

Determination of Work Schedule Based on Employee Data Classification Using the Decision Tree Algorithm C4.5 Method

Penentuan Jadwal Kerja Berdasarkan Klasifikasi Data Karyawan Menggunakan Metode Decision Tree Algorithm C4.5

Mas'ud Effendi*, Risna Dyah Ariani, Retno Astuti

¹Department of Agro-Industrial Technology Faculty of Agricultural Technology, Universitas Brawijaya,
Jl Veteran, Malang 65145, Indonesia

*mas.ud@ub.ac.id

Received: 9th July, 2021; 1st Revision: 15th October, 2021; 2nd Revision: 26th November, 2021; Accepted: 14th December, 2021

Abstract

This study aims to create a work shift scheduling system based on data classification, as well as to determine its level of accuracy and provide schedule recommendations. The method used was the Decision Tree Algorithm C4.5 which functions as a classification system to form a work shift schedule. The study included 128 employees and a total of 43 training data were obtained from a 1/3 split of the dataset, then it was processed using RapidMiner 5.3 data mining software. Furthermore, the rule of decision tree calculation results was used to classify employee and shift formation on the web system based on PHP and MySQL. The attributes of the decision-maker consist of gender, health records, age, and work unit. The classification found 96 employees who occupy the afternoon shift and 32 for the night shift. System testing was carried out using K-fold cross-validation, which produced an average accuracy of 93.39%, with the highest found in the six-fold cross-validation of 95.35%. Two-shift scheduling was proposed on the web system with a work shift rotation in form of a metropolitan plan (2-2-2 rota).

Keywords: scheduling, Decision Tree C4.5, employee

Abstrak

Penelitian ini bertujuan untuk membuat sistem penjadwalan shift kerja berdasarkan klasifikasi data, mengetahui tingkat akurasi sistem penjadwalan shift kerja berdasarkan hasil klasifikasi, serta memberikan usulan jadwal shift kerja berdasarkan klasifikasi dari sistem penjadwalan kerja berbasis web. Metode yang digunakan adalah Decision Tree Algorithm C4.5 yang berfungsi sebagai pengklasifikasi untuk pembentukan penjadwalan shift kerja. Kebutuhan data dalam penelitian ini berjumlah 128 data tenaga kerja. Data training diperoleh melalui split 1/3 data tenaga kerja (43 data). Data training diolah menggunakan software data mining RapidMiner 5.3. Rule hasil perhitungan decision tree digunakan sebagai dasar klasifikasi tenaga kerja dan pembentukan shift pada sistem web berbasis PHP dan MySQL. Atribut pengambil keputusan terdiri dari atribut jenis kelamin, catatan kesehatan, usia, dan unit kerja. Hasil klasifikasi diperoleh 96 tenaga kerja yang menempati shift siang dan 32 tenaga kerja yang menempati shift malam. Pengujian sistem menggunakan K-fold cross validation menghasilkan rata-rata akurasi sebesar 93.39% dengan akurasi tertinggi pada six-fold cross validation sebesar 95.35%. Usulan rekayasa pada sistem web terdiri dari 2 shift kerja. Rotasi shift kerja mengadopsi tipe rotasi metropolitan plan (2-2-2 rota).

Kata kunci: penjadwalan, Decision Tree C4.5, tenaga kerja

INTRODUCTION

A growing company needs to continuously increase its production capacity to meet market demand. Meanwhile, increased production capacity must be supported by sufficient employee numbers to meet the productivity targets. The growing number of employees and the limited machine capacity require companies to make proper work scheduling arrangements to

effectively and efficiently carry out delegated works.

Consequently, the company needs to schedule work by considering the factors that affect employees' performance. According to Fajarwati et al. (2011), employee performance is influenced by psychological, environmental, social, and physical factors. These factors are then used to form attributes in the work shift schedule. The attributes used are usually age, distance from home, marital status, and health records.

Psychological and physical factors significantly affect work safety, while environmental and social factors affect employees' convenience in carrying out their jobs. Ukhisia et al. (2013) stated that occupational health and safety play an important role in increasing employee productivity. The variables that affect occupational health and safety consist of the availability of personal protective equipment, workload, communication, and health condition. The work and social environment also significantly influence performance hence, these factors and variables need to be considered in scheduling work.

Schedule adjustment for the workforce causes numerous problems when conventionally carried out without using a computerized system. Classifications and schedules according to the criteria of the workforce tend to be difficult when the number of employees is large and continues to grow. In addition, resetting when there is a change in work schedule takes a long time. Constraints in preparing work schedules can be overcome by using a web-based application to classify employee criteria. Classification is made based on the employee's age, position, and the company shift system. According to Kusriani (2006), classification is a process in artificial intelligence that declares an object as one of the predefined categories

The methods commonly used in the employees' classification are the Genetic, Naïve Bayes, and the Iterative Dichotomize 3 (ID3) algorithm. These three methods have weaknesses in estimating incomplete attributes data, hence, the efficiency value cannot be maximized. Agrawal & Gupta (2013) compared the ID3 method with the C4.5 algorithm, the results showed that C4.5 can increase efficiency in the decision-making algorithm because it has a reasonably high-efficiency value compared to other methods. Besides, this method has advantages in processing continuous and discrete numerical data, understands missing attribute values, produces rules that are easy to interpret, and is the fastest algorithm to execute compared to others (Nofriansyah, 2014). The C4.5 algorithm allows classification without using all of the attributes because it reduces unused features by making predictions on the classified data.

