Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11796
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Decision Tree Algorithm to Measure Employee Performance Discipline

Linda Marlinda[1)*], Evita Fitri[2)], Siti Nurhasanah Nugraha[3)], Faruq Aziz[4)], Santoso Setiawan[5)]
[1,2,3,4,5)] Universitas Nusa Mandiri, Jakarta
[1)] linda.ldm@nusamandiri.ac.id , [2)] evita.etv@nusamandiri.ac.id , [3)] siti.nhg@nusamandiri.ac.id ,
[4)]faruq.fqs@nusamandiri.ac.id, [5)] santoso.sts@nusamandiri.ac.id

Submitted : Sep 19, 2022 | **Accepted** : Sep 30, 2022 | **Published** : Oct 3, 2022

Abstract: Performance appraisal is done to measure the performance of an employee on the work done. The company conducts performance appraisals on employees at least every six months, involving all employees. This study uses the Absenteeism_at_work dataset. The purpose of this research is to analyze the performance of the Decision Tree algorithm in the classification process. Classification will be grouped into two, namely: disciplined and undisciplined The classification process will be carried out using K-Nime. Algorithm performance measurement using Knime Analytics Platform is open-source software for creating data science models. Knime builds data understanding and designs data science workflows and reusable components using accuracy, recall, and precision parameters. From the research conducted, the results of the Decision Tree algorithm have an accuracy rate of 94.6% while the label No. 5.4%. Based on the nineteen attributes proposed, it can be concluded that the Decision Tree algorithm has better performance.

Keywords: Data Mining, Decision Tree Algorithm, Employee Performance, Knime

INTRODUCTION

Human resources are a very important factor in an organization or company. The more employees a company has, the heavier the burden on the Human Resources Division (HR) in carrying out the task of managing employees. To measure employee performance in the company, the HR Division has conducted a performance assessment of each employee at least every six months, but the results of the assessment have not been grouped according to certain classes. To facilitate the HR Division, in monitoring the results of employee performance appraisals that have been carried out, a data grouping method is needed to classify employee performance appraisal data according to their level of performance(Ozdemir, 2020) .

The measurement of employee performance has been carried out out by many companies in order to increase the level of discipline based on several assessment criteria. Measurement of employee performance can stimulate initiatives in improving operational performance, especially the level of employee discipline in daily work activities to achieve superior performance. The level of employee discipline can be seen from attendance by using several methods that can be trusted and carried out responsibly by employees(Wahid, 2019). Employee measurement can be carried out by the management, then evaluated, using an optimization method that is carried out from time to time every month.

One method of grouping data is classification, classification is a method of grouping data, by grouping data that have similar characteristics between one data and another. From several algorithms in the classification, the researcher will use the decision tree algorithm as an algorithm that will be compared in the process of grouping employee performance data(Oliveira & Fran, 2019). Researchers will use nineteen attributes that are used as parameters to measure the performance of the algorithm, including reasons for not being present, the month of absence, day of the week, season, transportation costs, distance from the place of residence to work, service time, age, average workload. -average/day, hit the target, education, child, social, weight, height, body mass index, time absent in hours. With the many parameters used to measure the performance of the algorithm, it is hoped that it can provide more accurate results when compared to using only a few parameters.

The decision tree method or decision tree is a classification method that has been widely used for solving classification problems. Decision tree classification provides a fast and effective method for classifying datasets. Even the development of the decision tree method has been carried out to produce an approach that can classify sensitive data. The decision tree method approach has been proven to be applicable in various fields of life(Jijo & Abdulazeez, 2021) (Achmad & Fauzi, 2012)

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11796
e-ISSN : 2541-2019
p-ISSN : 2541-044X

