.ids@bsi.ac.id, [2)] widiarina.wda@bsi.ac.id, [3)] kartika.kma@bsi.ac.id

Submitted : Mar 8, 2023 | **Accepted** : Apr 7, 2023 | **Published** : Apr 7, 2023

Abstract: All agencies, companies, and public spaces must employ security guards to maintain security. The problem with the security guard's schedule is that there is an imbalance in disciplinary issues, poor performance because there is no seniority that is emulated so that the guard is not optimal which results in the problem of losing employee items, working time is not according to the rules and there is a vacancy in personnel due to personnel not coming to work, suing With the existence of a policy in the placement and distribution of the right employee work schedule, it is hoped that it can synergize all elements in the institution so that the quantity and quality of the security guard's work can increase and be completed on time. One of the techniques in data mining is classification. By applying classification techniques to security guard data and work schedules, the Decision Tree method and C4.5 algorithm are developed. The results of data processing form the root node of the gender tree as the root, that those who get schedule A are men while those who get schedule B with high school and junior level education are women, besides that they get schedule A. The accuracy of all classifications of the correct number is 61, 53%.

Keywords: Security Guard Work Schedule; Data Mining; level of accuracy; C4.5 Algorithm

INTRODUCTION

Companies that provide security guard services to assist in securing the environment around the factory (Ansori Aan, 2016), given the rapid development of technology, a solution is needed in making security guard schedules to make it easier to divide employee work shifts according to qualifications (Robert Tua Siregar, 2020) for example, such as dividing the schedule for the night shift or day shift by dividing it according to gender, for example for male employees, the night shift is made and employees whose gender is female are made for the day shift or to be able to determine the level of workers such as recent education and so on. Every company wants its employees to work optimally according to their respective expertise (Margono, 2020). There are many factors that affect the performance of an employee, such as age, gender, level of discipline, and others (Kurniawan, 2019). Based on these factors, a work schedule must be prepared according to the criteria of each employee. If the number of employees reaches dozens or hundreds of people, it will be very difficult to make a work schedule taking into account the criteria of each employee.

To overcome these obstacles, a data mining analysis method was created using the C4.5 Algorithm (Kiu, 2021), which is capable of producing a work schedule that can divide employee schedules or shifts according to the criteria set by the company. By making a work schedule that fits the criteria of employees it will be easy to do, thus every employee will be able to do the job to the fullest (Julianto, 2019).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12247 p-ISSN : 2541-044X

LITERATURE REVIEW

Previous research that the author uses as a reference includes the following:

1. Research conducted by Triuli Novianti, et al, 2016. Application of the C4.5 Algorithm. This research is a Data Mining application that can be used to display useful information about Employee Work Schedules with data mining techniques. The information displayed is in the form of a relationship between Employee Work Schedules and Age, Gender and Work Unit. This can be seen in the decision tree and decision tree based application design. From the 103 dirty data after cleaning the incomplete data and going through the selection process as required in the analysis of employee performance schedules, all complete data is obtained so that 103 clean data can be used. From the existing data, it was split to use 1/3 (34) training data and 2/3 (69) test data. The results of testing the entire data using 5-fold cross validation resulted in a test accuracy of 70% (Novianti, 2016).
2. Research conducted by Ikhsan Romli, et al, 2020. The classification used in data mining is Decision tree because it is a technique that is widely used and produces output with existing rules, with this it can present employee data to determine overtime schedules. In this study using the C4.5 algorithm to produce determine the overtime schedule. The results of the overtime schedule trial using the C4.5 algorithm with the Confusion matrix have good accuracy, precision and recall values, namely 91% accuracy, 86.05% precision and 92.5% recall (Romli, 2020).
3. Research conducted by G L Pritalia, et al, 2018. to maintain the stock available in the warehouse to be stable. The stable stock referred to in the warehouse is that the goods are not overstocked or the goods are not out of stock. Decision trees are a very powerful and well-known method of classification and prediction. The richer the information or knowledge contained in the training data, the greater the accuracy. This algorithm is used to analyze the time to purchase stocks that are already depleted by classifying which items are time to add stock or not, so that the availability of goods remains stable and maintained. The results of the analysis using the C4.5 algorithm are to determine when determining the availability of goods has an accuracy rate of 98.9% (Lukhayu Pritalia, 2018).

