DOI: https://doi.org/10.31933/dijdbm.v3i2
Received: 8 March 2020, **Revised:** 10 April 2022, **Publish:** 1 June 2022

HUMAN RESOURCES MANAGER'S DECISION MAKING TOWARDS EMPLOYEES USING DATA MINING METHOD

Nia Kusuma Wardhani[1], Windu Gata[2]
[1]Universitas Mercu Buana, Indonesia, nia.kusuma@mercubuana.ac.id
[2]Universitas Nusa Mandiri, Indonesia, windu@nusamandiri.ac.id

Corresponding Author: Nia Kusuma Wardhani

Abstract: Important decisions in decision making in organizations are often initiated by Human Resources Managers, where data objects in the form of personal or employee must be complete and processed not by using feelings in decision making. If the forecast results in making decisions that are not right, it can result in the company not being ready to meet the future so that it has the potential to harm the company itself. Based on previous studies on Human Resource Management, this research was conducted by testing the Kaggle repository dataset using the Datamining methodology, where the results obtained can be used as an initial reference for a manager who makes decisions based on datasets from the company. Usually for data processing using qualitative and quantitative methods, but in this study using the Data Mining method using algorithms from machine learning, namely: Estimation, Prediction, Cluster, and Classification. It is hoped that the results of this research will enable Human Resource Managers as a basis for analysis to make decisions using Data Mining or Machine Learning as needed.

Keywords: Decision-making; Resource Manager; Employee; Data Mining; Machine Learning

INTRODUCTION

Employees are important assets owned by the company, so it is necessary to have good human resource management so that they can improve the company's business development. In an effort to improve employee performance, there are various important decisions in the management of human resource management, such as decisions on hiring new employees, decisions on promotion results, career development decisions, decisions on salary increases and employee benefits, and so on (Pratiwi, 2018) (Azizah et al., 2020) (Azizah et al., 2020). Good decision shows work professionalism in the management of human resource management (Vira Afrianti et al., 2021).

The decision-making process in an organization is often initiated by the Human Resources Manager, where personal or employee object data must be complete and processed systematically so that decisions are made based on data not using feelings. If the results of the estimation (forecasting) are not correct, it will affect decision making so that it can result in the company not being ready to meet future needs which can ultimately harm the company itself (Catio et al., 2020).

Data processing that is usually used by Human Resources Managers uses basic descriptive statistical models, such as: the highest 10, trend, maximum, minimum, average, and other

calculation models presented in the form of tables and graphs. Likewise with the use of Quantitative models,

Human Resources Managers can compare variables with one another in numerical form, which previously had a hypothesis as an initial guess and data in the form of a test or not that used statistics to represent the population, or a qualitative model based on phenomena and facts. empirical which has natural properties without being engineered(Ismail, 2018). From the Quantitative and Qualitative models, it is expected that Human Resources Managers can make good and correct decisions for the survival of the company today and in the future.

Since the internet began to develop, the data is more and more spread out and can be used by managers to see or assess something. For example, if a Human Resources Manager wants to assess someone who is a candidate for employment, he can view that person's profile on social media or view other data available on the internet as study material. Likewise, companies that implement Resource Management Applications to carry out their company operations will have data that will continue to grow and increase according to the company's growth.

A Human Resources Manager must be observant with these data, where the data can be processed using Machine Learning (Machine Learning) so that patterns and accuracy can be seen that can help a manager to make decisions (Ranjan et al., 2008). This can also be a guide in providing a return on investment (ROI) called Human Resources based Organizational Data Mining (HRODM) (Hila Chalutz Ben-Gal, 2020). Indeed, at this time, especially research in the Portuguese country, Data Mining methods are still not a priority for most companies, especially with the challenges of privacy, ethics, and having to have an analysis team from different backgrounds (Alves, 2020).

Data Mining is a combination of statistics, artificial intelligence (Artificial Intelligence) and machine learning (Machine Learning) to obtain patterns (Pattern) from mined data so that it can be used for better decision making (Marco Mellia et al., 2021).

This study provides an example of how Human Resource Managers can use the Data Mining method as a future research method, specifically: Estimation, Prediction, Clustering, and Classification for decision making.

LITERATURE REVIEW

Data Mining (Data Mining) is a collection of techniques for the automation of discovery, validation, novelty, use, and understanding of patterns in large databases (Bhartia, 2019). The well-known stage models in data processing are Knowledge Discovery in Databases (KDD)(Lavrac, 2021) and The Cross Industry Standard Process Data Mining (CRISP-DM)(Lopez, 2021). Where the use of Machine Learning Algorithms that can be used by Human Resources Managers has several categories, namely: Estimation, Cluster Prediction, Association, and Classification (Refaat et al., 2021).

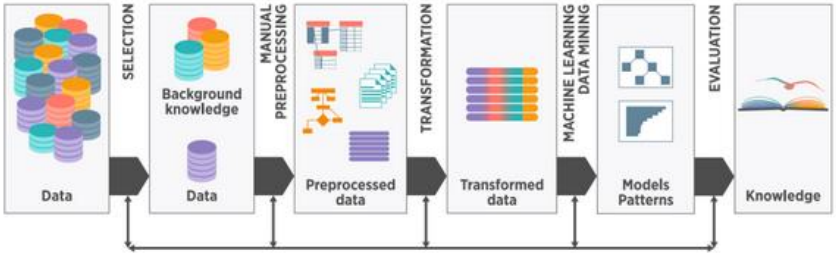

Figure 1. Data Mining Method: Knowledge Discovery in Database (KDD)

The stages that must be carried out by a Human Resources Manager if they want to use the KDD

Research Method, as shown in Figure 1, which has 6 (six) stages, namely: Data Selection is the initial stage of development to recognize the application used, the main knowledge , the goal of the user, selects a dataset that focuses on some variable or sample data; Preprocessing Data (Preprocessing Data) is the stage of eliminating noise or outliers, handling missing features, calculating the speed of information sequentially, and knowing changes in data; Data Transformation is the technological automation of features and formation of training datasets for use in machine learning; Machine Learning (Machine Learning) or Data Mining (Data Mining) is the stage of applying algorithms to machine learning or data mining such as classification, regression, association, or clusters in developing models and getting patterns in the data; and Evaluation is an assessment of the results that have been carried out such as evaluation of data models and patterns, explanation of data models and patterns, visualization, and reducing the potential for repetitive data.

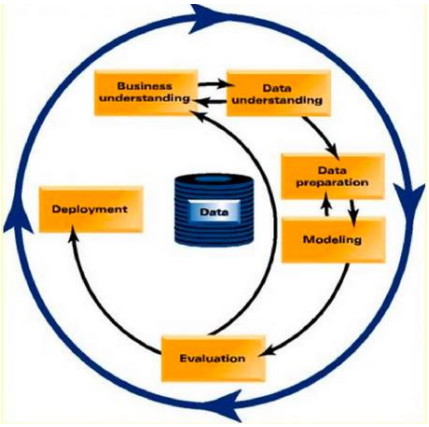

Figure 2. The CRoss Industry Standard Process Data Mining (CRISP-DM)

The most famous method in Data Mining that can also be chosen by Resource Management is CRISP-DM which is depicted in Figure 2. This CRISP-DM has 6 (six) stages, namely: Business Understanding is the stage of understanding project objectives and demand from the perspective of the business to be transformed into knowledge defined by the problem and shaping the design to achieve the goal; Data Understanding is the stage of initialization and data collection, identifying data problems, searching for everything in the data content, and hypothesizing the stored information; Data Preparation is the activity of preparing the Final Dataset from Data Raw Materials and carrying out transformation and cleaning; Modeling (Modelling) is the selection of Modeling techniques that use algorithms from Machine Learning; Evaluation is an assessment that brings out the highest quality of Data by doing a comparison and then comparing it with Business Objectives; and Deployment, namely distributing the results of the evaluation so that an organization can make decisions in the form of a Live model that can be in the form of an application.

Several studies on employees use the Data Mining method and obtain patterns and are used as the basis for decision making by Human Resources Managers, as follows:

1. Estimation, Employee Resignation Using a Machine Learning-Based Model using the Logistics Regression Algorithm in China, where the results of this study expect Human

Resources Managers to be able to predict when an employee resigns and how to prevent and handle it(Dai & Zhu, 2021).

2. Classification, Measurement of Employee Performance Development using the C4.5, Naive Bayes, and Fuzzy Logic algorithms where Human Resources Managers can monitor employee performance appraisals and performance developments(Santhosh & Mohanapriya, 2021). Another research regarding a hospital in China that uses Data Mining to analyze and organize and

   improvise employees using Decision Trees (Yang et al., 2018), research in several companies in Sukabumi uses the Classification and Regression Trees (CART) algorithm to describe Employee Performance (Pahmi et al., 2019).

3. Clusters, measuring employee motivation based on statistical analysis using clusters with questionnaires that are expected to find out various motivational profiles of employees in the company or industry(Zámečník, 2014). Another example is the clustering of company shareholdings to companies in Germany, the aim is to find out the attractiveness of retention related to all the company's needs(Orlieb et al., 2016).

4. Examples 1 to 4 also explain the use of predictions from each analysis using new data for implementation.

based on previous studies, this research was conducted by testing the Kaggle repository dataset related to Human Resource Management using the Datamining methodology, where the results obtained can be used as an initial reference for a manager who makes decisions based on datasets from the company.

RESEARCH METHODS

In this study, the research model is a case study method from various examples of Dataset – Repository models related to Human Resources which are processed in the Data Mining model using Machine Learning Algorithms with Estimation, Cluster, Association, Classification, and Prediction models as in Figure 3.

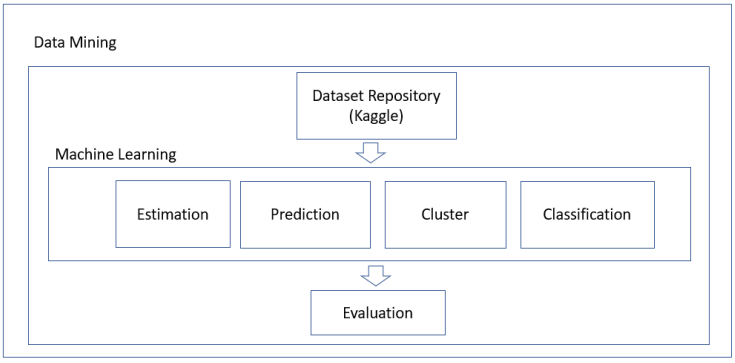

Figure 3. Research stage

The use of the estimation algorithm function has the aim of estimating something unknown from the output of discrete data. Prediction function to find out the future by using new data from patterns that have been obtained from old data. While the cluster aims to see the similarity

(heterogeneous) of the population, while the classification is consistent with the future state and the renewal of the object so that it leads to the label (J. A. Berry & S. Linoff, 2004).

FINDINGS AND DISCUSSION

In this study, the dataset used was taken from the repository www.kaggle.com with the link https://www.kaggle.com/rhuebner/human-resources-data-set with the name of the dataset in the form of CSV HRDataset_v14.csv which has 36 columns and total 312 employees. While the assistive device used is the Rapidminer Application. and the use of the Data Mining method where the use of algorithms from each Machine Learning will be discussed one by one, starting from Estimation then Prediction, Cluster, and Classification.

Estimation

Estimation Algorithm is used to estimate the actual situation with calculations estimated by the algorithm. For example: Human Resources Manager wants to know whether the salary (#Salary) received is in accordance with the Machine Learning Algorithm calculation with the supporting attributes of Department (#Departement), Employee Satisfaction (#EmpSatisfaction), Engagement Survey (#EngageSurvey), Length of Work (#LongWorking), Marriage (#MaritalDesc), and Performance Score (#PerformanceScore).

