Bit-Tech

Vol.2, No.3, Mei 2020
Available online at: http://jurnal.kdi.or.id/index.php/bt

Alleged Bad Credit at Saving Cooperatives Borrow Flamboyant Assistance PPSW Jakarta With Comparasion the Algorithms Naive Bayes and C4.5

Renaldi[1], Yusuf Kurnia[2]

[1) 2)]*Buddhi Dharma University*
Jl.Imam Bonjol No. 41 Karawaci Ilir, Tangerang, Indonesia

[1)]lim.renaldi123@gmail.com

[2)]yusuf.kurnia@ubd.ac.id

Article history:

Received 16 March 2020;
Revised 3 April 2020;
Accepted 8 April 2020;
Available online 30 Mei 2020

Keywords:

Data Mining
Comparison
Naïve Bayes
C4.5

Abstract

Data mining is often used in the financial sector, one of which is cooperatives. According to Law No. 25 of 1992, what is meant by cooperatives are business entities whose members are individual or cooperative legal entities based on activities based on the principles of cooperatives as well as as a people's economic movement based on the principle of kinship. One of the things that needs to be considered is the provision of credit or borrowing in the Flamboyan cooperative, which in this study there are many bad crediting occurrences that occur in the Flamboyan cooperative. By using a lot of data mining techniques, data can be utilized optimally. From the above problems, it can be overcome by utilizing data mining techniques, namely Predicting Bad Credit at the Flamboyant Savings and Loan Cooperative Fostered by PPSW Jakarta Using Comparative Algorithms Naive Bayes and C4.5. The algorithm used in the system is the best result of the Naive Bayes and C4.5 comparison based on data from the Flamboyan cooperative. The results obtained from the comparative data processing between the Naïve Bayes algorithm and the C4.5 using a dataset of 2282 transaction data obtained the results of the accuracy of the Naïve Bayes algorithm of 69.19% and the C4.5 algorithm of 71.87%, based on the accuracy results state that the C4 algorithm .5 is superior to the Naïve Bayes algorithm. Then the results from the C4.5 decision tree are translated into the bad credit prediction system in the Flamboyan cooperative.

I. Introduction

There are many definitions for the term Data Mining and none have been standardized or agreed upon by all parties. However, this term has the essence (Notion) as a scientific discipline whose main purpose is to discover, explore, or add knowledge from the data or information we have. This activity is the main concern or work of the data mining discipline[1]. Data Mining according to David Hand, Heikki Mannila and Padhraic from MIT is an analysis of data that is usually large in size to find clear relationships and conclude previously unknown in a way that is currently understood and useful for the owner of the data.

Data mining is often used in the financial sector, one of which is cooperatives. According to Law No. 25 of 1992, what is meant by cooperatives are business entities whose members are individual or cooperative legal entities based on activities based on the principles of cooperatives as well as as a people's economic movement based on the principle of kinship. Regulation of the Minister of Cooperatives and Small and Medium Enterprises (Permen Kop & UMKM) Number 15 / Per / M.KUKM / IX / 2015 which states that the KSP's own capital consists of principal savings, mandatory savings, reserves set aside from remaining business proceeds, grants and other savings that have the same characteristics as mandatory savings. A savings and loan cooperative is a cooperative that is engaged in the business of capital formation through the savings of members regularly and continuously for subsequent loans to members quickly, at low cost, facilitated and precisely for productive purposes and for welfare[2]. As a cooperative effort in supporting good service, an orderly, neat and thorough work procedure is needed so that it will produce information that is fast, accurate and timely as needed. Whereas in savings and loan cooperatives, of course there is a lot of data that increases every year, so that the data cannot be processed regularly and later it will only be an archive.

doi:

II. METHODS

In this study, the authors apply the Classification Method with the C4.5 algorithm in an application that can predict bad credit that will occur. In Data Mining there is what is called classification, the first classification applied to the field of plants that classifies a particular species, as was done by Carolus von Linne (also known as Carolus Linnaeus) who first classified species based on physical characteristics[3]. Classification is a job of assessing data objects to include them in a certain class from a number of classes that are willing. In classification, there are two main tasks that are carried out, namely the construction of the model as a prototype to be stored as memory and the use of the model to introduce / classify / predict other data objects so that it is known which class the data object is in the model that has been stored.