Data mining is a process for finding useful information from a large collection of databases stored in storage using pattern recognition techniques such as statistical techniques, mathematics, artificial intelligence, and machine learning (Turnip, 2018). This method is one of the existing methods of classification techniques in data mining. The decision tree method turns very large facts into decision trees that represent rules (Alkhairi, 2022). Decision trees are also useful for exploring data, finding hidden relationships between a number of candidate input variables and a target variable (Setio, 2020). The data in a decision tree is usually expressed in the form of a table with attributes and records. Attributes declare a parameter which is referred to as a criterion in the formation of a tree. For example, to decide to play tennis, the criteria to consider are weather, wind, and temperature. One of the attributes is an attribute that represents the solution data per data item called the result attribute. Many algorithms can be used in the formation of decision trees, including ID3, C4.5, CART. Algorithm C4.5 is a development of the ID3 algorithm (Pratama, 2020). The C4.5 and ID3 algorithms were created by a researcher in the field of artificial intelligence named j. Rose quinlan in the late 1970s. Algorithm C4.5 builds a decision tree from top to bottom, where the top attribute is the root, and the bottom attribute is called the leaf. There are several stages in making a decision tree in the C4.5 algorithm, namely (Azwanti, 2018):

1. Prepare training data. Training data is usually taken from historical data that has happened before and has been grouped into certain classes.
2. Counting tree roots. The roots will be taken from the attribute to be selected, by calculating the gain value of each attribute, the highest gain value will be the first root. Before calculating the gain value of the attribute, first calculate the entropy value.

To calculate the entropy value, the formula is used:

$$Entro(S) = \sum_{i=1}^{n} - pi * Log2\ pi \qquad (1)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12247 p-ISSN : 2541-044X

Information:
S : Set of Cases
n : Number of Partitions S
pi : The proportion of Si to S
Then after the entropy value for each attribute has been obtained, calculate the gain value using the formula:

$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} (|si|)/(|s|) * Entropy(Si)$ (2)

Information :
S : Set of Cases
A : Features
n : Number of partitions attribute A
Si| : The number of cases on the i-th partition
|S| : Number of cases in S

Rapid Miner is software that is open (open source). Rapid Miner is a solution for analyzing data mining, text mining and predictive analysis (Utami, 2021). Rapid Miner uses a variety of descriptive and predictive techniques to provide insights to users so they can make the best decisions. Rapid Miner has approximately 500 data mining operators, including operators for input, output, data preprocessing and visualization.

Rapid Miner is a stand-alone software for data analysis and as a data mining machine that can be integrated into its own product. Rapid Miner is written using the Java language so that it can work on all operating systems(Ardiansyah, 2019).

METHOD

The stages in this research include research steps. The framework in this research is described as follows:

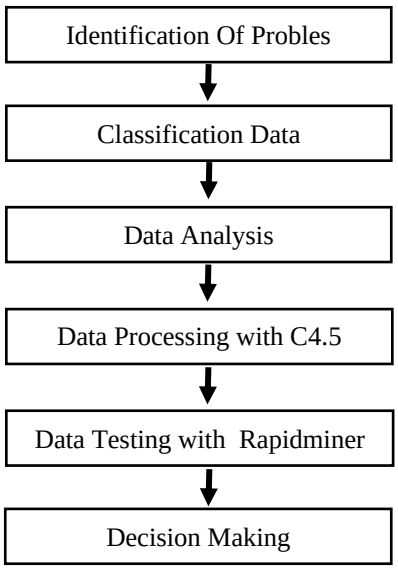

Figure 1: Research Framework

In figure 1. The steps of the research process with the stages of problem identification, data classification, data analysis, data processing with the C4.5 algorithm, data testing with the rapidminer application and decision making.
The data used in this study is data based on the criteria used in the calculation, namely the company used for the calculation of the highest alternative for determining the security guard's work schedule.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12247 p-ISSN : 2541-044X

The method proposed for the process as described above is the classification method with the algorithm used is the C4.5 algorithm with the criteria used as follows:

1. Security Name
2. Age
3. Gender
4. Education
5. Skills
6. Levels
7. Schedule

The existing dataset is then selected for the parameters to be analyzed. The parameters taken are attributes of the previously obtained overtime schedule data and will be used as input or input variables. There are 12 (twelve) of these parameters. Classification of Security Guard Data to Determine Work Schedules Using the Decision Tree Method.