LITERATURE REVIEW

Decision tree algorithm

The decision tree is a data structure consisting of nodes and edges of nodes in a tree divided into three, namely root nodes, branch nodes, and leaf nodes. In a decision tree, each node represents an attribute that has been tested, each branch is a distribution of test results, and a leaf node represents a particular class group. The top-level node of a decision tree is the root node, which is usually the attribute that has the greatest influence on a particular class(Informasi, n.d.). In general, the decision tree performs a top-down search strategy for the solution. In the process of classifying unknown data, attribute values will be tested by tracing the path from the root node to the end node (leaf) and then predicting the class that is owned by certain new data. The process in the decision tree is changing the shape of the data (table) into a tree model (tree) and then changing the tree model into a rule (rule). The decision tree is a set of IF-THEN rules. The initial stage is testing the root node which is marked with the symbol Q (Question), if the root node test or fulfills the requirements to choose which is symbolized by Y (Yes) and N (No) then produces something then the testing process is also carried out on each branch based on the results of testing. This also applies to internal nodes where a new test condition will be applied to the next leaf node (Raman, Kaliappen, & Suan, 2020)

Employee performance

Employee performance can be seen from several criteria such as achievement, and discipline that will support employee careers in the company. Company management can assess the level of professionalism of employees who are responsible and have a level of honesty and can assess fairly and objectively (Sabuhari, Sudiro, Irawanto, & Rahayu, 2020). This employee performance appraisal procedure is carried out by each work unit leader who has responsibility for managing the performance appraisal process by looking at several attributes including managing all coaching processes from formation, appointment, placement, transfer, and discipline to employee dismissal. Employee performance appraisals include loyalty, work performance, ability to master all forms of tasks and work, responsibilities, ability to complete tasks as well as possible and on time, obedience, obedience to the company, honesty, carrying out tasks sincerely, and not abusing authority and position. , able to work together, the ability to work in a team, maintain relationships with fellow employees, can take the initiative and be creative in carrying out tasks, leadership, and ability to lead members in a work unit. This element applies to those who have positions(Paais & Pattiruhu, 2020).

Data mining

Data mining is the process of extracting data or finding new information by looking for certain patterns or rules from data. In Data Mining, several processes must be carried out before data mining is carried out. In this study, data analysis was carried out using the decision tree method, so it is necessary to carry out the data analysis process first before the data used is in mining(Informasi, n.d.). The data to be used and analyzed is employee data based on predetermined criteria. In analyzing the employee data, there are attributes used, namely; reasons for absenteeism, months of absence, days of the week, season, transportation costs, distance from the place of residence to work, time of service, age, average workload/day, pressing targets, education, son, social, weight, height, body mass index, time absent in hours(Sedik, n.d.).

METHOD

In measuring employee performance using Knime, some stages must be carried out from research procedures starting from identifying and formulating problems that are the main objectives, along with the steps to perform data mining, namely

A. Data Pre-processing

1. The data will be cleaned in the pre-processing data process, to eliminate empty and imperfect values. Usually, in the pre-processing process, a note file reader, note column filter, and missing values are needed.
2. The file reader node is used for configuration looking for files that have been exported in CSV form, the column filter node is used to filter unused attributes so that only three 3 attributes are selected, namely delta_time, length, and label.
3. The data label process uses a missing value node, but the data must be clean from imperfect columns such as missing data or irrelevant attributes.

B. Data Transformation

In the data transformation process, the pre-processed data is converted into a format that is by the data mining process, and the notes used are normalizers. To form a range of data to recognize patterns.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11796
e-ISSN : 2541-2019
p-ISSN : 2541-044X

C. Data Mining

The Decision Tree Classification method with the Cross Validation technique is used in the data mining process, the nodes needed are X-Partitioner, Decision Tree Learner, Decision Tree Predictor, X-Aggregate

1. X-Partitioner functions to determine the number of iterations or repetitions in the cross-validation technique, this data will later be divided into 2, namely training data and testing data.
2. Decision Tree Learner functions as training data, because the Decision Tree method is supervised learning, so it requires training data to recognize patterns from each data.
3. After running the Decision Tree Learner, then proceed with the Decision Tree Predictor. This node functions to classify data by testing data testing with the results of the Decision Tree Learner process.
4. The X-Aggregator node serves as the end of the cross-validation process. This node will collect the results from the Node Predictor which will display the results of the predictions from several iterations.

D. Evaluation

The evaluation process is used to recognize patterns with node scores using the confusion matrix technique.