A previous study related to scheduling using the decision tree algorithm C4.5 method was carried out by Achmad & Slamet (2012), who created an application that can produce a reasonable work schedule according to the

employee's criteria. Romli & Zy (2020) also used the C4.5 algorithm to determine employee overtime schedules, and produced an accuracy value of 91%. Furthermore, Soriano et al. (2020) introduced an integrated employee scheduling problem that considers various real-life issues such as varying demands, different working conditions, and individual preferences. Novianti & Santosa (2016) also produced a decision tree with the highest validation test results using a 5-fold cross-validation of 70%. However, they indicated that the decision tree model produced is not accurate enough to be used as a classification basis because there are still data whose decisions are biased as well as attributes that are still asked more than once. The accuracy of this classification can be improved by pruning or deleting a dataset and replacing with new supporting data.

This study aims to implement the decision tree algorithm C4.5 method to create a work shift scheduling system and determine its accuracy based on the results of employee classification. The proposed schedule will be used to form a web-based work scheduling system based on workforce grouping using the C4.5 decision tree algorithm method. Furthermore, this study is expected to produce a tree with a validation value close to 100% through the use of pruning.

METHODS

This study was conducted at CV Sumber Horti Nasional (SHN), located at Jalan Sikatan Number 02, Mangiran Hamlet, Lamong Village, Badas District, Kediri Regency, East Java, Indonesia. The company produces self-breeding horticultural seeds, and for the last four years, the sales data shows an increase in sales, demand, and marketing areas. This condition is a challenge because the company currently is still constrained by production capacity and employees, hence, the company wants to add new employees and enforce work shifts to increase production capacity and meet consumer needs. The re-arrangement of the work schedule was carried out using an application that can produce an employee classification, while a web-based software was developed to classify the criteria for the workforce in the company.

The work schedule re-arrangement was completed in production units that apply work shifts. The inputs needed for scheduling were employee data, including name, gender, date of birth (age), status, distance from home to the

company, medical history records, and workstation. This study was only carried out to prepare a workforce schedule with a web application system as a proposal to the company, while calculation of productivity and efficiency of the work schedule was not conducted.

The hardware used in this study was a personal computer, while the software used includes Microsoft Windows 10, RapidMiner 5.3, HTML, MySQL, PHP, and XAMPP. The data used are 128 employees and other data shown in Table 1.

Establishment of Classification System

The system's formation was achieved using RapidMiner 5.3 which is a data mining application platform used for data preparation, training, text mining, and predictive analysis. The system design was executed in 2 stages, namely tree formation and system testing. Tree construction was performed by entering some of the training data from employee master and work schedule data from work units. Selection and transformation were then carried out on the employee master data to adjust the attributes used in the construction of the classification system and to simplify the mining process. The training data was obtained from a 1/3 split of the dataset using random sampling, hence, a total of 43 were obtained from 128 employees' data. Meanwhile, random sampling was performed using Microsoft Excel software to make the selected attributes represent the entire data. Labeling was further employed on the training data formed as a basis for classification in tree construction. This was conducted based on the work shift placements arrangement consisting of day and night shifts. The training data from the sampling was processed using the decision tree algorithm C4.5 method, then a validity test using K-fold Cross Validation was implemented on the results. The tree construction process and testing of classification results were performed with the RapidMiner 5.3

software. The data classification results were tested five times in the 10-fold area, namely 6-fold, 8-fold, 10-fold, 12-fold, and 14-fold. The average value of the accuracy results on each test data was then calculated to determine the level of validity for the tree formed. According to Febriana et al. (2018), validation testing uses several K-fold values around 10-fold cross-validation, and an average calculation is needed to ensure the data conditions are in good condition.

Decision Tree Algorithm C4.5

A decision tree is a tree-shaped flowchart structure that can divide an extensive data set into smaller record sets by applying decision rules with each set of division members becoming similar. The decision tree divides the attributes into nodes to be classified according to the classification label that has been selected (Agrawal & Gupta, 2013). The formation of the decision tree algorithm C4.5 has two main stages, namely the root search and the branch creation process. The root search process is carried out by calculating each attribute of the training data that has been labeled.

The entropy value is initially calculated followed by the gain value calculation which provides the information for each attribute. This information sets the initial root based on the highest gain value. The calculation of entropy and gain is then carried out again at the lower gain value, this recalculation is conducted repeatedly until the last node. The highest gain value category is used as node 1.1 then, the next node is formed until the calculation is complete for all attributes. This process is performed until no more branch nodes are left without a decision or have a gain value of 0. The entropy and gain values are calculated using formulas (1) and (2), while the decision tree algorithm C4.5 training process is shown in Figure 1. Meanwhile, the process of calculating and creating the root is demonstrated in Figures 2, 3, and 4.

Table 1. Study data requirements

No	Data Requirement	Data Source	Method	Data Use
1.	Employee master data which includes name, gender, place and date of birth, status, distance from home to the company, health records	Company data	Observation and documentation	Data for setting work shifts and forming training data
2.	Work shift condition data	Company data and literature	Observation and literature study	Data for consideration of work shift scheduling formation

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i \times \log_2 p_i \quad (1)$$

where,

S = case set

n = number of partitions S

p_i = the proportion of S_i to S where S_i is the case set on the i-th partition

$$\text{Gain (S, A)} = \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Entropy (S}_i) \quad (2)$$

where,

S = case set

A = attribute

n = number of attribute partitions A

$|S_i|$ = number of cases on partition i

$|S|$ = number of cases in S

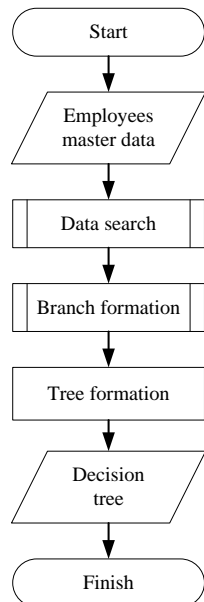


Figure 1. Process Flowchart of Training Decision Tree Algorithm C4.5

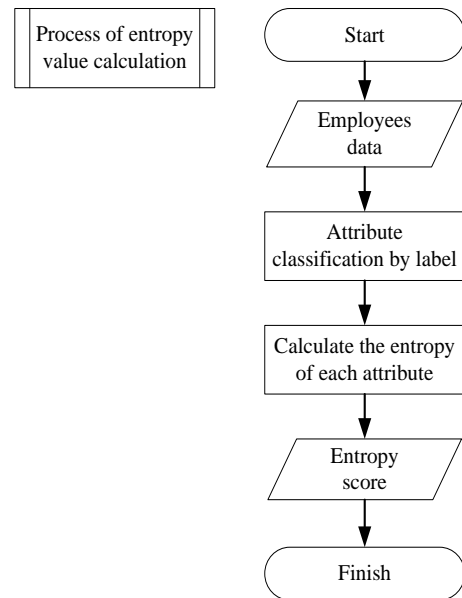


Figure 2. Entropy Value Calculation Training Process

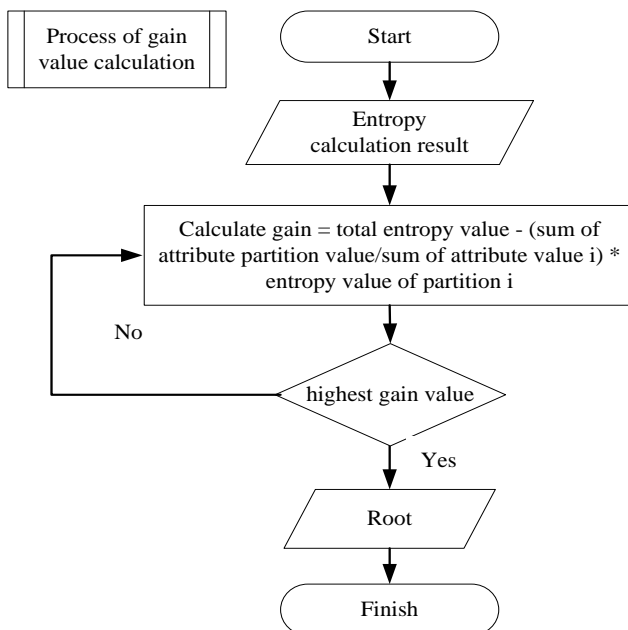


Figure 3. Gain Value Calculation Training Process

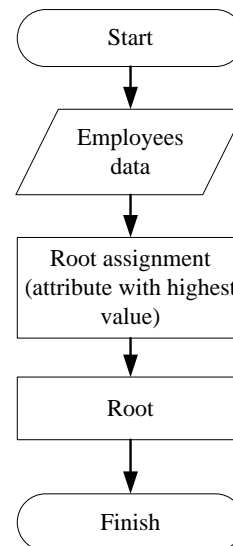


Figure 4. Flowchart of Root Formation Process

System Interface Creation

The system interface development was carried out by making a work shift scheduling application based on the results of the employee classification using the C4.5 decision tree algorithm. The web-based application created has an interface design that functions as a liaison between the system and the user. The system interface includes the home page, employee data, work schedule, tree, and tests. Furthermore, the software operating system used was Windows 10 with the PHP programming language, while the tools used were notepad++ and the Apache MySQL webserver, also display on the web used HTML. Tree construction and algorithm implementation in the system was performed by manual calculation based on rules formed from running data using RapidMiner 5.3. The expected output is data classification as a reference for companies to make work shift scheduling arrangements and recommendations.