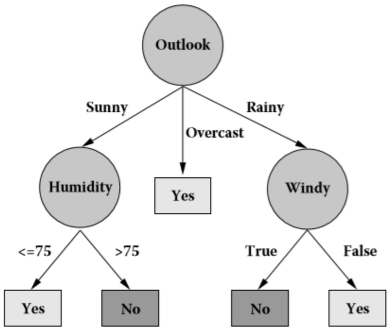


Fig 1. Example Of A Decision Tree

The image presents the classic "golf" dataset, bundled with the C4.5 installation. As previously stated, the aim is to predict whether the weather conditions on a pack can be used to play wolves. Note that some of these features are rated over time while others are categorical. The figure illustrates the tree induced by C4.5 as training data and error options[4].

Bayes is a simple probabilistic prediction technique based on the application of the Bayes theorem (or Bayes rule) with the assumption of strong (naive) independence. In other words, in naïve Bayes, the model used is "independent feature model". In Bayes (especially Naive Bayes), the meaning of strong independence on features is that a feature in a data is not related to the presence or absence of other features in the same data.

The relationship between Naive Bayes and classification, correlation hypothesis, and evidence with classification is that the hypothesis in the Bayes theorem is a class label that is input into the classification model. If X is an input vector containing features and Y is the class label, Naive Bayes is written as P (Y | X). The notation means the final probability (posterior probability) for Y, while P (Y) is called the initial probability (prior probability)[3].

Not data mining The name is if there is no data set that is processed in it. The word "data" in statistical terminology is a collection of objects with certain attributes, where the object is an individual form of data where each data has a number of attributes. These attributes affect the dimensions of the data, the more attributes / features the bigger the data dimensions. Data sets - the data make up the data set. In this book, sometimes mentioning data, sometimes mentioning vector, they both have the same meaning: Record: Data matrix, Transaction data, Document data. Graph: Word Wide Web (WWW), Struktur molekul. Ordered data set: Spatial data, Temporal data, Sequential data, Genetic sequence data.

Data sets containing data sets, with all data having the same number of numeric attributes (features), can be viewed as vectors (data) in a multidimensional area, where each dimension (feature) represents a different attribute that describes the object / data. Matrix data is the most common type of data record and is widely used in statistical applications [3].

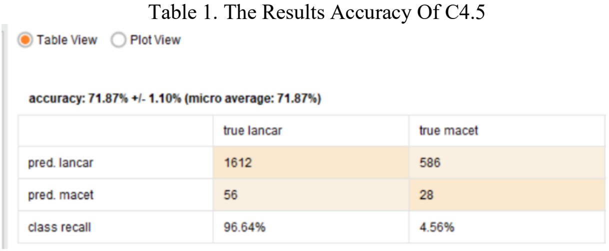
“Defining data mining as a process to get useful information from large database warehouses. The data we are talking about in the context of data mining is actually data that is managed in a database container”[4].

“The term Data Mining has several equivalents, such as knowledge knowledge or pattern recognition. The two terms actually have their own accuracy. The term knowledge discovery or knowledge discovery is appropriate because the main purpose of data mining is to get knowledge that is still hidden in chunks of data”[5].

III. RESULTS

From the comparison between the Naïve Bayes algorithm and C4.5 using a dataset of 2282 transaction data, the accuracy of the Naïve Bayes algorithm is 69.19% and the C4.5 algorithm is 71.87%. Based on the accuracy results, it states that the C4.5 algorithm is superior to the Naïve Bayes algorithm.

Table 1. The Results Accuracy Of C4.5

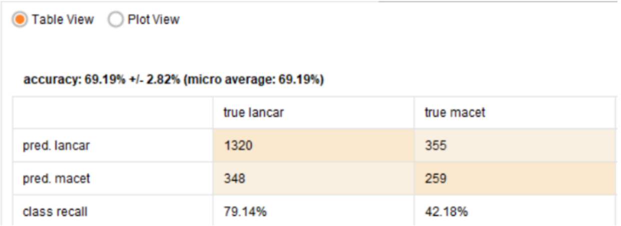


● Table View ○ Plot View

accuracy: 71.87% +/- 1.10% (micro average: 71.87%)

	true lancar	true macet
pred. lancar	1612	586
pred. macet	56	28
class recall	96.64%	4.56%

Table 2. The Results Accuracy Of Naïve Bayes



● Table View ○ Plot View

accuracy: 69.19% +/- 2.82% (micro average: 69.19%)

	true lancar	true macet
pred. lancar	1320	355
pred. macet	348	259
class recall	79.14%	42.18%

From the results of the comparison between C4.5 and Naive Bayes, the best accuracy is won by the C4.5 algorithm, therefore the authors use the best algorithm into manual calculations and the system created. C4.5 is part of the algorithm for classification in machine learning and data mining. The C4.5 algorithm uses the concept of information gain and entropy reduction to select the optimal division and produce a decision tree. The stages in making a decision tree using the C4.5 algorithm are as follows:

Prepare data, In this study the authors used 60 training data which had all classification conditions taken randomly from the data set.

Calculating Entropy, Determine the root of the tree by calculating the highest gain value of each attribute based on the lowest entropy index value. Previously, the entropy index value was calculated using the formula:

$$\text{Entropi } (S) = \sum_{j=1}^{k} - p_i \,.\, log_2 . p_i$$

Information:
S : case set
k : number of partitions i
pi : probability obtained from Sum

Calculating Gain, Calculate the gain using the formula:

$$Gain\ (S, A) = Entropy\ (S) - \sum_{i=1}^{n} *\ Entropy\ (Si)$$

To get a decision tree from the data used, calculations must be carried out by implementing the Entropy and Gain calculation formulas listed above, which are as follows:

At this stage it is a calculation to determine which attribute condition first determines which attribute has the highest gain. The results of Node 1 show that the greatest gain occurs in the loan attribute, so the loan is the first leaf in the decision tree.

Total Entropy Calculation,

$$Entropy\ (Total) = \left(-\frac{49}{60} * log_2\left(\frac{49}{60}\right)\right) + \left(-\frac{11}{60} * log_2\left(\frac{11}{60}\right)\right) = 0.687315093$$

Entropy Calculation of Work Status attribute,

$$Entropy\ (Status\ Kerja\ , Y) = \left(-\frac{26}{33} * log_2\left(\frac{26}{33}\right)\right) + \left(-\frac{7}{33} * log_2\left(\frac{7}{33}\right)\right) = 0.745517843$$

$$Entropy\ (Status\ Kerja\ , N) = \left(-\frac{5}{5} * log_2\left(\frac{5}{5}\right)\right) + \left(-\frac{0}{5} * log_2\left(\frac{0}{5}\right)\right) = 0$$

$$Entropy\ (Status\ Kerja\ , TJ) = \left(-\frac{18}{22} * log_2\left(\frac{18}{22}\right)\right) + \left(-\frac{4}{22} * log_2\left(\frac{4}{22}\right)\right) = 0.684038436$$

Calculation of Work Status Gain

$$Gain\ (Status\ Kerja) = 0.687315093 - \left(\left(\frac{33}{60} * 0.746\right) + \left(\frac{5}{60} * 0\right) + \left(\frac{22}{60} * 0.684\right)\right) = \mathbf{0.026466186}$$

From the results of the node calculation example above, the following decision tree results are created:

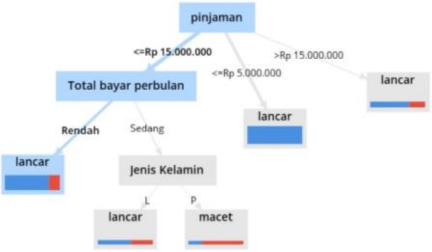
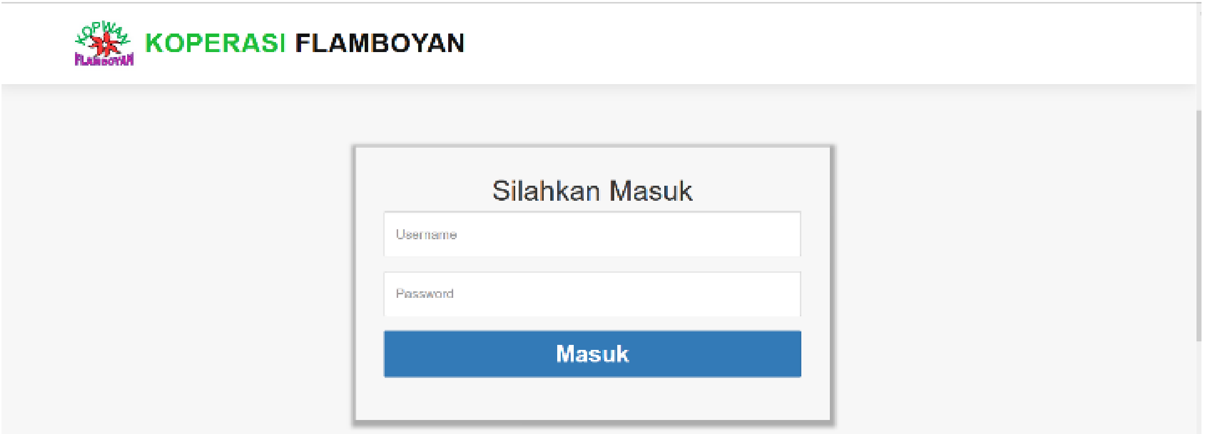


Fig 2. Rapid Miner Decision Tree Results

Login View



KOPERASI FLAMBOYAN

Silahkan Masuk

Username

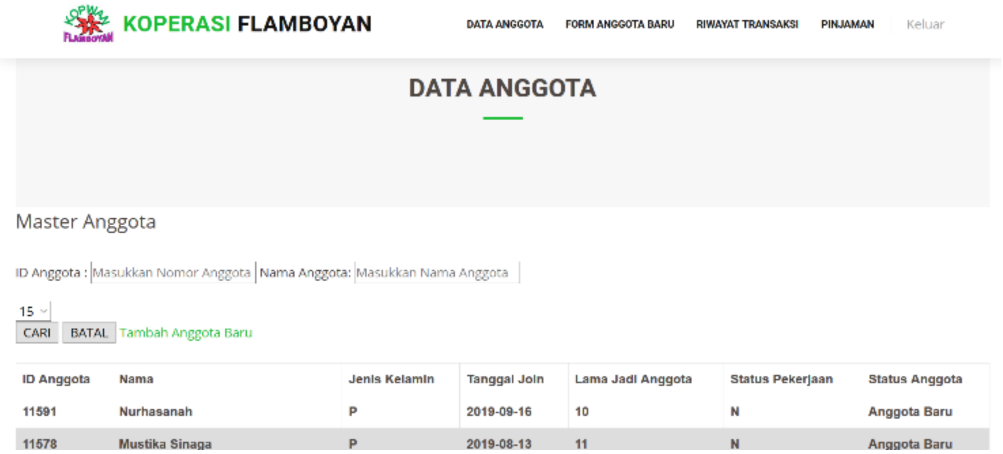
Password

Masuk

Fig 3. Login View

This login display is a display that will be used to enter the flamboyant cooperative credit prediction system. Where the system can help estimate cooperative borrowers.

Member Data Page View



KOPERASI FLAMBOYAN

DATA ANGGOTA | FORM ANGGOTA BARU | RIWAYAT TRANSAKSI | PINJAMAN | Keluar

DATA ANGGOTA

Master Anggota

ID Anggota : Masukkan Nomor Anggota | Nama Anggota: Masukkan Nama Anggota

15

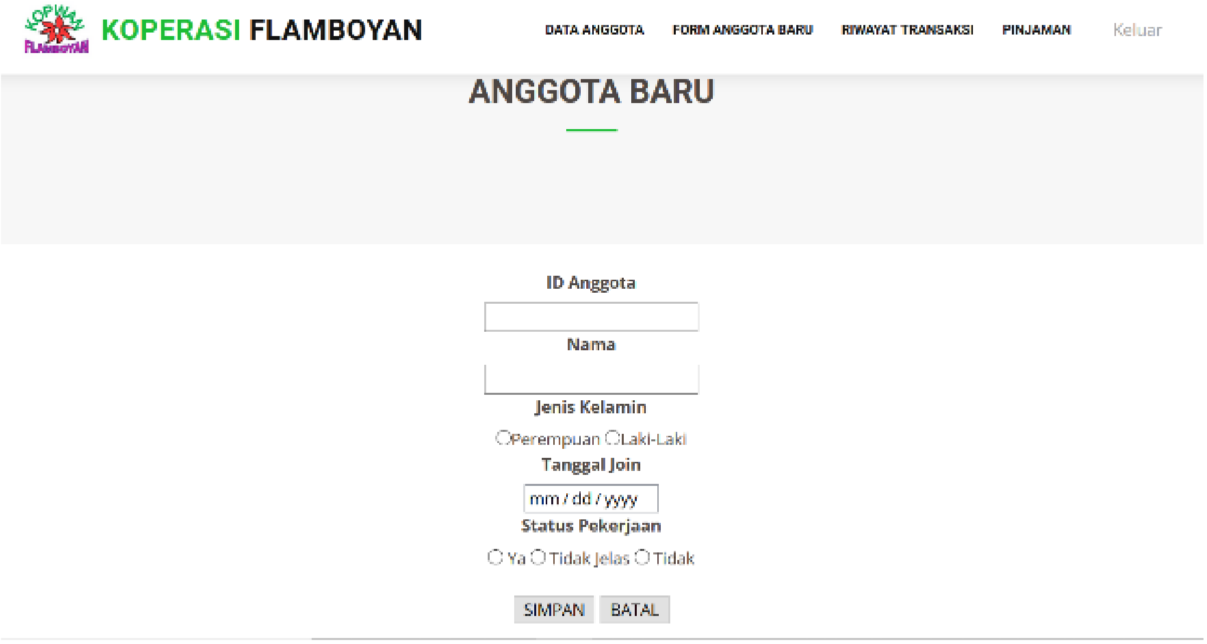
CARI | BATAL | Tambah Anggota Baru

ID Anggota	Nama	Jenis Kelamin	Tanggal Join	Lama Jadi Anggota	Status Pekerjaan	Status Anggota
11591	Nurhasanah	P	2019-09-16	10	N	Anggota Baru
11578	Mustika Sinaga	P	2019-08-13	11	N	Anggota Baru

Fig 4. Member Data Page View

Member Data Page Views are views to view cooperative member data, data that can be displayed such as name, gender, date of joining, length of time as a member, employment status and member status. in this view can also perform member data search.

New Member Form Page View



ANGGOTA BARU

ID Anggota

Nama

Jenis Kelamin

○Perempuan ○Laki-Laki

Tanggal Join

mm / dd / yyyy

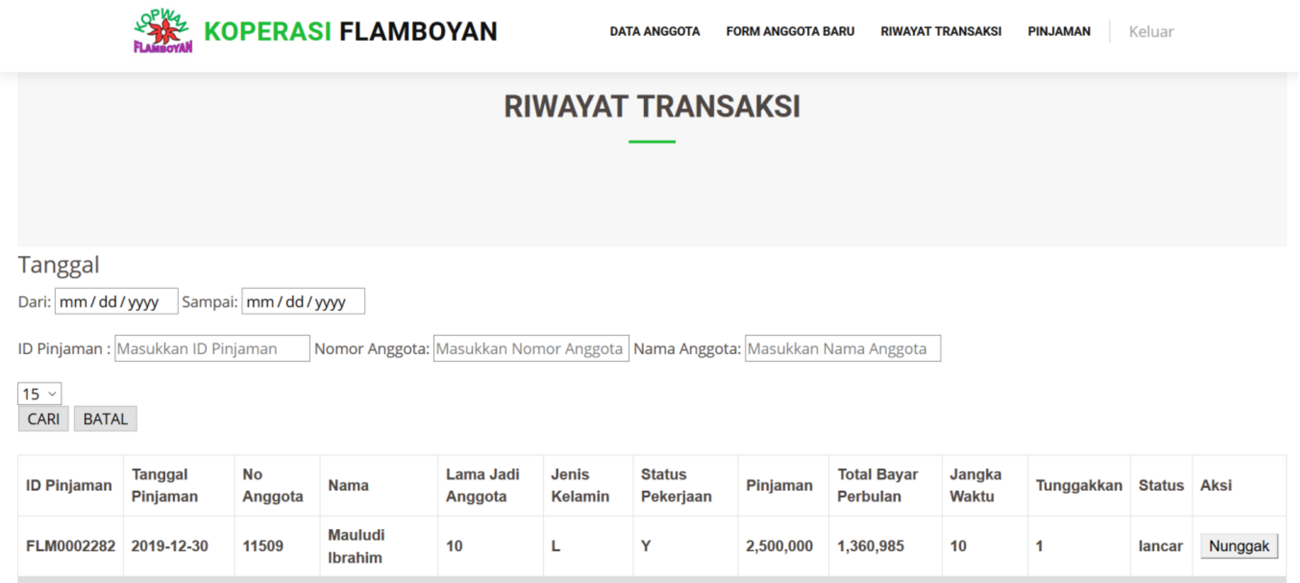
Status Pekerjaan

○ Ya ○ Tidak Jelas ○ Tidak

Fig 5. New Member Form Page View

This page functions to input data for prospective new cooperative members. where inputted data such as: member ID, name, gender, date of joining and employment status.

Transaction History Page Display



RIWAYAT TRANSAKSI

Tanggal

Dari: mm / dd / yyyy Sampai: mm / dd / yyyy

ID Pinjaman : Masukkan ID Pinjaman Nomor Anggota: Masukkan Nomor Anggota Nama Anggota: Masukkan Nama Anggota