RESULT

Pre-processing data in this study to execute Employee Work Attendance Dataset Using K-NIME using Decision Tree, thus forming a workflow/project Decision Tree

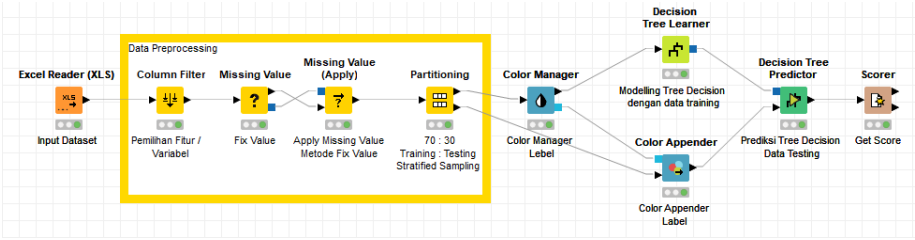
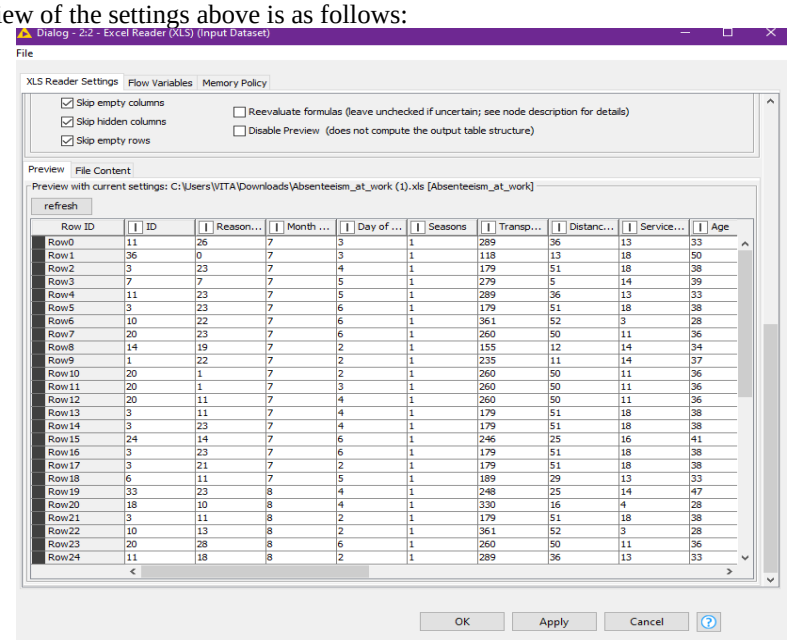


Figure 1. workflow/project Knime Employee performance

From the picture above, after the dataset is used as an excel reader when the data is pre-processed, the missing value process is carried out to be processed using the decision tree method, the process is as follows:

1. Process Dataset Reader

   Excel Reader as a process of reading dataset files in the form of an Excel extension. The configuration of the process is as follows: Absenteeism_at_work. The file is configured by checking the check box "Table contains column name in row number 1, meaning that the column name or attribute name will be on the first row, a preview of the settings above is as follows:



Row ID	ID	Reason...	Month ...	Day of ...	Seasons	Transp...	Distanc...	Service...	Age
Row0	11	26	7	3	1	289	36	13	33
Row1	36	0	7	3	1	118	13	18	50
Row2	3	23	7	4	1	179	51	18	38
Row3	7	7	7	5	1	279	5	14	39
Row4	11	23	7	5	1	289	36	13	33
Row5	3	23	7	6	1	179	51	18	38
Row6	10	22	7	6	1	361	52	3	28
Row7	20	23	7	6	1	260	50	11	36
Row8	14	19	7	2	1	155	12	14	34
Row9	1	22	7	2	1	235	11	14	37
Row10	20	1	7	2	1	260	50	11	36
Row11	20	1	7	3	1	260	50	11	36
Row12	20	11	7	4	1	260	50	11	36
Row13	3	11	7	4	1	179	51	18	38
Row14	3	23	7	4	1	179	51	18	38
Row15	24	14	7	6	1	246	25	16	41
Row16	3	23	7	6	1	179	51	18	38
Row17	3	21	7	2	1	179	51	18	38
Row18	6	11	7	5	1	189	29	13	33
Row19	33	23	8	4	1	248	25	14	47
Row20	18	10	8	4	1	330	16	4	28
Row21	3	11	8	2	1	179	51	18	38
Row22	10	13	8	2	1	361	52	3	28
Row23	20	28	8	6	1	260	50	11	36
Row24	11	18	8	2	1	289	36	13	33