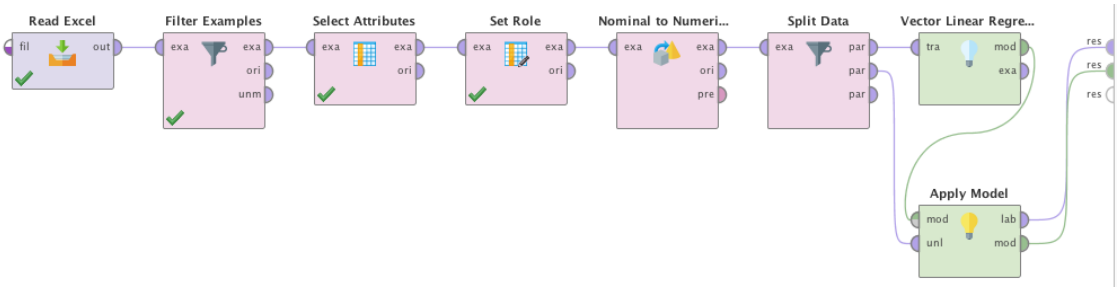

Figure 4. Employee Salary Estimation Model

There are many algorithms that can be used to estimate, such as Linear Regression, Polynominal Regression, Vector Linear Regression, and others. In this study, the example is Vector Linear Regression which can be seen in detail in the Rapidminer Model in Figure 4. The results of the model that have been made will be formed in the form of a formula for the suitability of employee salaries based on the dataset, as follows:

```
Salary =   (23754.988 * Sex = M   + 23749.104 * Sex = F  + 6808.532 * MaritalDesc = Single  +
7835.327 * MaritalDesc = Married  + 1596.259 * MaritalDesc = Divorced  + 24201.231 * MaritalDesc
= Widowed  + 7062.646 * MaritalDesc = Separated  - 41094.909 * Department = Production        -
7849.819 * Department = IT/IS  - 6290.624 * Department = Software Engineering  - 17943.465 *
Department = Admin Offices  - 30498.076 * Department = Sales  + 151180.732 * Department =
Executive Office  + 19744.193 * PerformanceScore = Exceeds  + 7139.680 * PerformanceScore = Fully
Meets  + 18634.156 * PerformanceScore = Needs Improvement  + 1985.991 * PerformanceScore = PIP  +
853.310 * LongWorking  - 490.138 * EngagementSurvey  + 2141.792 * EmpSatisfaction  + 47504.043)
```

While the results of the comparison of the suitability of employee salaries in the dataset are shown in table 1, as follows:

Table 1. Example of Estimation of Employee Salary Suitability

EmpID	Salary	Estimasi
10062	59,365	75,301
10250	50,178	86,377
10150	77,692	92,846
10085	93,396	92,727
10040	71,860	69,788
10108	110,929	92,007

After seeing the comparison of the suitability of the employee's salary as a whole, the Human Resources Manager can make a decision whether to make adjustments or as a reference for the future. For example: Employee with ID 10062 has a salary (#Salary) of USD 59,365 but the prediction is very different from the Vector Linear Regression algorithm of USD 75,301 then the Resource Manager can decide whether to increase or keep the salary of the Employee. Meanwhile, employees with IID 10108 have a salary of USD 110.929 which is greater than the prediction from the Vector Linear Regression algorithm with a value of USD 92.007 so that the Human Resources Manager can see in the background why the salary of the employee is higher than predicted so that they can make the right decision.

Prediction

A Human Resources Manager can also use the formula that has been obtained from the Vector Linear Regression Algorithm to make predictions which can later become decisions on employee salaries in the future. For example, if there is a new employee of the female gender (F) whose marriage status is Married, it is intended for the Sales Department, has a higher Performance Score (Exceeds) where the employee has been working (LongWorking) for 5 years, has a survey Involvement (Engagement Survey) is 4, and Employee Satisfaction is 5, it can be predicted that the salary value is as follows:

```
1. Salary = 23749.104 * Sex = F  + 7835.327 * MaritalDesc = Married - 30498.076 * Department =
    Sales  + 19744.193 * PerformanceScore = Exceeds + 853.310 * LongWorking  - 490.138 *
    EngagementSurvey  + 2141.792 * EmpSatisfaction  + 47504.043
2. Salary = 23749.104 * 1  + 7835.327 * 1 - 30498.076 * 1  + 19744.193 * 1  + 853.310 * 5  -
    490.138 * 4  + 2141.792 * 5  + 47504.043  = 81350
```

Based on the above formula, the estimated salary of the employee is 81,350 (eighty-one thousand three hundred and fifty) dollars. The Human Resources Manager can make a decision on the estimated salary value of the new employee.