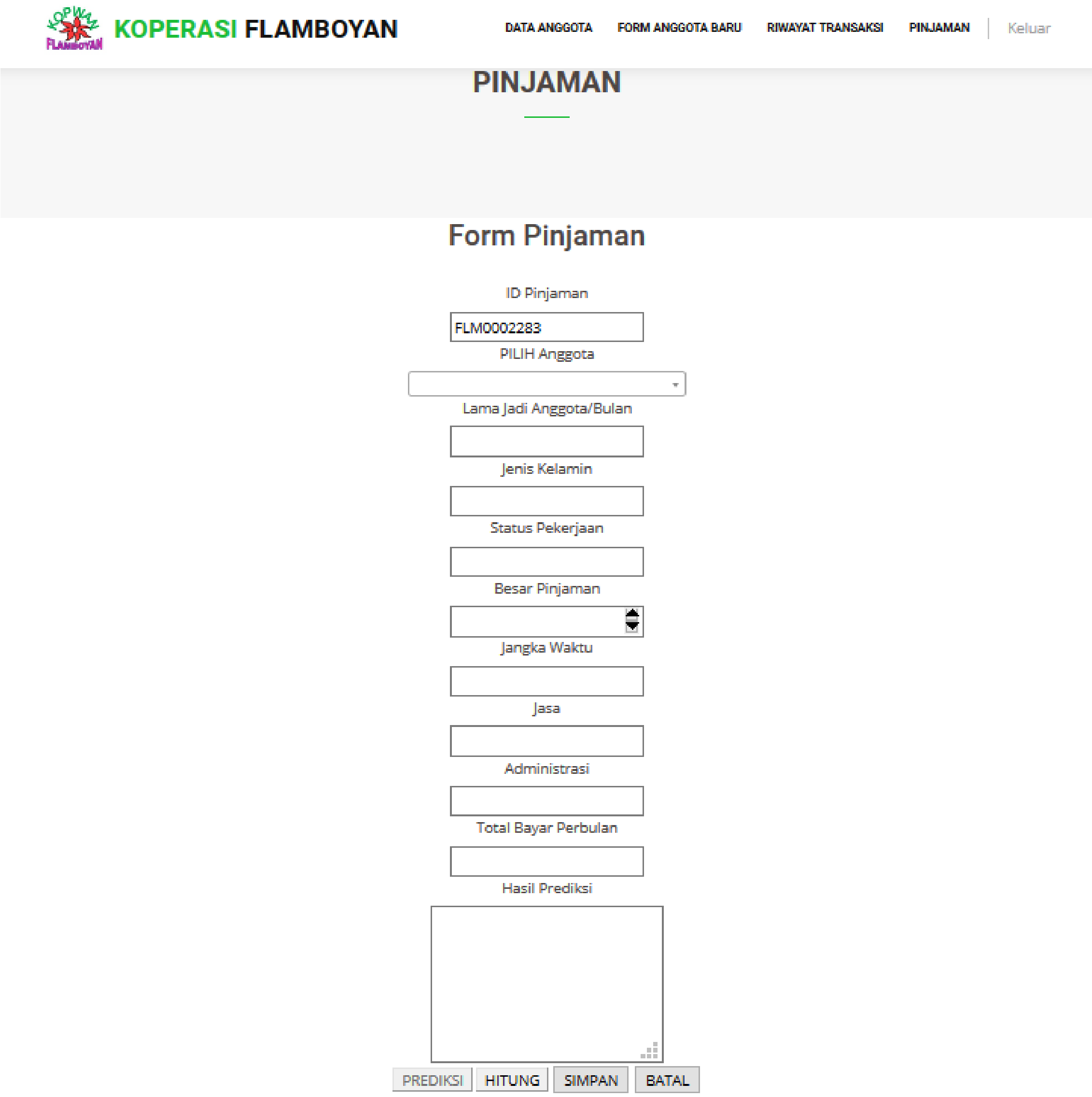
15

ID Pinjaman	Tanggal Pinjaman	No Anggota	Nama	Lama Jadi Anggota	Jenis Kelamin	Status Pekerjaan	Pinjaman	Total Bayar Perbulan	Jangka Waktu	Tunggakkan	Status	Aksi
FLM0002282	2019-12-30	11509	Mauludi Ibrahim	10	L	Y	2,500,000	1,360,985	10	1	lancar	Nunggak

Fig 6. Transaction History Page Display

This page can be used to find out the transaction history of cooperative members. the transaction in question is a borrowing transaction and payment history every month, whether the borrower is in arrears or not.

Loan Page View



PINJAMAN

Form Pinjaman

ID Pinjaman: FLM0002283
PILIH Anggota:
Lama Jadi Anggota/Bulan:
Jenis Kelamin:
Status Pekerjaan:
Besar Pinjaman:
Jangka Waktu:
Jasa:
Administrasi:
Total Bayar Perbulan:
Hasil Prediksi:

Fig 7. Loan Page View

This display functions to determine whether the cooperative member or prospective borrower will be in arrears or not. All prospective borrowers will have their data filled in first, data to be filled in such as member ID, loan size and loan term. After filling in the data, the system can predict whether the prospective borrower will commit arrears or not.

IV. Conclusions

Based on the results obtained from the analysis and discussion, the authors obtain conclusions that can be drawn from this study, this bad credit prediction system is made from the best algorithm results between the comparison of the Naive Bayes algorithm with an accuracy of 69.15% and C4.5 with an accuracy of 72%. , this web-based bad credit prediction system has been successfully established and is able to provide an accurate prediction of bad credit.

This bad credit prediction system is used as a decision support system for the Flamboyan cooperative in providing loans to members, so that bad credit that occurs in the Flamboyan cooperative can be minimized, this bad credit prediction system is easy to operate because it has a user friendly appearance and process.

REFERENCES

[1] Dedy Suryadi dan Sani Susanto. 2010. Pegantar Data Mining Menggali Pengetahuan dari Bongkahan Data. Yogyakarta : Andi.

[2] Ninik Widiyanti dan Sunindhia, 2009, Koperasi dan Perekonomian Indonesia, Jakarta, Rineka Cipta..

[3] Prasetyo, Eko. 2012. Data Mining Konsep Dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi.

[4] Xindong Wu dan Vipin Kumar. 2009. The Top Ten Algorithms in Data Mining. Minnesota USA : Chapman & Hall/CRC.

[5] Finn Lee S dan Juan Santana. 2010. Data Mining : Meramalkan Bisnis Perusahaan. Jakarta: PT. Elex Media Komputind.