Figure 2 Configuration File Reader

Preprocessing data as follows:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11796
e-ISSN : 2541-2019
p-ISSN : 2541-044X

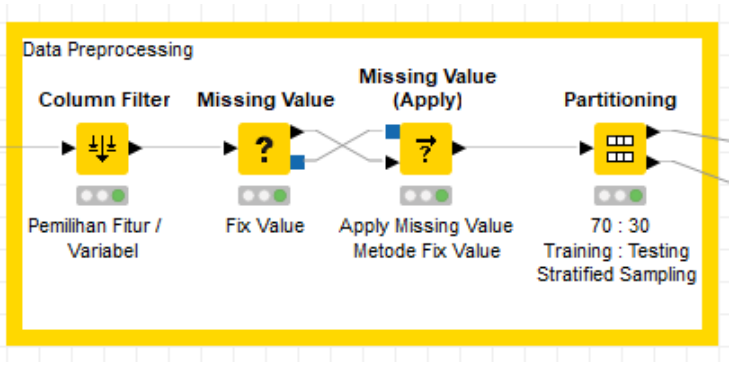


Figure 3. Flow for the data merging process

Column Filter is used in the process of selecting which variables or attributes are used in the next process. Configure the Column Filter process as follows:

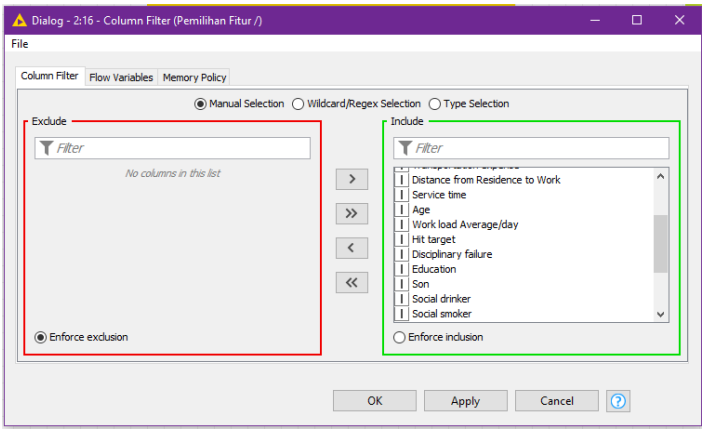


Figure 4. Configuration on Node Missing Value

Missing Value and Missing Value Apply in this process is carried out on the dataset as a check whether there is an error data or the value is unknown. The configuration for the program related to Missing Value is as follows:

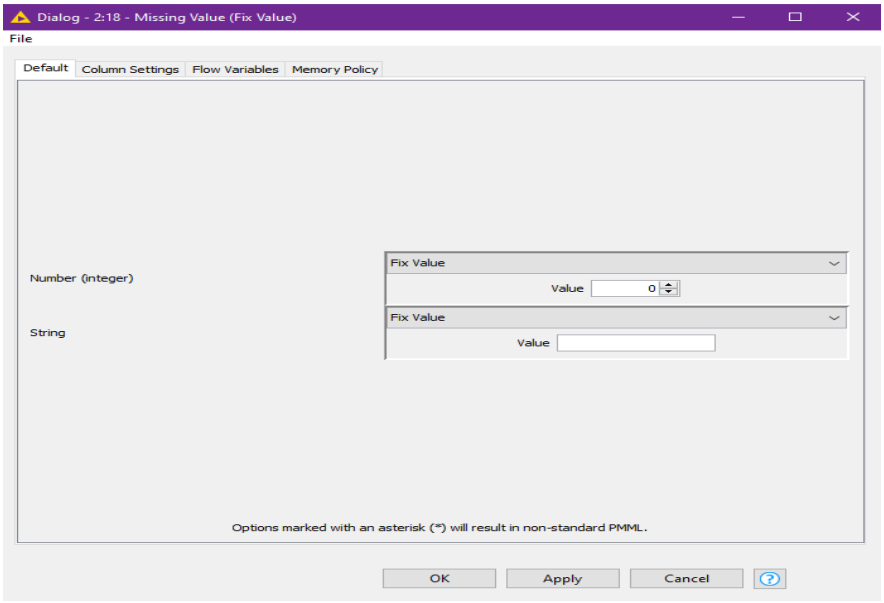
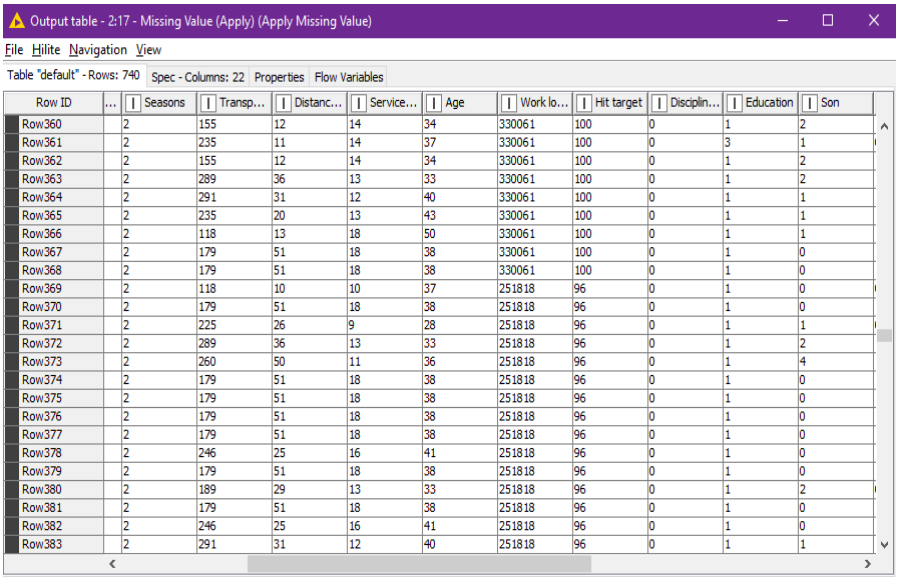


Figure 5 Configure the Joiner node to display latitude longitude

Some of the output tables from Missing Value Apply are as follows:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11796
e-ISSN : 2541-2019
p-ISSN : 2541-044X



Row ID	...	Seasons	Transp...	Distanc...	Service...	Age	Work lo...	Hit target	Disciplin...	Education	Son
Row360		2	155	12	14	34	330061	100	0	1	2
Row361		2	235	11	14	37	330061	100	0	3	1
Row362		2	155	12	14	34	330061	100	0	1	2
Row363		2	289	36	13	33	330061	100	0	1	2
Row364		2	291	31	12	40	330061	100	0	1	1
Row365		2	235	20	13	43	330061	100	0	1	1
Row366		2	118	13	18	50	330061	100	0	1	1
Row367		2	179	51	18	38	330061	100	0	1	0
Row368		2	179	51	18	38	330061	100	0	1	0
Row369		2	118	10	10	37	251818	96	0	1	0
Row370		2	179	51	18	38	251818	96	0	1	0
Row371		2	225	26	9	28	251818	96	0	1	1
Row372		2	289	36	13	33	251818	96	0	1	2
Row373		2	260	50	11	36	251818	96	0	1	4
Row374		2	179	51	18	38	251818	96	0	1	0
Row375		2	179	51	18	38	251818	96	0	1	0
Row376		2	179	51	18	38	251818	96	0	1	0
Row377		2	179	51	18	38	251818	96	0	1	0
Row378		2	246	25	16	41	251818	96	0	1	0
Row379		2	179	51	18	38	251818	96	0	1	0
Row380		2	189	29	13	33	251818	96	0	1	2
Row381		2	179	51	18	38	251818	96	0	1	0
Row382		2	246	25	16	41	251818	96	0	1	0
Row383		2	291	31	12	40	251818	96	0	1	1

Figure 6 The result of the Missing Value process

The Partitioning, Dataset in this process divides the dataset into training data and testing data with an amount of 70: 30 each with the Stratified Sampling technique on the Label attribute that is used as a reference, as for the configuration as follows:

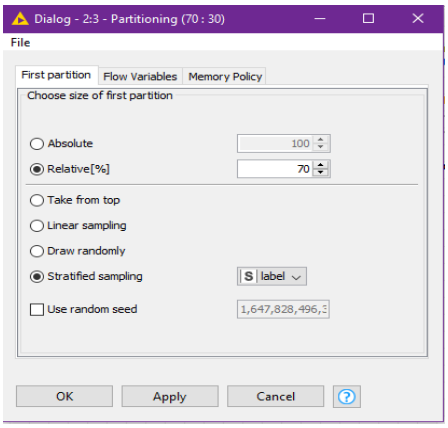
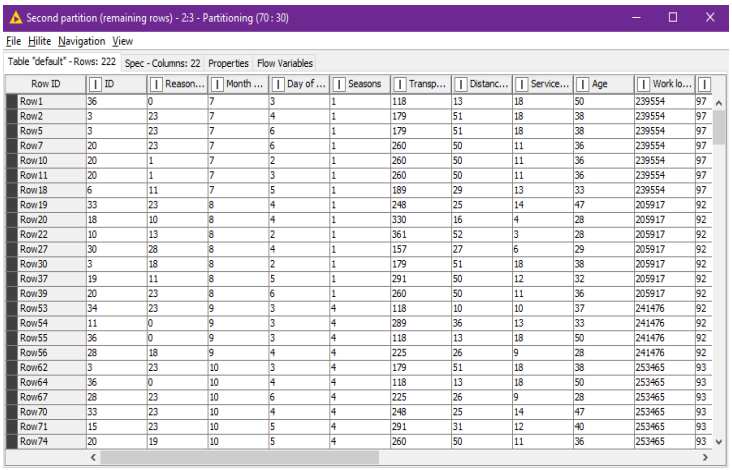


Figure 7 X-Partitioner node configuration

In this data sharing process, 70% of the training data is divided with the number of Rows 518 and Cols 22. And the testing data is 30% with the number of Rows 222 and Cols 22 as follows:



Row ID	ID	Reason...	Month ...	Day of ...	Seasons	Transp...	Distanc...	Service...	Age	Work lo...	
Row1	36	0	7	3	1	118	13	18	50	239554	97
Row2	3	23	7	4	1	179	51	18	38	239554	97
Row5	3	23	7	6	1	179	51	18	38	239554	97
Row7	20	23	7	6	1	260	50	11	36	239554	97
Row10	20	1	7	2	1	260	50	11	36	239554	97
Row11	20	1	7	3	1	260	50	11	36	239554	97
Row18	6	11	7	5	1	189	29	13	33	239554	97
Row19	33	23	8	4	1	248	25	14	47	205917	92
Row20	18	10	8	4	1	330	16	4	28	205917	92
Row22	10	13	8	2	1	361	52	3	28	205917	92
Row27	30	28	8	4	1	157	27	6	29	205917	92
Row30	3	18	8	2	1	179	51	18	38	205917	92
Row37	19	11	8	5	1	291	50	12	32	205917	92
Row39	20	23	8	6	1	260	50	11	36	205917	92
Row53	34	23	9	3	4	118	10	10	37	241476	92
Row54	11	0	9	3	4	289	36	13	33	241476	92
Row55	36	0	9	3	4	118	13	18	50	241476	92
Row56	28	18	9	4	4	225	26	9	28	241476	92
Row62	3	23	10	3	4	179	51	18	38	253465	93
Row64	36	0	10	4	4	118	13	18	50	253465	93
Row67	28	23	10	6	4	225	26	9	28	253465	93
Row70	33	23	10	4	4	248	25	14	47	253465	93
Row71	15	23	10	5	4	291	31	12	40	253465	93
Row74	20	19	10	5	4	260	50	11	36	253465	93

Figure 8. Result of merging Tables using Joiner node

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11796
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Execution of the Decision Tree Method on the Dataset

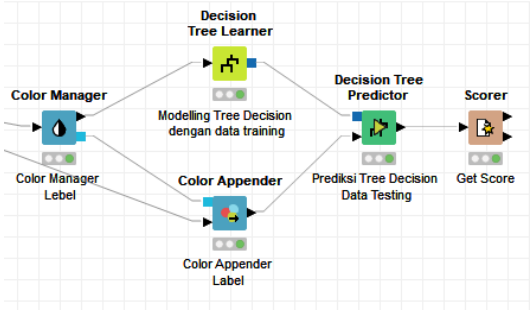


Figure 9 The flow of the process of making markers / marks on the map

Color Manager as the process of setting the color in the label used, while the data used is 70% training data and followed by a decision tree modeling.

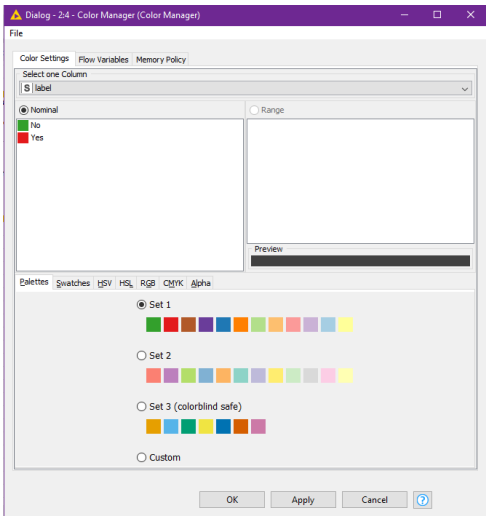


Figure 10 Configuration to specify marker color on map

Decision Tree Learner is the process of executing the decision tree model on the training dataset as a model learning process for the dataset being tested, as for the configuration of the Decision Tree Learner as follows:

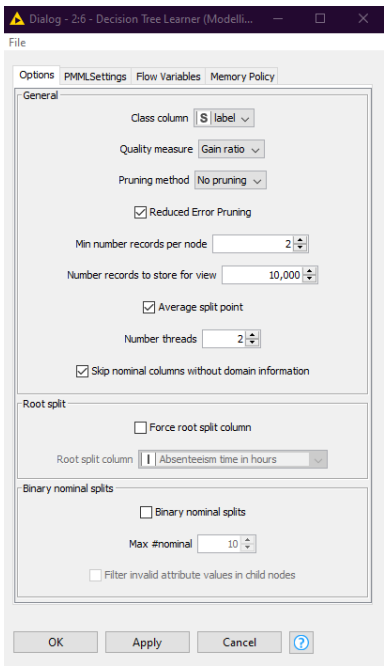


Figure 11 Node Scorer and Configuration

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11796
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Furthermore, on the Training data and Testing data, predictions are made using the Decision Tree Predictor through the Color Appended. In the Class Column this process is the "Label" column with Quality Measure using Gain Ratio. The results of the modeling and predictions in the above process are as follows:

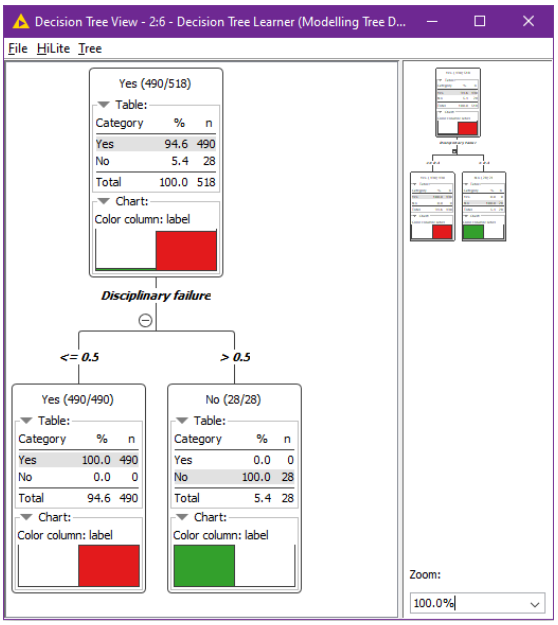


Figure 12 Configuration to specify marker color on map

From the training and testing data that was tested, it can be seen that the accuracy of the Yes label is 94.6% while the No. 5.4%. As for the Get Score process on the testing data, the results of the error are 0% with the following display:

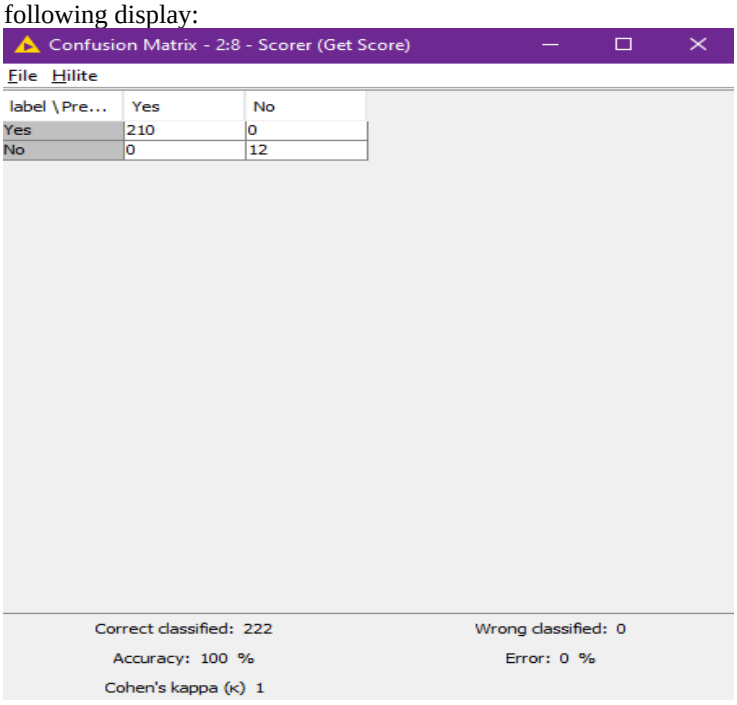


Figure 13. Result of Confusion Matrix

DISCUSSIONS

To predict employee performance, the best model is the decision tree mode compared to other models, the decision tree accuracy value shows a value that is superior to other models. With the decision tree model, the prediction results are more on the actual data in the calculation process. From the results of calculations using a decision tree in terms of employee assessment, it can provide a reference for employees to get promotion opportunities or promotions for an employee from the company. Performance calculations using KNIME are

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

very accurate and easy to get the desired results. Future research is expected to be able to add variables to predict employee performance and apply multi-class (Paais & Pattiruhu, 2020).

CONCLUSION

The decision tree method can be used for classification models from a set of input data, especially employee assessments that are focused on discipline. The decision tree uses a learning algorithm to obtain a classification model. So that the input data becomes an important influence in the formation of the decision tree and the resulting rules. The level of discipline will have an impact on the performance and careers of employees. The use of KNIME can provide a high level of acuteness from training data and test data, the accuracy of the Yes label is 94.6% while No is 5.4%. Meanwhile, for the Get Score process on the test data, the error result is 0%.

REFERENCES

Achmad, D. M., & Fauzi, B. S. (2012). Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree. Jurnal IPTEK, 16(1), 18–23. Retrieved from http://jurnal.itats.ac.id/wp-content/uploads/2013/06/3.-BUDANIS-FINAL-hal-17-23.pdf .

Jijo, B. T., & Abdulazeez, A. M. (2021). *Classification Based on Decision Tree Algorithm for Machine Learning*. *02*(01), 20–28. https://doi.org/10.38094/jastt20165

Oliveira, E. L. De, & Fran, R. A. (2019). *Absenteeism Prediction in Call Center*. *3*, 958–968. https://doi.org/10.1007/978-3-030-16181-1

Ozdemir, F. (2020). *Assessing Employee Attrition Using Classifications Algorithms*. 118–122.

Paais, M., & Pattiruhu, J. R. (2020). *Effect of Motivation , Leadership , and Organizational Culture on Satisfaction and Employee Performance*. *7*(8), 577–588. https://doi.org/10.13106/jafeb.2020.vol7.no8.577

Raman, M., Kaliappen, N., & Suan, C. L. (2020). *A Study on Machine Learning Classifier Models in Analyzing Discipline of Individuals Based on Various Reasons Absenteeism from Work*. 361–365.

Sabuhari, R., Sudiro, A., Irawanto, D. W., & Rahayu, M. (2020). *The effects of human resource flexibility , employee competency , organizational culture adaptation and job satisfaction on employee performance*. *10*, 1777–1786. https://doi.org/10.5267/j.msl.2020.1.001

Sedik, D. (n.d.). *Vol4no Data Mining : Evaluating Performance of Employee ' s using*. *4*.

Wahid, Z. (2019). *Predicting Absenteeism at Work Using Tree-Based Learners*.

*name of corresponding author