EMPLOYEE CREDIT CLASSIFICATION ANALYSIS USING DECISION TREE BASED CRISP-DM MODEL (CASE STUDY OF SAMSUNG INDONESIA COMPANY)

Riad Sahara[1], Syahid Abdullah[2], Cian Ramadhona Hassolthine[3], **Muhammad Ikhwani Saputra[4]**

[1,2,3,4]Program Studi PJJ Informatika, Universitas Siber Asia

Jl. Harsono RM, Ragunan, Pasar Minggu Jakarta Selatan 12550
[1]riadsahara@lecturer.unsia.ac.id, [2]syahidabdullah@lecturer.unsia.ac.id, [3]cianhassolthine@lecturer.unsia.ac.id, [4]muhammadikhwani@lecturer.unsia.ac.id

Abstract

Employee credit program is a form of employee retention as an effort to retain potential employees from the company. In its implementation, providing employee credit is not without risks that may occur. These risks include the inability or failure to pay credit installments when due. To minimize the risks that may occur, a survey and analysis with the right method is needed for cooperative members before providing employee credit. Researchers will use a Decision Tree-based algorithm as a tool for decision making in providing employee credit to Cooperatives at Samsung Indonesia Company. Researchers also use the CRoss-Industry Standard Process for Data Mining (CRISP-DM) model on the data mining development life cycle as a research step taken. This CRISP-DM model is very appropriate to use because it is a neutral model or method and can be used in various industries and combined with various tools. From the measurements that have been carried out using a sample data of 10 records from a total of 584 records, a classification model of 2 employees with non-performing employee credit status and 8 employees with performing employee credit status was obtained. The classification model was obtained based on the Gini Index Value of the Employee ID, Division, and Position attributes are 0.7, 0.3428571, and 0.2714286, respectively. So, the decision to grant credit to employees depends on the Position, after that the Division, and the last is the ID of the employee.

Keywords: *CRISP-DM, Data Classification, Data Mining, Decision Tree, Employee Credit*

Abstrak

Program kredit karyawan merupakan salah satu bentuk retensi karyawan sebagai upaya mempertahankan karyawan yang potensial dari perusahaan. Dalam pelaksanaannya, pemberian kredit karyawan bukan tanpa risiko yang mungkin terjadi. Risiko tersebut antara lain ketidakmampuan atau kegagalan membayar angsuran kredit pada saat jatuh tempo. Untuk meminimalisir risiko yang mungkin terjadi, diperlukan survei dan analisis dengan metode yang tepat bagi anggota koperasi sebelum memberikan kredit karyawan. Peneliti akan menggunakan algoritma berbasis *Decision Tree* sebagai alat bantu pengambilan keputusan dalam pemberian kredit karyawan pada Koperasi di Perusahaan Samsung Indonesia. Peneliti juga menggunakan model *CRoss-Industry Standard Process for Data Mining* (CRISP-DM) pada siklus hidup pengembangan *data mining* sebagai langkah penelitian yang dilakukan. Model CRISP-DM ini sangat tepat digunakan karena merupakan model atau metode yang netral dan dapat digunakan di berbagai industri serta dikombinasikan dengan berbagai *tool*. Dari pengukuran yang telah dilakukan dengan menggunakan data sampel sebanyak 10 record dari total 584 *record*, diperoleh model klasifikasi sebanyak 2 orang karyawan dengan status kredit karyawan macet dan 8 orang karyawan dengan status kredit karyawan lancar. Model klasifikasi diperoleh berdasarkan Nilai *Gini Indeks* dari atribut ID Karyawan, Divisi, dan Jabatan masing-masing sebesar 0.7, 0.3428571, dan 0.2714286. Jadi, keputusan pemberian kredit kepada karyawan bergantung pada Jabatan, setelah itu Divisi, dan terakhir ID karyawan.