Table 1. Classification of security guard data to determine work schedules

NO	NAMA	AGE	JK	EDUCATION	SKILL	LEVEL	SCHEDULE
1	Sugeng	23	Man	SENIOR HIGH SCHOOL	EXPERIENCE	juniors	A
2	Firman	24	Man	SENIOR HIGH SCHOOL	NON EXPERIENCE	juniors	A
3	Nuryati	29	Woman	SENIOR HIGH SCHOOL	EXPERIENCE	senior	A
4	Ismail	33	Man	SENIOR HIGH SCHOOL	EXPERIENCE	senior	B
5	Wika	25	Woman	JUNIOR HIGH SCHOOL	NON EXPERIENCE	juniors	A
6	Yuli	26	Woman	JUNIOR HIGH SCHOOL	EXPERIENCE	senior	B
7	Dewi	28	Woman	SENIOR HIGH SCHOOL	EXPERIENCE	senior	A
8	Saiful	27	Man	SENIOR HIGH SCHOOL	NON EXPERIENCE	juniors	A
9	Dodo	36	Man	SENIOR HIGH SCHOOL	NON EXPERIENCE	juniors	A
10	Desi	25	Woman	SENIOR HIGH SCHOOL	NON EXPERIENCE	juniors	B
11	Tere	31	Woman	SENIOR HIGH SCHOOL	EXPERIENC	juniors	B
12	Kusniati	24	Woman	JUNIOR HIGH SCHOOL	NON EXPERIENCE	juniors	B

In Table 1. Classification of security guard data to determine work schedules based on attributes of age, gender, education, skills, levels with divided status for seniors and juniors.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12247
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

Based on the data classification, look for entropy and gain calculations:

Table 2. Calculation of Entropy and Gain on Work Schedule

		Amount (S)	A (Si)	B (Si)	Entrophy	Gain
TOTAL		12	7	5	0,979869	
AGE						0,061573
	Age 2	9	6	3	0,918296	
	Age 3	3	1	2	0,918296	
GENDER						0,104349
	Man	5	4	1	0,721928	
	Woman	7	3	4	0,985228	
EDUCATION						0,061573
	SENIOR HIGH SCHOOL	9	6	3	0,918296	
	JUNIOR HIGH SCHOOL	3	1	2	0,918296	
SKILLS						0,020721
	EXPERIENCE	6	3	3	1	
	NON EXPERIENCE	6	4	2	0,918296	
LEVELS						0,010246
	Juniors	8	5	3	0,954434	
	Senior	4	2	2	1	

In Table 2. The results of the Gain calculation, the highest Gain value is obtained, namely in the Gender/JK attribute of 0.104349, the Gender attribute becomes root. The next step is to create a table based on the value or value in the attribute, because the attribute only contains Male and Female, so only 2 tables need to be made

.Table 3. Calculation of Entropy and Gain in Men's Work Schedules

		Amount (S)	A (Si)	B (Si)	Entrophy	Gain
Total		12	7	5	0.979869	
AGE						0.813202
	Age 2	3	3	0	0	
	Age 3	2	1	1	1	
EDUCATION						0.679065
	SENIOR HIGH SCHOOL	5	4	1	0.721928	
	JUNIOR HIGH SCHOOL	0	0	0	0	
SKILLS						0.813202
	EXPERIENCE	2	1	1	1	
	NON P EXPERIENCE	3	3	0	0	
LEVELS						0.979869
	Juniors	4	4	0	0	
	Senior	1	0	0	0	

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12247
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In Table 3. The results of the Gain calculation for the level attribute have the highest value of 0.979869, so the formation of a decision tree is under the Gender attribute. Then proceed with entropy and gain calculations on women's work schedules.

Table 4. Calculation of Entropy and Gain in Women's Work Schedules

		Amount (S)	A (Si)	B (Si)	Entrophy	Gain
Total		12	7	5	0,979869	
AGE						0,979869
	Age 2	9	6	3	0	
	Age 3	3	1	2	0	
EDUCATION						0,646535
	SENIOR HIGH SCHOOL	4	2	2	1	
	JUNIOR HIGH SCHOOL	3	1	2	0	
SKILLS						0,416961
	EXPERIENCE	4	2	2	1	
	NON EXPERIENCE	3	1	2	0,918296	
LEVELS						0,979869
	Juniors	4	1	3	0	
	Senior	3	2	1	0	

Based on Table 4. the results of the gain calculation show that the age attribute and the Level attribute have the highest gain value, namely 0.979869, because the gain calculation for the highest level attribute both on male and female work schedules, the Level attribute is below the Gender attribute. For the Education attribute it becomes a leaf attribute because the value is between 0 and 1, so that based on the calculation results above, it is obtained that those who get schedule A are men while those who get schedule B with high school education and junior level are women, besides that they get a scheduleA. Implementation is done using Rapid Miner software. By generating a decision tree as follows:

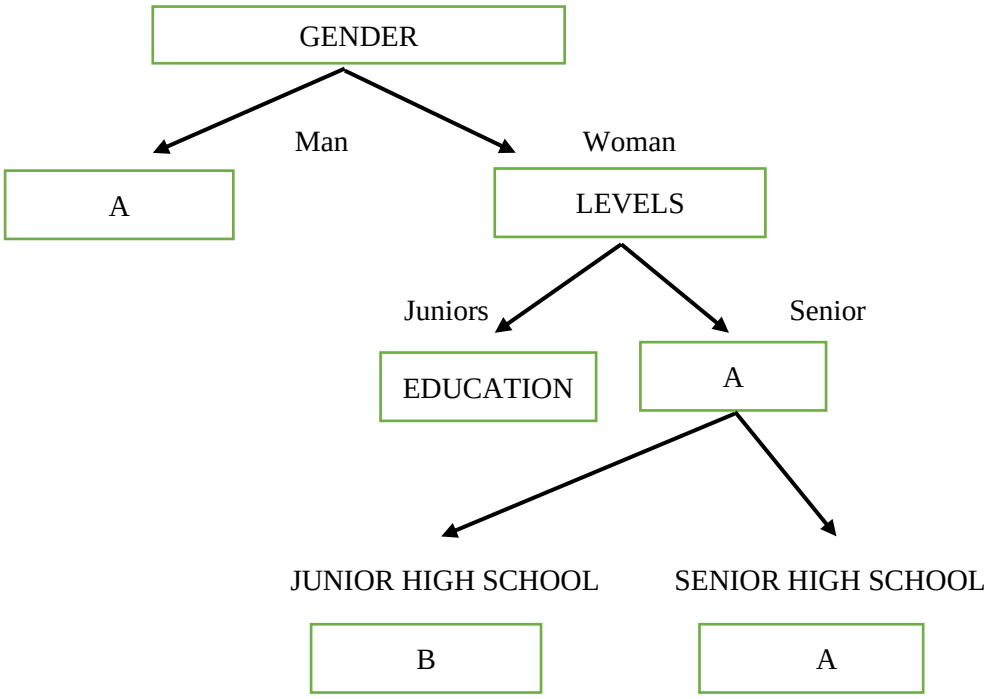

Figure 2. Security guard work schedule decision tree

*name of corresponding author

In Figure 2, it can be seen that the root node of the tree is gender, that is, those who get schedule A are men, while those who get schedule B with high school and junior level education are women, besides that they get schedule A. The accuracy of all classifications is determined by the number of classifications used. correctly divided by the number of classification records by 61.53%.

DISCUSSIONS

The results of the decision tree for determining the security guard's work schedule produce a decision that those who get schedule A are men while those who get schedule B with high school and junior level education are women with the accuracy of all classifications determined by the number of correct classifications divided by the number of classification records by 61.53%. . In research conducted by Triuli Novianti, the relationship between Employee Work Schedules with Age, Gender and Work Unit with the results of testing all data using 5 fold cross validation resulted in a testing accuracy of 70%. So the previous research with the current research with the same object division of labor on employees average accuracy rate of 60% to 70%. Meanwhile, research conducted by Ikhsan Romli, with the same object but different attribute choices, the level of accuracy is higher with an accuracy of 91 %.

.

CONCLUSION

The conclusion of the research results shows that the root node of the tree is gender, that those who get schedule A are men, while those who get schedule B with high school and junior level education are women, besides that they get schedule A. The accuracy of all classifications is determined by the number of correct classifications divided by the number of records classification accuracy of 61.53%.

REFERENCES

Alkhairi, P. (2022). Penerapan Data Mining Untuk Menganalisis Kepuasan Pegawai Terhadap Pelayanan Bidang SDM dengan Algoritma C4.5. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, *7*(1), 40. https://doi.org/10.30645/jurasik.v7i1.414

Anestiviya, V. (2021). Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar). *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, *2*(1), 80–85. Retrieved from http://jim.teknokrat.ac.id/index.php/JTSI

Ansori Aan. (2016). Digitalisasi Ekonomi Syariah. *ISLAMICONOMIC: Jurnal Ekonomi Islam*, *7*(1), 1–18. https://doi.org/10.32678/ijei.v7i1.33

Ardiansyah, D. (2019). Algoritma C4.5 Untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa Smp Dengan Menggunakan Aplikasi Rapid Miner. *Jurnal Inkofar*, *1*(2), 5–12. https://doi.org/10.46846/jurnalinkofar.v1i2.29

Azwanti, N. (2018). Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, *13*(1), 33. https://doi.org/10.30872/jim.v13i1.629

Julianto, P. (2019). Pengaruh Disiplin Kerja Terhadap Prestasi Kerja Pegawai Pada Puskesmas Di Kecamatan Depati Vii Kabupaten Kerinci. *Jurnal Administrasi Nusantara*, *2*(1), 42–58. https://doi.org/10.51279/jan.v2i1.24

Kiu, V. (2021). Penerapan Algoritma C4.5 Untuk Menentukan Reward Karyawan Pada Pt Indoland Batam. *Jurnal Comasie*, *04*(04).

Kurniawan, I. S. (2019). Faktor-faktor yang mempengaruhi loyalitas karyawan. *Jurnal Ekonomi Dan Manajemen KINERJA*, *16*(1), 85–97.

Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, *1*(1), 47–56. https://doi.org/10.24002/ijis.v1i1.1727

Margono, P. (2020). Pengaruh Kompensasi dan Pengembangan Karir terhadap Kinerja Karyawan. *Journal of Economics and Business UBS*, *9*(1), 83–88. https://doi.org/10.52644/joeb.v9i1.36

Novianti, T. (2016). Penentuan Jadwal Kerja Berdasarkan Klasifikasi Data Karyawan Menggunakan Metode Decision Tree C4.5 (Studi Kasus Universitas Muhammadiyah Surabaya). *Jurnal Komunika : Jurnal Komunikasi, Media Dan Informatika*, *5*(1), 1. https://doi.org/10.31504/komunika.v5i1.633

*name of corresponding author

Pratama, T. G. (2020). Penerapan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Pada Algoritma C4.5 Dalam Menentukan Blogger Profesional. *Jurnal Bisnis Digitasl Dan Sistem Informasi*, *1*, 49–55.

Robert Tua Siregar. (2020). *Manajemen Sumber Daya Manusia Dalam Organisasi* (edisi 1, J). Retrieved from https://kitamenulis.id/

Romli, I. (2020). Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *4*(2), 694–702.

Setio, P. B. N. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, *3*, 64–71.

Turnip, S. M. (2018). Analisis Pola Penyebaran Penyakit dengan Menggunakan Algoritma C4.5. *Jurnal Teknik Informatika Unika St. Thomas*, *3*(1), 1–5.

Utami, D. S. (2021). Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM). *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, *1*(1), 299–305.

*name of corresponding author