1278 Employee Credit Classification Analysis Using Decision Tree Based CRISP-DM Model (Case Study of Samsung Indonesia Company)

By Riad Sahara

WORD COUNT 2685 TIME SUBMITTED 20-AUG-2024 01:46PM

PAPER ID 111174540

Template ini berlaku untuk terbitan Volume 6 Nomor 1 April 2023

Employee Credit Classification Analysis Using Decision Tree Based CRISP-DM Model (Case Study of Samsung Indonesia Company)

Riad Sahara

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 PT. Samsung Indonesia. 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 bad employee credit status and 8 employees with smooth employee credit status was obtained. The classification model was obtained based on the Gini Index Value of the Employee ID attribute of 0.32, and the Gini Index Value of the Division and Position attributes of 0.26666667.

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

Abstrak

Program kredit karyawan merupakan bentuk dari retensi karyawan sebagai upaya untuk mempertahankan karyawan yang potensial dari perusahaan. Pada penerapannya pemberian kredit karyawan bukan tanpa risiko yang mungkin akan terjadi. Risiko tersebut seperti ketidakmampuan atau kegagalan atas kewajiban pembayaran angsuran kredit pada saat jatuh tempo. Untuk meminimalisir risiko yang mungkin akan terjadi, diperlukan survei dan analisis dengan metode yang tepat terhadap anggota koperasi sebelum diberikan kredit karyawan. Peneliti menggunakan algoritma berbasis *Decision Tree* sebagai alat bantu untuk pengambilan keputusan dalam pemberian kredit karyawan pada Koperasi di PT. Samsung Indonesia. Peneliti juga menggunakan model *CRoss-Industry Standard Process for Data Mining (CRISP-DM)* tentang siklus hidup pengembangan penambangan data sebagai langkah penelitian yang dilakukan. Model *CRISP-DM* ini sangat tepat digunakan karena merupakan model atau metode yang bersifat netral serta dapat digunakan dalam berbagai industri dan dikombinasikan dengan berbagai *tools*. Dari pengukuran yang sudah dilakukan dengan menggunakan *sample data* sebanyak 10 *records* dari total 584 *records*, didapatkan model klasifikasi 2 karyawan dengan status kredit karyawan macet dan 8 karyawan dengan status kredit karyawan lancar. Model klasifikasi tersebut didapatkan berdasarkan *Gini Indeks Value* dari atribut Id Karyawan sebesar 0.32, serta *Gini Indeks Value* dari atribut Divisi dan Posisi sebesar 0. 26666667.

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

1. Introduction

PT. Samsung Indonesia, 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. These risks include inability or failure to pay credit installments when they are due [2]. 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 [3]. 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 [4].

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 [5]. 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 [6]. Researchers will use a Decision Tree-based algorithm as a tool for decision making in granting employee credit to Cooperatives at PT. Samsung Indonesia. The Decision Tree algorithm has many advantages, especially for managing numeric and discrete data [7], as well as other advantages that make the performance of this algorithm the fastest when compared to other similar algorithms [8]. 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 [9] and can be used in various industries and combined with various tools [10] as well as appropriate ICT products for the needs of the financial sector, especially credit [11].

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 Figure 1. below this [12]. An industry-neutral process model for data mining is called the Cross-Industry Standard Process for Data Mining (CRISP-DM). 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 [13].

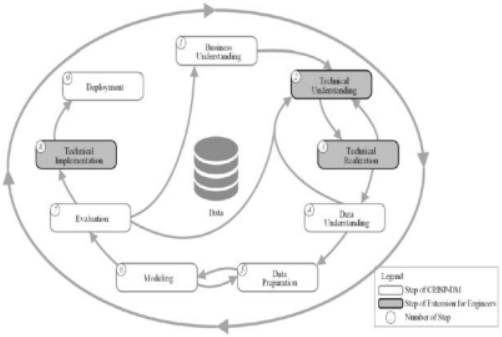

Figure 1. Research Stages based on the CRISP-DM Business Process [14].

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.

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

1. Knowledge of Business. At this point, PT. 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 pl 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 unk own. In this data mining technique, it is formed based on experience and qualified assumptions.
5. Data Preparation. Data preparation is carried o t using the preprocessing stage, namely data reduction and filtering, as ell 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 modeli 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 mplementation 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. This employee credit program certainly provides benefits for employees to obtain loans that are based on kindness and usefulness [16].

2.3. Decision Tree

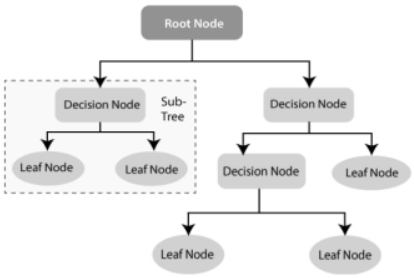

Figure 2. Decision Tree Structure [17]

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 [18]. 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. 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 [19].