Kata kunci: *CRISP-DM, Data Mining, Decision Tree, Klasifikasi, Kredit Karyawan*

1. INTRODUCTION

Samsung Indonesia Company, through its cooperatives, functions to manage savings and loan activities, one of which is the employee credit program. The employee credit program is a form of employee retention as an effort to retain potential employees from the company. It is hoped that this program will have a positive impact that will provide benefits for employees and profits for the company [1].

In practice, granting employee credit is not without risks that may occur [2]. These risks include inability or failure to pay credit installments when they are due [3]. This can happen because Cooperative Companies, as organizations that provide credit, do not analyze the level of credit risk based on their members' credit history data [4]. If this risk occurs the impact will be on the financial health of the cooperative company, which will also be felt by all its members [5].

To minimize the risks that might occur as in the previous description, a survey and analysis using appropriate methods is needed for cooperative members before being given employee credit [6]. The survey and data analysis carried out apart from minimizing risk, can also speed up the analysis time for granting employee credit when compared to data analysis using conventional models and appropriate tools [7]. Researchers will use a Decision Tree-based algorithm as a tool for decision making in granting employee credit to Cooperatives at Samsung Indonesia Company. The Decision Tree algorithm has many advantages, especially for managing numeric and discrete data [8], as well as other advantages that make the performance of this algorithm the fastest when compared to other similar algorithms [9]. Researchers also used the Cross-Industry Standard Process for Data Mining (CRISP-DM) model regarding the data mining development life cycle as a research step. The CRISP-DM model is very appropriate to use because it is a neutral model or method [10] and can be used in various industries and combined with various tools [11] as well as appropriate ICT products for the needs of the financial sector, especially credit [12].

The hypothesis of the research that the researcher will carry out is: (1) The results of an analysis of the correct calculation of the level of credit risk based on employee credit history data, as well as recommendations for Standard Operating Procedures (SOP) for applying and granting employee credit in a company. (2) The level of accuracy of the Decision Tree algorithm in data modeling for the feasibility of granting credit to employees in company cooperatives.

The results of the research carried out are an algorithm recommendation which is intended as a tool to help make decisions regarding the classification of employee credit from an agency or organization. This is done so that credit is provided on target and minimizes risks that may occur in the credit business process.

2. RESEARCH METHODS

2.1. Research Methods and Stage

The research carried out was a type of experimental research. The experimental design used refers to the CRISP-DM model, with six main phases or stages added with three stages for developers, with a business process model as shown in Figure 1. below this [13]. An industry-neutral process model for data mining is called the Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM is the most popular framework used in data mining projects. It provides a structured approach to data mining, guiding analysts through various phases, including business understanding, data understanding, data preparation, modeling, evaluation, and deployment. This framework ensures consistency and efficiency in data mining projects, making it a valuable tool for organizations seeking to extract valuable insights from their data.

It comprises six iterative stages, starting with business insight and ending with deployment. Based on the CRISP-DM user guide, Table I provides a quick summary of the major concepts, assignments, and results of this phase [14].

TABLE I. CRISP-DM PROCESS MODEL DESCRIPTION[15]

Phase	Description
Business Understanding	To determine the resources that are available and required, the business condition must be evaluated. One of the most crucial things to do at this stage is to decide what your data mining objectives are. It is necessary to first describe the type of data mining (such as classification) and the requirements for its success

	(such as precision). The creation of a project plan is required.
Data understanding	Important responsibilities at this phase include gathering data from data sources, examining, characterizing, and verifying the accuracy of the data. The work of describing data using statistical analysis and determining features and their order is explained in more detail in the user guide.
Data preparation	Data selection must be accomplished by establishing inclusion and exclusion criteria. Data cleansing is a useful tool for addressing poor data quality. Derived characteristics need to be constructed based on the model (specified in the first phase) that is used. Various approaches are conceivable for each of these steps, depending on the model.
Modeling	Choosing modeling approaches and creating models and test cases comprise the data modeling stage. You can apply any data mining technique. The data and the business problem determine the decision. The way the decision is explained is more crucial. Some parameters need to be set before a model can be built. The best way to evaluate a model is to compare it to the evaluation criteria and choose the best model.
Evaluation	During the assessment stage, the outcomes are compared to the predetermined business goals. As a result, it is necessary to evaluate the data and decide what needs to be done next. In addition, a broad examination of the process is necessary.
Deployment	The user guide provides a general description of the deployment phases. This could be a software element or

the final report. Deployment planning, monitoring, and maintenance comprise the deployment phase, as the user guide describes.
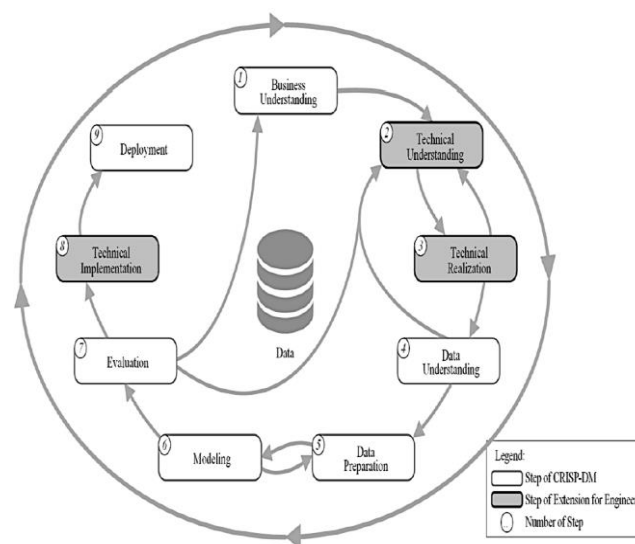

Figure 1. Research Stages based on the CRISP-DM Business Process [16].

The following is an implementation of research steps based on the methods used:

1. Knowledge of Business. At this point, Samsung Indonesia's cooperatives were receiving employee credit using the Decision Tree Algorithm, which allowed for the classification of data.
2. Technical Understanding. The data understanding phase in this research starts with data collection. This data understanding stage will involve understanding data needs related to achieving business goals. Employee credit history data was taken from 2021 and 2022, with a total of 1,083 (one thousand eighty-three) employees or customers.
3. Technical Realization. Technical realization is used to execute experimental programs and to test and choose measurement concepts. To meet the predetermined business objectives, the data generates all the features and information needed for ensuing data analysis tasks. At this point, the measurement concept is used to determine the best data acquisition method. An experimental plan is then implemented, and every step of the data acquisition process—including any

technical constraints, potential sources of error, and the quality of the data—is documented.

4. Data Understanding. Understanding data in practice uses hypotheses for information that is still unknown. In this data mining technique, it is formed based on experience and qualified assumptions.

5. Data Preparation. Data preparation is carried out using the preprocessing stage, namely data reduction and filtering, as well as feature creation. Additionally, data sets can be labeled according to the targeted knowledge, so that different stages and errors can be represented differently in the data set.

6. Modelling. At this modeling stage, the Decision Tree algorithm is applied to calculate the gain value of each attribute. The gain value calculation uses sample data, namely 10 records out of a total of 584 records.

7. Evaluation. In the evaluation phase, the trained model is tested against real data sets in production cases and the data mining results are assessed according to the determined business objectives.

8. Technical Implementation. The technical implementation allows evaluated models to be supplemented with run-time data during production. Therefore, the data acquisition method from technical realization in the DMME phase is transformed into a run-time capable infrastructure.

9. Deployments. After a successful evaluation, the trained model will be used in actual production to produce a better product than before. Better product results are also supported by historical data.

2.2. Employee Credit

Providing credit to employees in an organization or company is a form of retention program to retain potential employees which can also encourage loyalty and increase employee productivity. In an organization or company, employee credit is usually managed by a cooperative whose members are all employees of the organization or company [17]. This employee credit program certainly provides benefits for employees to obtain loans that are based on kindness and usefulness [18].

2.3. Decision Tree

Decision Tree is a powerful method commonly used in various fields, such as machine learning, image processing, and pattern identification. Decision Tree is a sequential model that combines a series of basic tests efficiently and cohesively where numerical features are compared with threshold values in each test [19]. Conceptual rules are much easier to construct than numerical weights in a neural network of connections between nodes, as seen in Figure 2. above. Mainly for grouping purposes, Decision Trees are used [20].
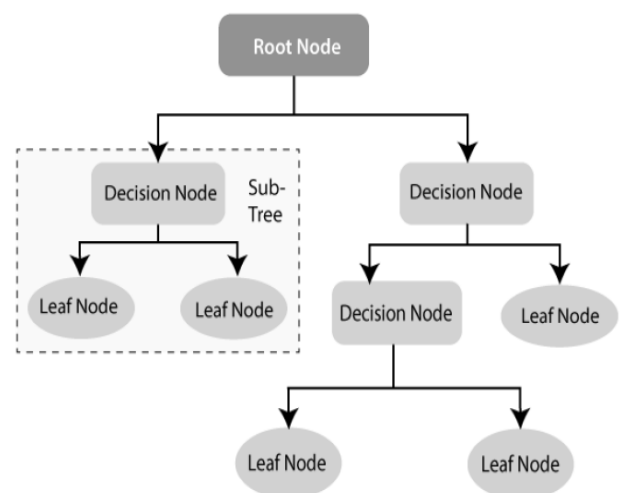

Figure 2. Decision Tree Structure [21]

Apart from that, Decision Tree is a classification model that is usually used in Data Mining. Nodes and branches consist of each Tree. Each node represents a feature in a category to be classified and each subset defines a value that can be retrieved by that node. Due to its simple analysis and accuracy on various forms of data, Decision Trees have found many areas of implementation [22]. To create a proper decision tree, Gain and Entropy values are required, which can be obtained as in Equations (1) and (2).

$$Gini\ Index = \ 1 - \sum (p_i)^2 \tag{1}$$

$$Entropy = \ -\sum p_i * \log_2(p_i) \tag{2}$$

where p_i represents the probability of an element belonging to a particular class.
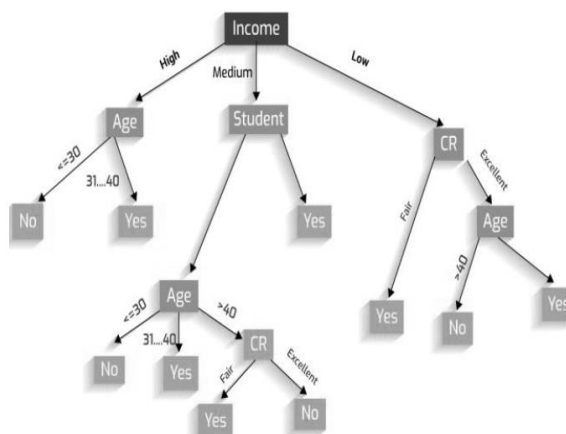
Figure 3. One of the implementations of using the Decision Tree Algorithm [23]

3. RESULTS AND DISCUSSION

3.1. Classification Results

The algorithm used in this research is the Decision Tree algorithm. The first step is to calculate the gain value of each attribute. The gain value calculation uses sample data, namely 10 records out of a total of 584 records (Table II). The use of sample data aims to facilitate the process of calculating the gain value for each attribute. Following are es of the Decision Tree algorithm calculation to determine the gain value (Gini index) for each attribute (Table III).

TABLE II. COOPERATIVE CUSTOMER SAMPLE DATA

No	Employee ID	Division	Position	Credit Status
1	IDAC-005332	HHP	SEC	Performing
2	AC-000677	TAB	SEC	Performing
3	IDAC-005382	HHP	TL	Non-Performing
4	IDAC-005331	HHP	SEC	Performing
5	KS-0002833	TAB	SEC	Performing
6	AC-000322	HHP	SEC	Performing
7	AC-000388	TAB	SEC	Performing
8	KS-0002852	TAB	GM	Performing
9	AC-000190	HHP	GM	Non-Performing
10	AC-000355	HHP	SEC	Performing

TABLE III. GINI VALUE OF EACH ATTRIBUTE

Attribute	Class		Gini Index value
	Performing	Non-Performing	
Id Karyawan			0.32
>IDAC-005332	8	2	
Gini Index (>AC-000355)	0.32		
Nama Karyawan			0
Division			0.26666667
HHP	6	2	
Gini Index (HHP)	0.66666666		
TAB	4	0	
Gini Index (TAB)	0		
Position			0.26666667
SEC	7	0	
Gini Index (SEC)	0		
TL	1	1	
Gini Index (TL)	0		
GM	2	1	
Gini Index (GM)	0		

The calculation of the gain value (Gini index) ends if the value of the attribute is already in the same class/label. The value for each attribute is in the same class/label (Figure 4). Then it is represented in full in Figure 5 in Complete Data Classification Mode.
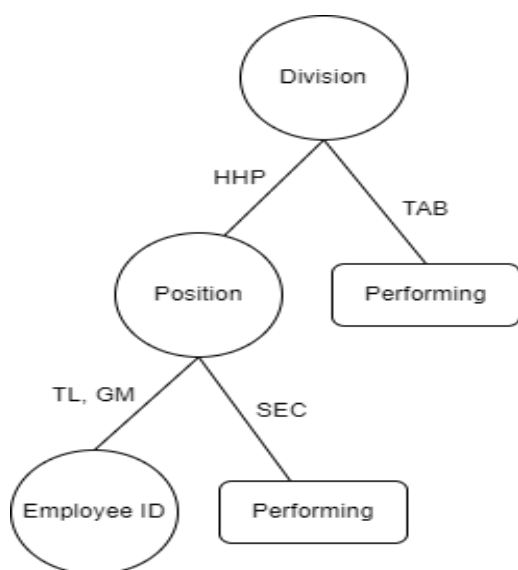
Figure 4. Sample Data Classification Model

Based on Figure 5, it is known that the root is Position. If an employee has an SEC position, it can be said that he has a Performing status. Meanwhile, if the position is TL or GM, it must be seen first then the Division needs to be seen. If it is in the HHP Division, then it can be said that the Credit Status is Performing. But if an employee is in the TAB Division, then the chance that his Credit Status is Performing is 2:1.
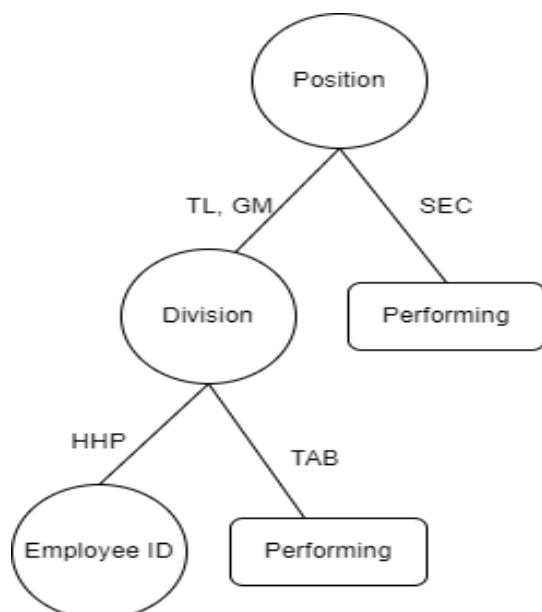

Figure 5. Full Data Classification Mode

3.2. Evaluation

The evaluation process carried out in this study is to measure the level of accuracy of the Decision Tree algorithm in modeling data for the creditworthiness of employee cooperatives. The use of RapidMiner tools helps the process of calculating the accuracy of the Decision Tree algorithm. The next evaluation process is obtained from the results of measurements and direct observations of employees at Samsung Indonesia Company starts from the legal team providing information regarding credit facilities which is continued by the preparation of Google Forms, which will later be forwarded to employees. After filling out Google Forms, employees will be given Terms and Conditions explaining the payment procedures and personal documents which will later be submitted to the company as collateral. From the measurements that have been carried out using sample data of 10 records from a total of 584 records, a classification model of 2 employees with non-performing employee credit status and 8 employees with performing employee credit status was obtained.

```
> gini_impurities(data_train1)[13:15]
          Var1           Var2 GiniImpurity
        <char>         <char>        <num>
1: CreditStatus  Employee ID   0.7000000
2: CreditStatus      Division   0.3428571
3: CreditStatus      Position   0.2714286
```

Figure 6. Calculation of Gini Index

Figure 6 is the calculation result of the Gini Index for each attribute. It is obtained that the Gini Index values of the Employee ID, Division, and Position attributes are 0.7, 0.3428571, and 0.2714286, respectively. Because the Position attribute has the smallest Gini Index, the decision to grant credit to employees depends on the Position, after that the Division, and the last is Employee ID.

4. CONCLUSION

The employee credit granting procedure at Samsung Indonesia Company has followed the standard credit granting procedure set with the aim of helping both in terms of process and results. Matters related to credit granting are regulated from various aspects taken, such as Submission of Google Forms Process, Share Google Forms to employees then data validation and the credit process can be approved by considering the credit repayment history in the

previous unit if any. The results of the study showed the results of the classification model of 2 employees with non-performing employee credit status and 8 employees with performing employee credit status. These results were carried out using a sample data of 10 records from a total of 584 records. From the results of the classification model, it was also obtained based on the Gini Index Value of the the Employee ID, Division, and Position attributes are 0.7, 0.3428571, and 0.2714286, respectively. So, the decision to grant credit to employees depends on the Position, after that the Division, and the last is Employee ID.

References:

[1] Samsung, "Inspiring and Empowering Women at Samsung: How Diversity, Equity and Inclusion Drive Innovation," https://news.samsung.com/global/inspiring-and-empowering-women-at-samsung-how-diversity-and-inclusion-drive-innovation. Accessed: Apr. 22, 2024. [Online]. Available: https://news.samsung.com/global/inspiring-and-empowering-women-at-samsung-how-diversity-and-inclusion-drive-innovation

[2] E. Retnowati, U. P. Lestari, Jahroni, D. Darmawan, and A. R. Putra, "Retensi Karyawan yang Ditinjau dari Kepercayaan dan Motivasi Kerja," *Jurnal Manajemen, Bisnis, dan Kewirausahaan*, vol. 1, no. 1, pp. 65–76, 2021, Accessed: Dec. 13, 2022. [Online]. Available: https://mada.indonesianjournals.com/index.php/mada/article/view/11

[3] N. Nurajijah and D. Riana, "Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah," *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 2, pp. 77–82, Apr. 2019, doi: 10.14710/jtsiskom.7.2.2019.77-82.

[4] N. Handayani, H. Wahyono, J. Trianto, and D. S. Permana, "Prediksi Tingkat Risiko Kredit dengan Data Mining Menggunakan Algoritma Decision Tree C.45," *Jurnal Riset Komputer)*, vol. 8, no. 6, pp. 2407–389, 2021, doi: 10.30865/jurikom.v8i6.3643.

[5] A. Faiz and Benyamin, "Penerapan Data Mining untuk Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4.5 pada Bank Bri Unit Anduonohu Kendari," *Router Research Jurnal Sistem Komputer dan Sistem Informasi*, vol. 1, no. 1, pp. 22–28, 2019, doi: https://doi.org/10.29239/j.router.2019.31 3.

[6] R. Kandayu, W. Silaban, M. Maulana, and E. Sastra Ompusunggu, "KLASIFIKASI KELANCARAN PEMBAYARAN PINJAMAN KOPERASI DENGAN METODE DECISION TREE," *Jurnal Sistem Informasi dan Ilmu Komputer Prima)*, vol. 5, no. 1, pp. 19–23, 2021.

[7] I. Riswanto and R. H. Laluma, "KLASIFIKASI KELAYAKAN PINJAMAN PADA KOPERASI KARYAWAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS WEB," *Jurnal Infotronik*, vol. 5, no. 1, pp. 11–16, 2020, doi: 10.32897/infotronik.2020.5.1.2.

[8] A. P. Wibawa, M. Guntur, A. Purnama, M. Fathony Akbar, and F. A. Dwiyanto, "Metode-metode Klasifikasi," in *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 2018, pp. 134–138.

[9] M. F. Arifin and D. Fitrianah, "Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada," *IncomTech, Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 2, pp. 87–102, 2018, doi: 10.22441/incomtech.v8i1.2198.

[10] N. Wahyuningsih and Hendry, "PERBANDINGAN METODE KLASIFIKASI DALAM ANALISIS SENTIMEN MASYARAKAT TERHADAP IDENTITAS KEPENDUDUKAN DIGITAL (IKD)," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 4, pp. 1218–1227, 2023, doi: 10.29100/jipi.v8i4.4155.

[11] R. Fatmasari, V. M. Ayu, H. Anto, W. Gata, and L. D. Yulianto, "Analisis Sentimen Dalam Pengkategorian Komentar Youtube Terhadap Layanan Akademik dan Non-Akademik Universitas Terbuka Untuk Prediksi Kepuasan," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, pp. 395–404, Sep. 2022, doi: 10.47065/bits.v4i2.1738.

[12] R. Sahara, S. Abdullah, and M. I. Saputra, "Integration Model of Academic Information Systems and Learning Management Systems with REST Web Services Using External Databases," *Journal of Information and Organizational Sciences*, vol. 47, no. 2, pp. 373–384, Dec. 2023, doi: 10.31341/jios.47.2.7.

[13] S. N. Luqman *et al.*, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," *Jurnal Sistem Cerdas*, vol. 04, no. 02, pp. 114–125, 2021,

Accessed: Dec. 07, 2023. [Online]. Available: https://apic.id/jurnal/index.php/jsc/articl e/view/162/86

[14] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.

[15] F. N. Dhewayani *et al.*, "Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *Jurnal Teknologi dan Informasi*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.

[16] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model," in *Procedia CIRP*, Elsevier B.V., 2019, pp. 403–408. doi: 10.1016/j.procir.2019.02.106.

[17] I. Rahmianti, "ANALISIS KELAYAKAN PEMBERIAN KREDIT KOPERASI DENGAN METODE DATA MINING DECISION TREE," *Jurnal Informatika & Rekayasa Elektronika)*, vol. 5, no. 2, pp. 153–161, 2022, doi: https://doi.org/10.36595/jire.v5i2.663.

[18] H. Hasugian, H. Mursyidin, and M. D. Handayani, "SISTEM PENUNJANG KEPUTUSAN PEMBERIAN KREDIT DENGAN METODE SIMPLE ADDITIVE WEIGHTING (SAW) STUDI KASUS: KOPERASI KARYAWAN GATERA PT PLN (PERSERO) AREA KEBAYORAN," in *Prosiding SINTAK*, 2018, pp. 465–471.

[19] A. Alibasyah, A. Ajiz, G. Dwilestari, and E. Wahyudin, "Penerapan Algoritma Decision Tree dalam Penentuan Karyawan Kontrak.," *MEANS (Media Informasi Analisa dan Sistem)*, vol. 7, no. 1, pp. 123–129, 2022, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_ Means/

[20] A. Setiawan, R. Febrio Waleska, M. Adji Purnama, Rahmaddeni, and L. Efrizoni, "KOMPARASIALGORITMAK-NEAREST NEIGHBOR(K-NN), SUPPORT VECTOR MACHINE(SVM), DAN DECISION TREEDALAM KLASIFIKASI PENYAKIT STROKE," *Jurnal Informatika & Rekayasa Elektronika)*, vol. 7, no. 1, pp. 107–114, 2024, doi: https://doi.org/10.36595/jire.v7i1.1161.

[21] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," May 01, 2021, *Springer*. doi: 10.1007/s42979-021-00592-x.

[22] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[23] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.