RESULTS AND DISCUSSION

System Planning

Design is the stage used to clarify the need for an employee data classification system as the basis for scheduling work shifts. The requirements

include the software and design of an employee data classification system using the C4.5 algorithm. The process performed in this design is shown in the flow chart (Figure 5).

Software Requirements Analysis

This is used to identify the actors involved in the system and functional requirements. Actors involved in the system are shown in Table 2, the functional requirements analysis is a step taken to explain the system's needs. This analysis describes the name of the process used to indicate the function of each system requirement. The list of functional requirements is presented in Table 3.

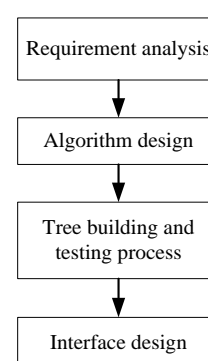


Figure 5. Design Flow Chart

Table 2. Identify actors in the system

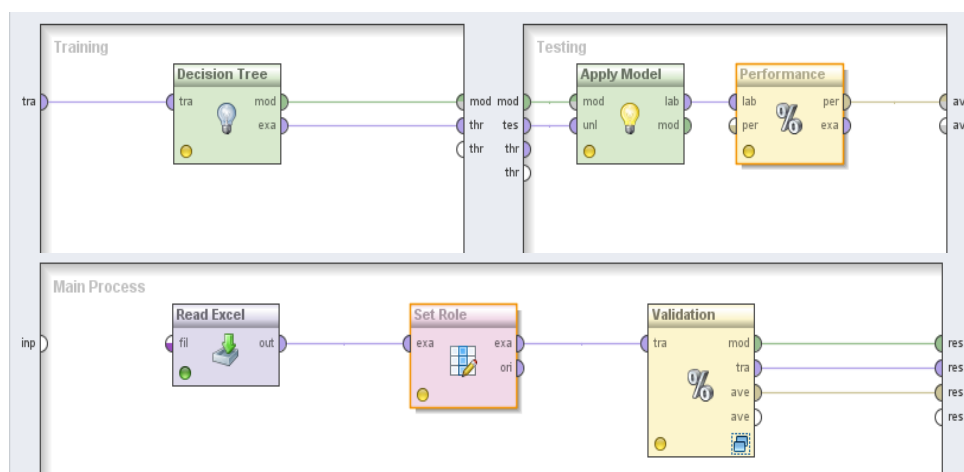
No	Actor	Actor Information	Description
1	User	Users are actors who access and use the system fully.	Users are employees who can access work schedules.
2	Admin	Admin is an actor who has the authority to fill in data, change data, and set up web applications. Admin is divided into two types with different ids, namely "superadmin" and "developer".	Admin is a person under the auspices of the Human Resource Development (HRD) and Top Management divisions. Admin with id "superadmin" is in charge of managing employee data, compiling work schedules. "Superadmin" can access all data. The admin with the "developer" id is in charge of making changes to the settings on the web application.

Table 3. Software specification functional requirements

ID	Functional Needs	Actor	Data Flow
SDK_001	The system provides an interface to edit data	Admin	Edit data
SDK_002	The system provides an interface for deleting data	Admin	Clear data
SDK_003	The system provides an interface for displaying detailed data	User, Admin	Data details
SDK_004	The system provides an interface to input data	User, Admin	Data entry
SDK_005	The system provides an interface for making data selections to determine the desired results	User, Admin	Select data
SDK_006	The system provides an interface to view the classification results from the C4.5 algorithm calculation	User, Admin	Determination result
SDK_007	The system provides an interface to view work shift schedule recommendations based on the classification C4.5 algorithm results	User Admin	Results of work schedule recommendations

Table 4. List of system attributes

No	Attribute	Information	Description	Domain	Value
1	Age	Classification Basis	Employee age	Integer	20-50
2	Gender	Classification Basis	Employee gender	Binomial	- Male (L) - Female (F)
3	Marital status	Classification Basis	Employee marital status	Binomial	- Single - Married
4	Medical history	Classification Basis	Employee medical history	Binomial	- Good - Poor
5	Work unit	Classification Basis	Work units occupied by employees	Polynomial	Top management Research and Development (R&D) Finance and Accounting Production Process Quality Human Resource Development (HRD) Marketing Warehousing Security
6	Distance from home	Classification Basis	Distance from home to company	Integer	1. (≤ 5 km) 2. (5.1 km-10 km) 3. (10.1 km-15 km) 4. (>15 km)
7	Shift work	Label	Shift work employees at the company	Binomial	- Day - Night

**Figure 6.** Validation Model Design of C4.5 Algorithm Decision Tree

Algorithm Design

The algorithm design process begins by analyzing the knowledge base which contains rules of the attributes used in the system. Each attribute in the workforce data is classified using the decision tree algorithm C4.5, the attributes used are shown in Table 4.

Tree Construction

Tree construction was performed using RapidMiner 5.3 software by designing the model and then testing the validity of the results. The

input to the model design is the labeled training data, while the attributes and labels that have been inputted are then set to separate and ensure that the classified data form the basis for classification in a tree formation. Data classified as an attribute and become the basis for classification is then marked with a label and used for the work shift schedule. Subsequently, the model is set to run the decision tree algorithm C4.5, the settings made are shown in Figure 6, and the parameter settings are demonstrated in Figure 7. The cross-validation test was carried out five times in 6-fold, 8-fold, 10-

fold, 12-fold, and 14-fold, while the model validation test results are presented in Table 5.

The highest validation level with the lowest standard deviation value was selected based on K-fold Cross-Validation results. According to Susanto et al. (2015), the standard deviation shows the level of strength of the model formed. The lower the dependence level of the model's performance on the test data set, the smaller the standard deviation. The test results on the 6-fold cross-validation were chosen in this study as the basis for tree formation. Details of accuracy testing on the 6-fold cross-validation are shown in Table 6. The tree formed in this study is demonstrated in Figure 8.

The accuracy value obtained cannot be 100% because not all attributes are included in the decision-making in the formed tree. The attributes that become nodes in the tree consist of age, gender, health records, and work units. Meanwhile, the marital status and home distance are not included because they have been selected with other attributes. These attributes have predictable decisions, hence, there is no need for decision-making in the C4.5 decision tree algorithm.

Interface Design

System development requires tools in the form of hardware and software, the software used in the interface development is shown in Table 7, while the hardware is shown in Table 8.

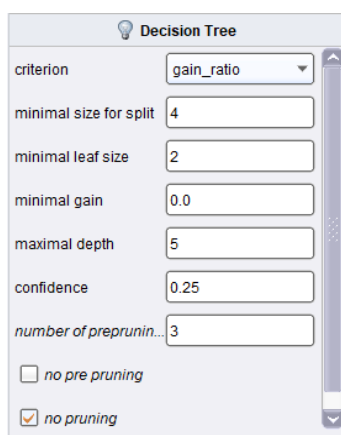


Figure 7. Parameter Settings on RapidMiner 5.3

Table 5. Validation test results with k-fold cross-validation

K-Fold	Accuracy	Standard Deviation
6	95.24%	± 6.73%
8	90.83%	± 13.62%
10	90.50%	± 16.19%
12	95.14%	± 11.00%
14	95.24%	± 11.66%
Average	93.39%	

Table 6. Model accuracy values with six-fold cross-validation

Accuracy: 95.24% ± 6.73% (micro: 95.35%)			
	True Day	True Night	Class Precision
Day Prediction	28	1	95.55%
Night Prediction	1	13	92.85%
Class Recall	95.55%	92.85%	-

Table 7. System development hardware requirements

Name	Specification
Processor	Intel(R) Core™i3-2365U CPU @ 1.40GHz
RAM	2.00 GB
Sistem Type	64-bit operating system, x64- based processor

Table 8. System development software requirements

Name	Specification
Operating system	Windows10
Programming language	PHP
Tools	Notepad++
Web server	Apache MySQL
Appearance	HTML

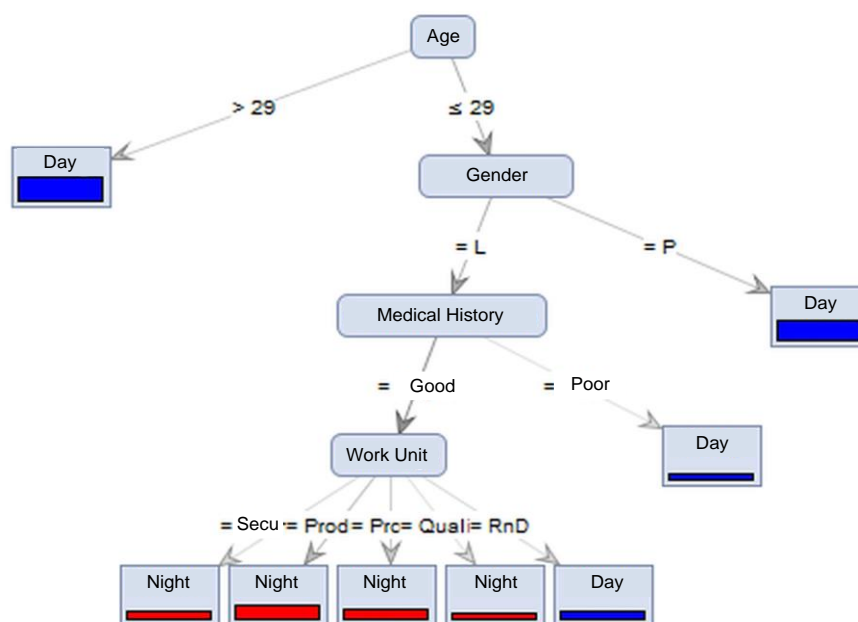


Figure 8. Running Results Tree on RapidMiner 5.3

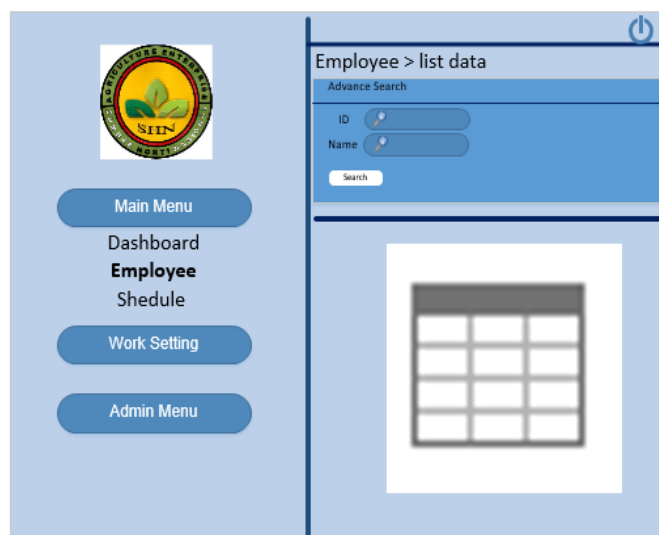


Figure 9. Interface Design Example

The system interface displays the web design, it consists of the system start and homepage or main menu, as well as the work settings, and the admin menu page. The start page contains the login section or how to enter the system, while the home page contains menus on the system in the main menu, tree, work settings, and admin menu page. Furthermore, the main menu page consists of a dashboard, an employee, and a schedule page, the dashboard contains information on the recapitulation of the employee based on each data criteria, such as the percentage of health status and the number of employees. The employee page consists of data on the entire workforce which the admin can only input, while the schedule page contains a recapitulation of work shift rotation and all data for each employee. The work settings page displays three subpages namely the work unit, the work shift, and the decision tree page. The work unit page displays data for the company's work units, the work shift page shows the current work shift and the decision tree page displays the results based on the decision tree algorithm. Moreover, the admin menu page contains settings for accessing the web system, an example of the interface design is presented in Figure 9.

Implementation of the Web Application System Interface at CV SHN

Web access starts with the login page, login can be done by entering the email address and password that has been registered in the system. System users are distinguished as admin and user, they both enter the home page's main menu after the login is successful, this page contains the menu

offered by the web. The system processing results on the dashboard page show that CV SHN currently has 128 employees consisting of 95 men and 33 women with 88 married status. Furthermore, the health level at the company is relatively good, with 95% of the total workforce having a good medical history. The employee page contains the inputted data and work shifts classification results from the system based on the processing results from RapidMiner 5.3. The classification shows that 32 employees are classified in night shifts and 96 are in day shifts.

The work settings page contains information integrated from the main menu, namely the dashboard and employee page. Meanwhile, the work unit page displays the units in CV SHN, namely Top Management, HRD, Production, Process, Security, Warehousing, Marketing, Finance and Accounting, R&D, and Quality. The work shift page contains the respective shifts applied to the company, namely day and night, while the decision tree page shows the results of employee data calculations using the C4.5 algorithm. The tree results are used as the basis for employee classification for the work shift system. The predictive C4.5 decision tree algorithm application on the web system is used to predict whether the data input on the employee page is incomplete to directly determine which workforce can enter the day or night shift.

The admin menu page consists of the admin list, groups, and the Content Management System (CMS) menu. The entire menu on this page can only be accessed by the admin using the super admin id. General users can only access the admin

list page for changes to user information, this page contains a list of ids or registered users on the system that can enter the website. Furthermore, the web user data display contains number information, user profile photo, email, group criteria, data change date, status, and activation id information. The admin group's page is for setting up admins on a web system, while the CMS page is used for system management. CMS is a content-oriented web-based application, according to Siambaton & Fakhriza (2016), CMS is added to web applications to make it easier for website admins to arrange and update the web from anywhere without contacting the webmaster. Website management can be done by managing content, categories, and users, an example of the interface implementation is shown in Figure 10.

Employee Schedule Analysis

A company carries out a work shift schedule to meet targets and maintain production continuity, the most common arrangements are two and three work shifts. The company where this study was conducted plans to adjust workforce scheduling to increase production capacity without violating applicable laws and regulations. According to Syahputri et al. (2017), the current work scheduling arrangements imposed by the company are less suitable for providing proportionality in days off and working time. Scheduling is sometimes arranged contrary to the applicable employees' laws in Indonesia. This condition promotes companies to advocate healthy work shift schedules. The company pays

more attention to night shifts because it potentially causes fatigue problems that lead to work accidents.

Work Shift Scheduling Settings

The company considers several aspects in arranging work shifts, these aspects are adopted into the classification system to avoid negative outcomes in relation to employee regulations. CV SHN considers work shift schedule preparation in line with the Decree of Manpower Minister of the Republic of Indonesia No. 102/MEN/IV/2004 and RI Law No. 13 of 2003 concerning employment. Some other considered aspects are:

1. Work shifts should be arranged in a forward rotation pattern with a period fewer than two weeks, and a day-off average of two days/week. The work shift duration must be set to a maximum of 8 hours per day and when an employee is forced to do additional hours, the workload must be reduced.
2. When it is quite risky to carry out the night shift, then the work should be done before 4 AM.
3. Demographic aspects, such as gender, age, and distance from home, must be considered in preparing work shift schedules.
4. Employees with gastric disease and unstable emotions are not to be placed on night shifts. Abdominal disease and other internal organs can be triggered due to excessive stress levels, making it dangerous to work at night (Susetyo et al., 2012).

The screenshot shows a web application interface for managing employees. The page title is "Employees" and it displays a list of 3 employees. The interface includes a sidebar with navigation options like "Dashboard", "Employee", "Work Unit", "Work Shift", and "Decision Tree". The main content area has an "Advance Search" section with fields for "ID" and "Name", and a table listing employee details.

No	Name	Age	Gender	Marital Status	Health Record	House	Work Unit	Work Shift	Actions
51	Lucky Andriawan	29	Male	Single	Good	2 Km	RnD	Malam	[Edit] [Delete]
52	Sandi Wibowo	29	Male	Married	Good	2 Km	RnD	Malam	[Edit] [Delete]
53	Anton Setiawan	27	Male	Single	Good	3 Km	RnD	Malam	[Edit] [Delete]

Figure 10. Example of Interface Implementation

Table 9. Classification results recapitulation of the two work shifts application

Division	Number of Employees	
	Day Shift	Night Shift
Security	2	2
Production	22	14
Process	4	5
Quality Control	9	6
Warehousing	10	5

Proposed Shift Work Schedule

CV SHN plans to increase its workforce and apply two work shifts namely day and night with a period of 8 hours each. The data from work shift classification using decision tree algorithm C4.5 is set to apply rotation through the web-based application. The setting carried out provides a proposal to CV SHN to increase the number of employees. This addition is aimed at maintaining the work formation in each section. Meanwhile, the web application offers two options for implementing work shifts namely two and three.

The two-shift work arrangement system offered at CV SHN has a work duration of 8 hours each. The employees rotate shifts that require a long time, with the morning shift starting at 8 AM - 4 PM and the night from 8 PM - 4 AM with a break duration of 1 hour. The break schedule for the day shift starts from 12 PM - 1 PM, while the night begins from 12 AM - 1 AM. Moreover, employees get two days off on weekends to implement the two-shift work system. The days off are expected to balance work time with rest to avoid fatigue, which leads to work accidents. The non-shift or permanent work system on the afternoon shift is proposed to be applied to the Top Management, Marketing, HRD, finance and accounting, as well as the R&D sections, while the proposal for two shifts was given to the Production, Process, Warehousing, Quality, and Security divisions. The classification results show that 96 employees are scheduled on the day shift in the divisions affected, while 32 are scheduled on the night shift, as shown in Table 9.

The suggestion of 3 work shifts was given because the schedule of 2 was only applicable to a long shift pattern. Long rotations on a 2-shift schedule tend to burden the workforce because of the workload and long working time. The long shift pattern is not recommended because it triggers fatigue, which leads to work accidents. In the proposed three shifts, the morning shift starts at 8 AM - 4 PM, the afternoon from 4 PM - 12 AM and the night from 12 AM - 8 AM. The break schedule for the morning shift starts at 12 PM - 1

PM, the afternoon at 6 PM - 7 PM, and the night at 4 AM - 5 AM. Furthermore, the metropolitan plan rotation pattern (2-2-2 rota) was chosen to shorten the adaptation level of employees when they have to change work shift schedules from night to morning. This fast rotating shift pattern uses four teams and three 8 hour shifts to provide 24/7 coverage. Each team rotates through a sequence of 2-day shifts, two swing shifts, 2-night shifts, and two days off over an 8-day cycle (Roberts, 2016). The advantage of this metropolitan rotation plan is that consecutive night shifts only occur for a maximum of two times and the workforce has two consecutive days off in one rotation cycle (8 days). This work shift scheduling is applied similarly to divisions that use a two-shift arrangement. However, CV SHN needs to add to the number of employees on the night shift in the proposed two work shifts. The addition can be done by recruiting 32 employees, 2 for the Security division, 14 for Production division, 5 to the Process Division, and 6 for the Quality division. The company can consider the proposed application of three work shifts when all aspects that support development have been met.

CONCLUSIONS

Work shift scheduling can be carried out by classifying employee data using the C4.5 decision tree method. The work schedule obtained from the classification results of 128 employees at CV SHN was 32 occupying the night shift and 96 occupying the day shift. Employees who run two work shifts are the Security, Warehousing, Production, Process, and Quality divisions, while others in Top Management, R&D, HRD, Marketing, as well as Finance and Accounting divisions only occupy the afternoon shift. The accuracy test conducted using K-fold cross-validation shows that the average accuracy is 93.39% with the highest obtained at 6-fold with 95.35%. Furthermore, the web system accessible with the address shn.redonesia.com displays employee data along with the rotation of their work shifts, also it displays settings for two

and three work shifts. The web system shows that the company needs to add 32 new employees to enforce three work shifts including 2 for Security, 5 at Warehouse, 5 for Process, 14 at Production, and 6 at the Quality division.

The process of calculating and forming work shift schedules on the web system still requires updates following technological developments and changes in company policies. The accuracy calculation results still have not reached 100%. Meanwhile, higher results close to 100% can be obtained by considering the data type from the attributes used. Further studies are recommended to use other data mining software as a comparison. It is also recommended that productivity on the results of scheduling in the web system be calculated to identify the need for adding new employees for the implementation of the proposed work shift.

References

- Achmad, B. D. M., & Slamet, F. (2012). Klasifikasi data karyawan untuk menentukan jadwal kerja menggunakan metode decision tree. *Jurnal Iptek*, 16(1), 17–23.
- Agrawal, G. L., & Gupta, H. (2013). Optimization of C4.5 decision tree algorithm for data mining application. *International Journal of Emerging Technology and Advanced Engineering*, 3(3), 341–345.
- Fajarwati, F. D., Hidayat, R., & Agustina, F. (2011). Pengaturan sistem shift kerja untuk meningkatkan performance serta mengurangi keluhan karyawan. *Jurnal Teknologi Technoscintia*, 4(1), 37–45.
- Febriana, L., Furqon, M. T., & Rahayudi, B. (2018). Klasifikasi penyakit typhoid fever (tf) dan dengue haemorrhagic fever (dhf) dengan menerapkan algoritma decision tree C4.5 (Studi kasus: Rumah Sakit Wilujeng Kediri). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(3), 1275–1282.
- Kusrini. (2006). *Sistem Pakar Teori dan Aplikasi*. Yogyakarta: Andi.
- Nofriansyah, D. (2014). *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Sleman: Deepublish.
- Novianti, T., & Santosa, I. (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), 1–10. <https://doi.org/10.31504/komunika.v5i1.633>
- Roberts, C. (2016). *Psychology for Cambridge International AS and A Level Revision Guide* (2nd ed.). Oxford: Oxford University Press.
- Romli, I., & Zy, A. T. (2020). Penentuan jadwal overtime dengan klasifikasi data karyawan menggunakan algoritma C4.5. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 4(2), 694–702.
- Siambaton, M. Z., & Fakhriza, M. (2016). Aplikasi content management system (cms) pada Joomla untuk membuat web service. *InfoTekJar : Jurnal Nasional Informatika Dan Teknologi Jaringan*, 1(1), 11–13.
- Soriano, J., Jalao, E. R., & Martinez, I. A. (2020). Integrated employee scheduling with known employee demand, including breaks, overtime, and employee preferences. *Journal of Industrial Engineering and Management*, 13(3), 451–463. <https://doi.org/10.3926/jiem.3126>
- Susanto, Mulyani, E. D. S., & Nurhasanah, I. R. (2015). Penerapan data mining classification untuk prediksi perilaku pola pembelian terhadap waktu transaksi menggunakan metode Naïve Bayes. In *Konferensi Nasional Sistem & Informatika 2015* (pp. 313–318). Bali.
- Susetyo, J., Oesman, T. I., & Sudharman, S. T. (2012). Pengaruh shift kerja terhadap kelelahan karyawan dengan metode bourdon wiersma dan 30 items of rating scale. *Jurnal Teknologi*, 5(1), 32–39.
- Syahputri, K., Leviza, J., & Devi, T. K. (2017). Penjadwalan tenaga kerja untuk menentukan regular days off (rdos) dengan menggunakan algoritma Monroe. *Media Teknika : Majalah Teknologi Universitas Sanata Dharma Yogyakarta*, 12(1), 10–16.
- Ukhisia, B. G., Astuti, R., & Hidayat, A. (2013). Analisis pengaruh keselamatan dan kesehatan kerja terhadap produktivitas karyawan dengan metode partial least squares. *Jurnal Teknologi Pertanian*, 14(2), 95–104.