Cluster

The Cluster Algorithm or grouping can be used by Human Resources Managers for groups of each employee in terms of Marriage (#MarriedID), Marital Status (#MaritalStatusID), Gender (#GenderID), Employee Status (#EmpStatusID), Department ID (#DeptID), Performance Score (#PerfScoreID), and Position (#PositionID) with a total of 4 Groups.

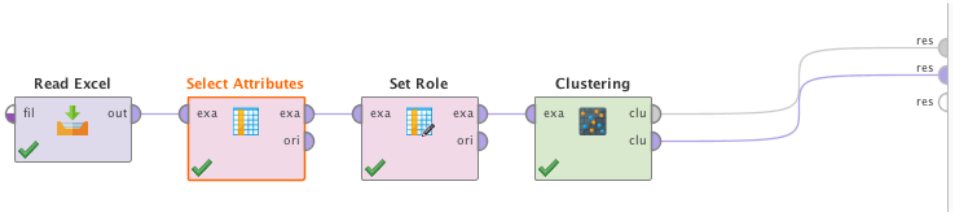

Figure 5. Employee Cluster Model using the K-Medoids Algorithm

The cluster modeling shown in Figure 5 using the K-Medoids Algorithm has the following results: Cluster 0 as many as 90 employees, Cluster 1 as many as 14 employees, Cluster 2 as many as 30 employees, Cluster 3 as many as 177 employees.

The details of each cluster can be seen in Table 2, where employees in Cluster 0 are grouped based on Employee Status, Performance Score, Position, and Gender. Cluster 1 is grouped based on Employee Status and Performance Score. Cluster 2 is grouped based on Marital Status, Performance Score, and Employee Status. And the largest group in Cluster 3 is grouped based on Position, Performance Score, Department, and Employee Status. Table 2 explains that those who have the highest number of 177 employees in Cluster 3 make the Employee Position worthy of attention just looking at the performance side. Meanwhile, Cluster 0 with 90 employees seems to focus on the new

Employee Status Employee Performance.

Table 2. Centroid Table Employee

Atribute	Cluster_0	Cluster_1	Cluster_2	Cluster_3
MarriedID	0	0	0	0
MaritalStatusID	0	0	4	0
GenderID	1	0	0	0
EmpStatusID	5	5	1	1
DeptID	0	0	0	1
PerfScoreID	4	1	3	3
PositionID	2	0	0	5

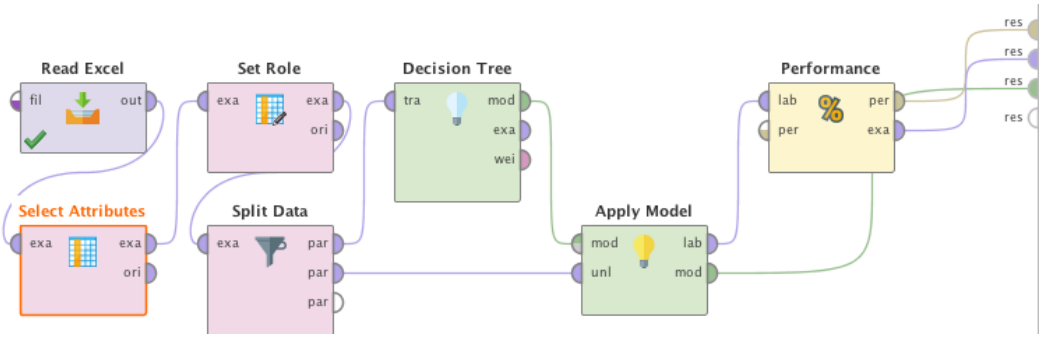

Figure 6. Decision Tree Classification Algorithm Model Based on Performance Score

Classification

The use of a Classification Algorithm by a Human Resources Manager to get a pattern from an employee who has a label of a Performance Score (#PerformanceScore) which consists of Exceeds, Full Meets, Needs Improvement, and Provides Opportunities (Performance Improvement Plan – PIP) by looking at attributes such as: Age (#Age), Department (#Departement), Employee Status (#EmploymentStatus), Employee Convenience (#EmpSatisfaction), Engagement Survey (#EngagementSurvey), Length of Work (# LongWorking), Marital Status (#MaritalDesc), Position (#Position), Recruitment Source (#RecruitmentSource), Salary (#Salary), Gender (#Sex), and Project Special Calculation (#SpecialProjectsCount). In this study using the Decision Tree Classification Algorithm with a model as shown in Figure 6, where the results obtained are in the form of a tree image depicted in Figure 7.

In figure 6 at level 0 it can be seen that the Engagement Survey (#EngagementSurvey) at the beginning of the Tree, as well as at level 1 the Engagement Survey (#EngagementSurvey) is still part of the choice, only at level 2 in the right position it can be seen if the value of the Engagement Survey (#EngagementSurvey) is shown. less than equal to 1,965 have a PIP Performance Score (#PerformanceScore) while the Engagement Survey (#EngagementSurvey) > 1,965 has a Needs Improvement Performance Score if they are under the age of 46.5 while those above 46.5 have a PIP Performance Score.

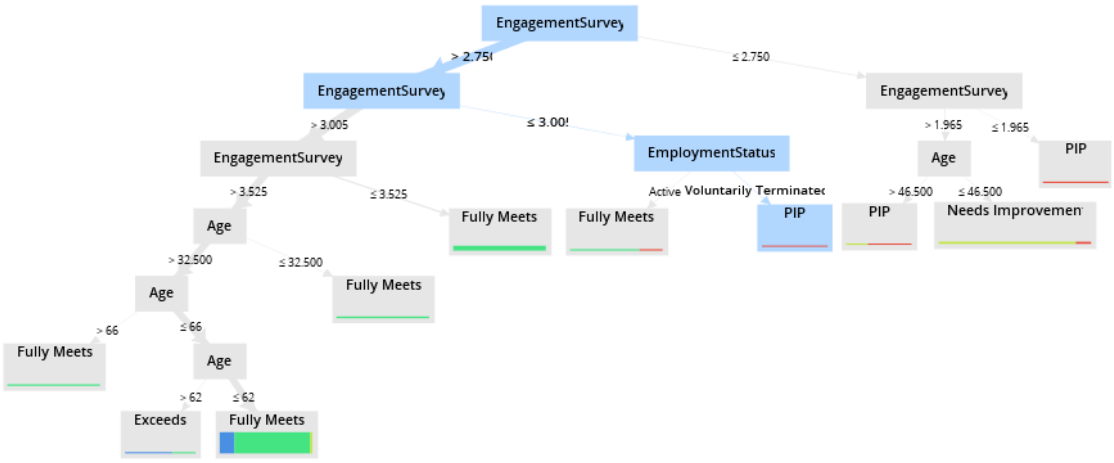

Figure 7. Results of Performance Score Classification using the Decision Tree Classification Algorithm

In the left position of the Engagement Survey (#EngagementSurvey) with a value below 3 it will be seen from the Employee Status (#EmploymentStatus) if Active (#Active) then the Performance Score (#PerformanceScore) is appropriate (#FullyMeet) if it is completed it will have a Performance Score (#PerformanceScore) PIP. If the Engagement Survey (#EngagementSurvey) is above 3 then age (#Age) below 32.5 has an Appropriate Performance Score (#PerformanceScore) (#FullyMeet). And if the Age (#Age) is above 32.5 and above 66 will be appropriate, while the age between 62 to 66 will have an Over value (#Exceed) and the others will be Suitable (#FullyMeets).

CONCLUSION AND RECOMMENDATION

Based on these various examples of data mining, Human Resources Manager can use the Data Mining method to find out patterns from Employee Datasets with Estimation, Prediction, Cluster, and Classification Algorithms as a basis for making decisions in the company, especially strategic decisions in the management of the field of Human Resources management.

BIBLIOGRAPHY

Alves, R. I. da C. V. (2020). *Data Mining Methods In Human Resources Management In Portugal: Purposes, Challenges And Opportunities*. Universidade Catolica Portuguese.

Azizah, N., Pringgabayu, D., Ajibroto, K., Studi Manajemen Perhotelan Politeknik Pajajaran ICB Bandung, P., & Studi Keuangan dan Perbankan Politeknik Pajajaran ICB Bandung, P. (2020). *Pengaruh Sistem Informasi Manajemen Kepegawaian Terhadap Pengambilan Keputusan Pimpinan dalam Pengembangan Karier Pegawai*.

Bhartia, P. (2019). *Data Mining and Data Warehousing: Principal and Partical Techniques*. Cambridge University Press.

Catio, M., Sarwani, & Ruknan. (2020). *Manajemen Sumber Daya Manusia*. Indigo Media.

Dai, W., & Zhu, Z. (2021). Employee Resignation Prediction Model Based on Machine Learning. *Springer*, 367–3734.

Hila Chalutz Ben-Gal. (2020). Human Resources based Organizational Data Mining (HRODM): Themes, Trends, Focus, Future. *SSRN*, 1–39.

Ismail, F. (2018). *Statiska Untuk Penelitian Ilmu Pendidikan dan Ilmu-Ilmu Sosial*. Prenadamedia Group.

J. A. Berry, M., & S. Linoff, G. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relation Management*. WILEY.

Lavrac, N. (2021). *Representation Learning: Propostionalization and Embeddings*. Springer.

Lopez, C. P. (2021). *Data Mining. The CRISP-DM Methodology. The CLEM Languange and IBM SPS Modeller*. Lulu.com.

Marco Mellia, Nur Zincir-Heywood, & Yixin Diao. (2021). Communication Networks and Service Management in the Era of Artificial Intelligence and Machine Learning. In *Communication Networks and Service Management in the Era of Artificial Intelligence and Machine Learning*. Wiley. https://doi.org/10.1002/9781119675525

Orlieb, R., Mastiake, W., & Simon Fietsze. (2016). Employee Share Ownership in Germany: A Cluster Analysis of Firm's Aim. *JSTOR*.

Pahmi, S., Saepudin, S., Maesarah, N., Solehudin, U. I., & Wulandari. (2019). Implementation of CART (Classification and regression trees) algorithm for determining factors affecting employee performance. *Proceedings - 2018 4th International Conference on Computing, Engineering, and Design, ICCED 2018*, 57–62. https://doi.org/10.1109/ICCED.2018.00021

Pratiwi, A. (2018). Kepegawaian (Simpeg) Sebagai Dasar Pengambilan Keputusan Bidang Sumber Daya Manusia (SDM) Di Bkdiklatda Kota Salatiga. *Jalan Kapten Suparman No. 39 Magelang*, *02*(02).

Ranjan, J., Goyal, D. P., & Ahson, S. I. (2008). Data Mining Techniques For Better Decisions In Human Resource Management Systems. *International Journal of Business Information Systems*, *3*(5), 464–481.

Refaat, S. S., Ellabban, O., Bayhan, S., Abu-Rub, H., Blaabjerg, F., & Begovic, M. M. (2021). *Smart Grid and Enabling Technologies*. WILEY.

Santhosh, R., & Mohanapriya, M. (2021). Generalized Fuzzy Logic Based Performance Prediction In Data Mining. *Materials Today: Proceedings*, *45*(xxxx), 1770–1774. https://doi.org/10.1016/j.matpr.2020.08.626

Vira Afrianti, P., Bayu Putra, R., Yona Sari, S., & Korespondensi Vira Afrianti, P. (2021). Budaya Organisasi, Pengaruh Kepemimpinan, Perilaku, Keputusan Dan Profesionalisme Kerja Terhadap Kinerja Karyawan Pada Kantor Radio Republik Indonesia Kota Padang. *Strategic: Journal of Management Sciences*, *1*(2). http://jurnal.stiesultanagung.ac.id/index.php/strategic

Yang, L., Xu, H., & Jiang, Y. (2018). Applied Research Of Data Mining Technology In Hospital Staff Appraisal. *Procedia Computer Science*, *131*, 1282–1288. https://doi.org/10.1016/j.procs.2018.04.340

Zámečník, R. (2014). The Measurement of Employee Motivation by Using Multi-factor Statistical Analysis. *Procedia - Social and Behavioral Sciences*, *109*(1998), 851–857. https://doi.org/10.1016/j.sbspro.2013.12.553