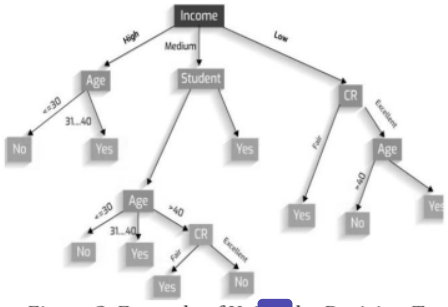

Figure 3. Example of Using the Decision Tree Algorithm [19]

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 the results 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	Name	Division	Position	Credit Status
1	IDAC-005332	Sri Wahyuni	HHP	SEC	Lancar
2	AC-000677	Deynada Syavira Putri	TAB	SEC	Lancar
3	IDAC-005382	Krisanda Alvasyera	HHP	TL	Macet
4	IDAC-005331	Sarwono	HHP	SEC	Lancar
5	KS-0002833	Mohammad Efendi	TAB	SEC	Lancar
6	AC-000322	Hanni Herlianti	HHP	SEC	Lancar
7	AC-000388	Martha Arista	TAB	SEC	Lancar
8	KS-0002852	Muhammad Ardiansyah	TAB	GM	Lancar
9	AC-000190	Gita AyuKusuma W.	HHP	GM	Macet
10	AC-000355	Regina OkyDelila	HHP	SEC	Lancar

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
Sri Wahyuni	1	0	
Deynada Syavira Putri	1	0	
Krisanda Alvasyera	0	1	
Sarwono	1	0	
Mohammad Efendi	1	0	
Hanni Herlianti	1	0	
Martha Arista	1	0	
Muhammad Ardiansyah	1	0	
Gita Ayu Kusuma W.	0	1	
Regina Oky Delila	1	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.

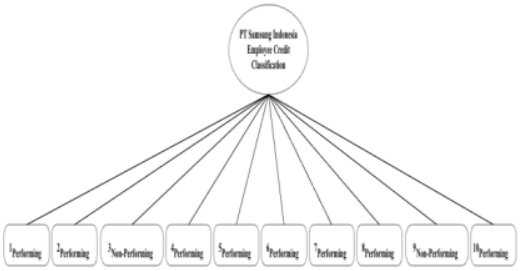

Figure 4. Sample Data Classification Model

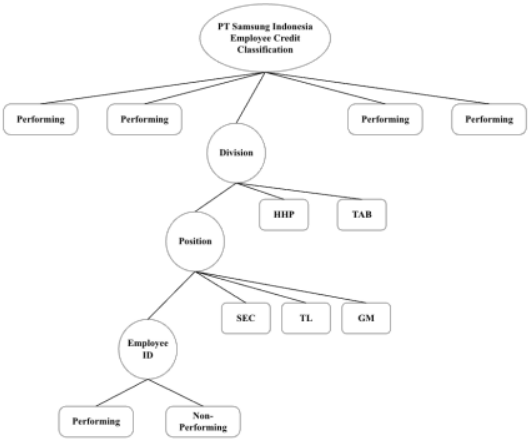

Figure 5. Full Data Classification Mode

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 PT employees. Samsung Indonesia 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 bad employee credit status and 8 employees with smooth employee credit status was obtained. The classification model was obtained based on the Gini Index Value of the Employee ID attribute of 0.32, and the Gini Index Value of the Division and Position attributes of 0.26666667.

4. Conclusion

The employee credit granting procedure at PT. Samsung Indonesia 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 bad employee credit status and 8 employees with smooth 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 Employee Id attribute of 0.32, and the Gini Index Value of the Division and Position attributes of 0. 26666667.

References:

1278 Employee Credit Classification Analysis Using Decision Tree Based CRISP-DM Model (Case Study of Samsung Indonesia Company)

ORIGINALITY REPORT

12%
SIMILARITY INDEX

PRIMARY SOURCES

1	ijeecs.iaescore.com Internet	66 words — 2%
2	journalofbigdata.springeropen.com Internet	43 words — 2%
3	Hendriyo Mokodompit, Nurnaningsih Nico Abdul, Elvie Fatmah Mokodongan. "PONDOK PESANTREN MODERN DARUL MADINAH WONOSARI KABUPATEN BOALEMO DENGAN PENDEKATAN ARSITEKTUR TROPIS", JAMBURA Journal of Architecture, 2024 Crossref	36 words — 1%
4	dataconomy.com Internet	28 words — 1%
5	Armando Plasencia Salgueiro, Ileana Dopico Mateo, Arlety García García, Ystria Rivero Ripoll, Ileana Suárez Blanco. "Chapter 18 Relevance of the National Program of Automatic, Robotics and Artificial Intelligence Projects in Applications", Springer Science and Business Media LLC, 2022 Crossref	24 words — 1%
6	Steffen Huber, Hajo Wiemer, Dorothea Schneider, Steffen Ihlenfeldt. "DMME: Data mining	21 words — 1%

methodology for engineering applications – a holistic extension to the CRISP-DM model", Procedia CIRP, 2019
Crossref

7	www.sciencegate.app Internet	19 words — 1%
8	link.springer.com Internet	16 words — 1%
9	www.ichpersd.org Internet	16 words — 1%
10	Jyoti B. Deone, Rahat Khan Afreen, Viraj R. Jadhao. "Techniques applied to increase soil fertility in smart agriculture", Elsevier BV, 2023 Crossref	14 words — < 1%
11	www.scirp.org Internet	9 words — < 1%
12	5dok.net Internet	8 words — < 1%
13	listens.online Internet	8 words — < 1%
14	www.open-access.bcu.ac.uk Internet	8 words — < 1%
15	repository.unair.ac.id Internet	7 words — < 1%
16	Zahra Maghsoodzadeh Sarvestani, Jasem Jamali, Mehdi Taghizadeh, Mohammad Hosein Fatehi Dindarloo. "A novel machine learning approach on texture analysis for automatic breast microcalcification diagnosis	6 words — < 1%

classification of mammogram images", Journal of Cancer Research and Clinical Oncology, 2023
Crossref

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF