Creditworthiness system with k-means model based z-score

D S Maylawati1,2*, Y A Gerhana1, S C Kusuma1, W B Zulfikar1 and M A Ramdhani1

1Department of Informatics, UIN Sunan Gunung Djati, Bandung, Indonesia
2Department of Informatics, Sekolah Tinggi Tekologi Garut, Garut, Indonesia

*Corresponding author’s email: dsaadillah@sttgarut.ac.id

Abstract. Evaluation of the creditor's ability needs to be done before the debtor provides the loan. The purpose of system design in this study is to establish a creditworthiness system by implementing K-Mean based Z-Score. The method of system design that used in this research is Rational Unified Process (RUP). The case study of those research is in the Savings and Loans Cooperative (KSP) at Junior High School 1st Cileunyi, Bandung. K-Means method-based Z-Score is used by the author as a model for solving problems, wherein the author uses three clusters with four parameters that applied to the method. Based on the results of the experiments, the application can classify prospective credits into three groups based on the four parameters. Those three groups among others: "OK" as many as 37 members, "Warning" as many as 28 members and "Bad" as many as 5 members. The results of those groups are used as a credit recommendation to the KSP.

1. Introduction

The Indonesian Cooperative (Koperasi) is a business entity whose members are individuals or cooperative legal entities by basing their activities based on cooperative principles, as well as the people's economic movement based on the principle of kinship, based on Act. No. 25 of 1992 Article 1 paragraph 1 concerning Cooperatives

In the implementation of lending, cooperatives often experience congestion in payments and loans. One of them happened at Saving and Loan Cooperative (KSP) SMP (Junior High School) 1 Cileunyi, which was then used as an object of research. At KSP SMP 1 Cileunyi, there were problems with several debtors who unable to repay the loan. From the data that collected from KSP SMP 1 Cileunyi, as many as 50% of the number of cooperative members have a history of poor instalments. 22% of them did not pay in instalments, 41% were late in instalments, and 15% paid in nominal terms under the provisions so that the distribution of loan funds for other members was hampered. Factors that can influence that problem are the weakness of the KSP management in selecting members for the ability to repay loans.

To solve that problem, information about the ability profile of the debtor is needed to be analyzed. This research develops Decision Support System (DSS) using the Z-Score-based K-Means Method. This method is used because it is easy to adapt, implement and run. In addition, this method is a development of the basic K-Means Algorithm that combines basic K-Means with the Z-Score calculation for each attribute. This combination is conducted because the standard K-Means Algorithm does not guarantee produce a good results, in terms of accuracy in the final cluster depends on the selection process in the initial centroid [1].

Previous studies which are similar to this research, among others: DSS for Potential Customers Using the K-Means Method on the “Tekad Usaha Mandiri” Savings and Loans Cooperative in Boyolali [2];
the implementation of K-Means clustering method for grouping competency-based student graduation [3]; Clustering drop out potential students using the K-Means method [4]; assessment of cooperative financial performance at the Pamekasan Cooperative and UMKM Office using K-Means algorithm [5]; and market segmentation analysis based on the behaviour characteristics of credit takers’ customers in Citra Abadi Savings and Loans (KSP) in Babat Lamongan [6].

2. Methods
The application design method for this study uses the Rational Unified Process (RUP) model approach. RUP is an software development methodology that collect various best practices in the software development area [7]. The main character of this approach is using use-case driven and iterative cycles. RUP uses object-oriented model with its activities focus on the use of Unified Model Language (UML) as an analysis and design modelling. To reach the reliability of system, this research used analytical, logical, conceptual, and operational verification [8].

3. Result and Discussion
Technology are made to enhance the quality of human life [9], information system can be used to increase the efficiency and effectivity of savings and credit. Because, information systems are a combination of information technology and the activities of people that that utilized computerized technology [10], which is generally used to support management and operational of institution or industry [11].

Information system has organized manner to manage the data [12], that based on user requirements [13], information systems have a good flexibility that allows to be developed [14], and able to solve complex data processing problems [15]. One of information system that intended to support decision making is the Decision Support System (DSS) [16]. In many previous research, information systems have several advantages, among others: good data accessibility [7], accurate [17], decision support precisely [18], more economical [19], efficient in time [20], widely used [21], improve user understanding [22], provide data and information well [23], improve productivity [24], and as data storage medium [25]. This study uses Z-Score-based K-Means method to analyse the creditor status through grouping the parameter values of each attribute of each cooperative member to determine the ability of the debtor to repay the loan.

3.1. K-Means Clustering based on Z-Score
Clustering is a collection of random data that has similar characteristics so that can be grouped and can be a DSS which is a computer-based system, clustering aims to gain knowledge so that it helps make decisions [26]. The purpose of clustering is to make a group of several data that has characteristics as closely as possible but the data have different characteristics with other groups [27]. One of clustering method is K-Means algorithm which is an unsupervised learning method. Unsupervised learning is an unattended learning, one of which is clustering that does not have class or data label. So, the data will be grouped based on the size of the similarity between data. It is different with supervised learning which is attended learning process, such as classification that must have a data label so that it can only be classified [28].

The K-Means algorithm was first discovered by several researchers through several disciplines, the most dominant being Lloyd (1957, 1982), Forrey (1965), Friedman & Rubin (1967), and McQueen (1967) [29]. The purpose of the K-Means algorithm is to separate a set of data into clusters or groups, where the number of k has been determined at the beginning. The steps of clustering data using z-score-based k-means are as follows [1]:

**Algorithm 1:** Find initial centroid

Input: D = {d₁, d₂, ..., dₙ} // collection of n data  
  k // number of clusters

Output: Series of k-initial centroid

Steps:
    a. Calculate Z-Score value for each data;
    b. Sort the data based on Z-Score value;
c. Separate data into k subset;
d. Calculate the average value for each subset;
e. Take one of the data values that closest to the average as the initial centroid for each subset.

**Algorithm 2: K-Means algorithm**

**Input:** D = \{d_1, d_2, ..., d_n\} // collection of n data

k // number of clusters

**Output:** Series of k-cluster.

**Steps:**

a. Use the result of algorithm 1 as an initial centroid of cluster;
b. Repeat;
c. Set each data in the cluster that has the nearest centroid;
d. Calculate the average value if each cluster until the data does not move cluster.

For the needs of status of credit debtor’s analysis, as the aim in this study, a software design flowchart is prepared as shown in Figure 1. Implementation of Z-Score-based K-Means Clustering considers the deposit nominal, the amount of late payment, net salary, and the number of dependents can affect the output. The nominal deposit is used because the funds that saved can expedite cash flow to other members who want to borrow [30]. Late in paying installments are taken from compulsory payments such as compulsory contributions and instalments. The more late it is, the more broken the promise of the credit that the cooperative has given [31]. Net salary also affects the ability of prospective borrowers to make instalments on loans [32]. Finally, the number of family dependents related to capacity, the more dependents, the smaller the capacity of prospective credit recipients to make credit instalments [33].

![Flowchart of Clustering K-Means based on Z-Score Implementation](image)

**Figure 1. Flowchart of Clustering K-Means based on Z-Score Implementation**

All parameter values are calculated using the z-score with the equation (1):

$$Z_{score} = \frac{X - \bar{x}}{StDev}$$  \hspace{1cm} (1)

Where, X represents the data being tested, \(\bar{x}\) shows the average and StDev shows the standard deviation (equation (2))

$$StDev: \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$  \hspace{1cm} (2)

Where, n is sample size; \(x_i\) is x value begin from i; and \(\bar{x}\) is average value.

3.2. **Application Design**

Application is designed so that follow the procedures below:
a. Log in to Admin page.
b. Register members by filling in the biodata along with the net salary and the number of family dependents.
c. Fill the mandatory initial deposit and initial voluntary savings as a registration fee.
d. If members submit a credit application, the system will calculate the late payment per month that the admin inputs every time a member pays instalments.
e. The system calculates the delay in payment of mandatory member deposits that entered by the admin in each month.

In this study, system testing is conducted using 17 data sample of member that described in Table 1.

Table 1. Data Sample for System Testing

| Name       | Total Late (day) | Savings (IDR) | Family Dependent | Net Salary (IDR) | Salary Cut (IDR) | Gross Salary (IDR) | Net Salary Ratio |
|------------|------------------|---------------|------------------|------------------|------------------|-------------------|-----------------|
| Applicant 1 | 0                | 2,710,000     | 1                | 38,800           | 5,250,900        | 5,289,700         | 0.99266         |
| Applicant 2 | 0                | 8,570,000     | 1                | 2,427,300        | 2,222,700        | 4,650,000         | 0.478           |
| Applicant 3 | 0                | 5,280,000     | 1                | 2,589,300        | 2,409,300        | 4,998,600         | 0.48199         |
| Applicant 4 | 244              | 4,270,000     | 2                | 2,481,200        | 2,381,800        | 4,863,000         | 0.48978         |
| Applicant 5 | 0                | 5,045,000     | 1                | 2,425,300        | 2,181,700        | 4,607,000         | 0.47356         |
| Applicant 6 | 64               | 4,270,000     | 2                | 2,481,200        | 2,381,800        | 4,863,000         | 0.48978         |
| Applicant 7 | 40               | 5,045,000     | 1                | 2,425,300        | 2,181,700        | 4,607,000         | 0.47356         |
| Applicant 8 | 220              | 5,200,000     | 1                | 1,403,800        | 3,751,200        | 5,155,000         | 0.72768         |
| Applicant 9 | 274              | 2,461,000     | 4                | 476,000          | 4,506,900        | 4,982,900         | 0.90447         |
| Applicant 10| 303              | 7,641,300     | 2                | 1,401,000        | 3,470,000        | 4,871,000         | 0.71237         |
| Applicant 11| 154              | 5,110,700     | 3                | -282,800         | 5,298,400        | 5,015,600         | 1.05638         |
| Applicant 12| 40               | 5,285,000     | 2                | 721,700          | 4,124,700        | 4,846,400         | 0.85108         |
| Applicant 13| 14               | 5,896,000     | 1                | 218,000          | 4,780,900        | 4,998,900         | 0.95639         |
| Applicant 14| 192              | 4,275,000     | 1                | 162,000          | 4,756,500        | 4,918,500         | 0.96706         |
| Applicant 15| 560              | 4,580,000     | 3                | 2,381,100        | 2,082,900        | 4,464,000         | 0.46659         |

The ratio used to avoid net salary is zero and to be in line with the other determinants (column 8). Where, the ratio of net salary is obtained from salary deductions divided by gross salary. Next, the total of Z-Score value which is calculated using equation (1). The example of calculation of Z-Score value results presented in Table 2.

Table 2. Z-Score Value for Applicant 1

| No | Item         | Total of Z-Score Value |
|----|--------------|------------------------|
| 1  | Late         | -0.875                 |
| 2  | Savings      | -0.9359351             |
| 3  | Family Dependant | -0.9870747       |
| 4  | Net Salary   | 1.1309272              |

After the Z-Score of each applicant data is calculated on each determinant, then sort the data based on the value for each determinant, then divide it into 3 subsets with the same amount of data. The average values is calculated for each subset and select the data that has the closest value to the average subset as initial centroid. The 'Nearest Data' column shows the selected data as initial centroid based on the closest Z-Score value to the average for each subset. For the example, Table 3 shows the process and the result of finding initial centroid for total late data.
Table 3. Initial Centroid for Total Late Data

| Subset | Name          | Z-Score Value | Average Value | Closest Data Value |
|--------|---------------|---------------|---------------|--------------------|
| 1      | Applicant 01  | -0.87830      | -0.86328      | -0.87830           |
|        | Applicant 02  | -0.87830      |               |                    |
|        | Applicant 03  | -0.87830      |               |                    |
|        | Applicant 05  | -0.87830      |               |                    |
|        | Applicant 06  | -0.87830      |               |                    |
|        | Applicant 15  | -0.78816      |               |                    |
| 2      | Applicant 08  | -0.62076      | -0.12284      | 0.11324            |
|        | Applicant 14  | -0.62076      |               |                    |
|        | Applicant 07  | -0.46623      |               |                    |
|        | Applicant 13  | 0.11324       |               |                    |
|        | Applicant 16  | 0.35791       |               |                    |
|        | Applicant 11  | 0.49956       |               |                    |
| 3      | Applicant 09  | 0.53819       | 1.18334       | 1.07260            |
|        | Applicant 04  | 0.69272       |               |                    |
|        | Applicant 10  | 0.88588       |               |                    |
|        | Applicant 12  | 1.07260       |               |                    |
|        | Applicant 17  | 2.72733       |               |                    |

The same thing (nearest data) is calculated based on the attributes of savings, total family dependents, and net salary. Data that has the closest distance to the average of subset is made as an initial centroid and the process of calculating the distance of the data with the centroid is done using the Euclidean Distance Space (Euclidean) formula in equation (3) [26].

$$D(X_2, X_1) = \sqrt{\sum_{j=1}^{p} (X_{2j} - X_{1j})^2}$$  \hspace{1cm} (3)

For the example of manual calculation for data distance to centroid of Applicant 1, then it is found value of Cluster 1 is 1.957145, Cluster 2 is 1.895186, and Cluster 3 is 3.697827. Those calculation result show that data distance is closest with Cluster 2, so that the Applicant 1 is grouped in Cluster 2. The calculation is applied for all applicant data. After the data distance is calculated and the data is grouped by the closest centroid, then the new centrois is calculated with summing the determinant values of each cluster member and dividing it by many members in the cluster. The result of clustering all of applicant data is shown in Table 4, where e1, e2, and e3 is euclidean distance value to centroid, then C1, C2, and C3 is cluster one until cluster 3.

Table 4. Grouping Data into Cluster

| Applicant | e1    | e2    | e3    | C1   | C2   | C3   | Applicant | e1    | e2    | e3    | C1   | C2   | C3   |
|-----------|-------|-------|-------|------|------|------|-----------|-------|-------|-------|------|------|------|
| 1         | 1.95714 | 1.89518 | 3.69782 | 10   | 2.01158 | 0.52655 | 2.44013 | 2.37053 | 2.85274 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |
| 2         | 1.16772 | 1.53311 | 3.76892 | 11   | 3.43151 | 2.65724 | 3.16331 | 2.30536 | 2.85274 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |
| 3         | 2.00356 | 2.04706 | 3.93599 | 12   | 3.18650 | 1.96982 | 2.20357 | 3.56361 | 2.79091 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |
| 4         | 3.56361 | 2.38094 | 2.79091 | 13   | 3.18650 | 1.96982 | 2.20357 | 3.56361 | 2.38094 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |
| 5         | 2.37053 | 1.63276 | 2.85274 | 14   | 3.18650 | 1.96982 | 2.20357 | 3.56361 | 2.38094 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |
| 6         | 4.06174 | 3.28712 | 3.56722 | 15   | 3.18650 | 1.96982 | 2.20357 | 3.56361 | 2.38094 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |
| 7         | 2.40935 | 1.45994 | 2.87697 | 16   | 3.18650 | 1.96982 | 2.20357 | 3.56361 | 2.38094 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |
| 8         | 1.89139 | 1.56024 | 3.50970 | 17   | 4.23815 | 2.99866 | 2.66384 | 2.30536 | 2.85274 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |
| 9         | 2.05418 | 0.90797 | 2.86452 | 18   | 3.18650 | 1.96982 | 2.20357 | 3.56361 | 2.38094 | 3.16331 | 2.47627 | 3.58798 | 3.16331 |

The parameters value are processed and transformed into a z-score to find the initial centroid then the K-Means process will be started. The results of the Z-Score K-Means process are divided into three clusters, among others: not feasible (Bad), not too feasible (Warning), and feasible (OK). This information becomes a recommendation that given by the system to the cooperative management.

3.3. Testing
The results test process provides the information about the application success in conducting clustering of cooperative applicant data. The data is collected from KSP SMPN 1 Cileunyi note. Testing process
uses 77 cooperative member data with four parameters that influenced the clustering process. The result of testing is conducted by knowing how much data can be grouped into the cluster or not. From the results of the experiment show that 31 members were in 'OK' cluster; 33 members were in 'Warning' cluster; 14 members were in 'Bad' cluster; and no members are not included in the cluster.

4. Conclusion
The Savings and Loans Cooperative application that has been designed can provide recommendations through clustering to the management about the feasibility of members for getting credit using the Score ranking process. The result of testing shows that 31 members were in 'OK' cluster; 33 members were in 'Warning' cluster; 14 members were in 'Bad' cluster; and no members are not included in the cluster.

5. References
[1] Kathiresan V and Sumathi P 2012 An efficient clustering algorithm based on Z-Score ranking method Proc. 2012 Int. Conf. Comput. Commun. Informatics 1–4
[2] Saputro Y E 2017 Sistem Pendukung Keputusan Potensi Nasabah Menggunakan Metode K-Means pada Koperasi Simpan Pinjam Tekad Usaha Mandiri Boyolali (Surakarta: STMIK Sinar Nusantara)
[3] Suryadi S and Irmayani D 2018 Penerapan Metode Clustering K-means untuk Pengelompokan Kelulusan mahasiswa Berbasis Kompetensi INFORMATIKA 6 1 52–72
[4] Vhallah I, Sumijan S and Santony J 2018 Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering K-Means J. RESTI (Rekayasa Sist. dan Teknol. Informasi) 2 2 572–577
[5] Prasetyowati E and Rofiq A A 2016 Penilaian Kinerja Keuangan Koperasi pada Dinas Koperasi dan UMKM Pamekasan dengan K-Means J. Simantec 5 2 67–74
[6] Dwiarta I M B and Sugijanto S 2016 Analisis Segmentasi Pasar Berdasarkan Karakteristik Perilik Nasabah Pengambil Kredit pada Koperasi Simpan Pinjam (KSP) Citra Abadi di Babat Lamongan Maj. Ekon. 21 2 295–303
[7] Maylawati D S, Ramdhani M A and Amin A S 2018 Tracing the Linkage of Several Unified Modelling Language Diagrams in Software Modelling Based on Best Practices Int. J. Eng. Technol. 7 2.29 776–780
[8] Suhendi H Y, Ramdhani M A and Irwansyah F S 2018 Verification Concept of Assessment for Physics Education Student Learning Outcome Int. J. Eng. Technol. 7 3.21 321–325
[9] Ramdhani M A, Aulawi H, Ilkhwana A and Mauluddin Y 2017 Model of green technology adaptation in small and medium-sized tannery industry J. Eng. Appl. Sci. 12 4 954–962
[10] Pamoragung A, Suryadi K and Ramdhani M A 2006 Enhancing the implementation of e-Government in Indonesia through the high-quality of virtual community and knowledge portal Proceedings of the European Conference on e-Government, ECEG 341–348
[11] Ramdhani M A 2013 Metodologi Penelitian untuk Riset Teknologi Informasi.(Bandung: UIN Sunan Gunung Djati)
[12] Maylawati D S, Darmalaksana W and Ramdhani M A 2018 Systematic Design of Expert System Using Unified Modelling Language IOP Conf. Ser. Mater. Sci. Eng. 288 1 012047
[13] Ramdhani M A, Maylawati D S, Amin A S and Aulawi H 2018 Requirements Elicitation in Software Engineering Int. J. Eng. Technol. 7 2.29 772–775
[14] Aulawi H, Ramdhani M A, Slamet C, Ainisysifa H and Darmalaksana W 2017 Functional Need Analysis of Knowledge Portal Design in Higher Education Institution Int. Soft Comput. 12 2 132–141
[15] Darmalaksana W, Ramdhani M A, Cahyana R and Amin A S 2018 Strategic Design of Information System Implementation at University Int. J. Eng. Technol. 7 2.29 787–791
[16] Suryadi K and Ramdhani M A 2002 Sistem Pendukung Keputusan. (Bandung: Remaja Rosdakarya)
[17] Slamet C, Andrian R, Maylawati D S, Darmalaksana W and Ramdhani M A 2018 Web Scraping
and Naïve Bayes Classification for Job Search Engine *IOP Conf. Ser.: Mater. Sci. Eng* **288** 1–7

[18] Gerhana Y A, Zulfikar W B, Ramdani A H and Ramdhani M A 2018 Implementation of Nearest Neighbor using HSV to Identify Skin Disease *IOP Conf. Ser. Mater. Sci. Eng.* **288** 1 012153

[19] Rahaman C, Slamet C, Darmalaksana W, Gerhana Y A and Ramdhani M A 2018 Expert System for Deciding a Solution of Mechanical Failure in a Car using Case-based Reasoning *IOP Conf. Ser. Mater. Sci. Eng.* **288** 1 012011

[20] Slamet C, Rahman A, Sutedi A, Darmalaksana W, Ramdhani M A and Maylawati D S 2018 Social Media-Based Identifier for Natural Disaster *IOP Conf. Ser. Mater. Sci. Eng.* **288** 1 012039

[21] Slamet C, Rahman A, Ramdhani M A and Darmalaksana W 2016 Clustering the Verses of the Holy Qur’an Using K-Means Algorithm *Asian J. Inf. Technol.* **15** 24 5159–5162

[22] Maylawati D S, Ramdhani M A, Zulfikar W B, Taufik I and Darmalaksana W 2017 Expert system for predicting the early pregnancy with disorders using artificial neural network *2017 5th Int. Conf. Cyber IT Serv. Manag.* CITSM 2017

[23] Maylawati D S, Ramdhani M A, Rahman A and Darmalaksana W 2017 Incremental technique with set of frequent word item sets for mining large Indonesian text data *2017 5th Int. Conf. Cyber IT Serv. Manag.* CITSM 2017 1–6

[24] Zulfikar W B, Jumadi, Prasetyo P K and Ramdhani M A 2018 Implementation of Mamdani Fuzzy Method in Employee Promotion System *IOP Conf. Ser. Mater. Sci. Eng.* **288** 1 012147

[25] Taofik A, Ismail N, Gerhana Y A, Komarujaman K and Ramdhani M A 2018 Design of Smart System to Detect Ripeness of Tomato and Chili with New Approach in Data Acquisition *IOP Conference Series: Materials Science and Engineering* **288** 1 012018

[26] Manaf K 2016 Implementasi Sistem Pakar Diagnosa Kerusakan pada Canon Np6650Xx *JOIN (Jurnal Online Inform.)* **1** 2 92–97

[27] Sulastri H and Gufroni A I 2017 Penerapan Data Mining dalam Pengelompokan Penderita Thalassaemia *Teknol. dan Sist. Inf.* **3** 2 299–305

[28] Zulfikar W B and Lukman N 2016 Perbandingan Naive Bayes Classifier dengan Nearest Neighbor untuk Identifikasi Penyakit Mata *JOIN (Jurnal Online Inform.)* **1** 2 82–86

[29] Wu X et al. 2008 Top 10 algorithms in data mining *Knowl. Inf. Syst.* **14** 1 1–37

[30] Deniyanto R D 2014 *Analisis Kinerja Keuangan pada Koperasi Simpan Pinjam (KSP) Bina Usaha Utama Tahun 2012-2014* (Semarang: Universitas Dian Nuswantoro)

[31] Larasati A, Yasa I W and Widiyanti I D 2014 Aspek Hukum Pemberian Pinjaman oleh Koperasi Simpan Pinjam (KSP) Unit Simpan Pinjam (USP) Koperasi pada Calon Anggota Berdasarkan Peraturan Pemerintah Nomor 9 Tahun 1995 *Artikel Ilmiah Hasil Penelitian Mahasiswa 2013*

[32] Ruri, “SOP Koperasi Simpan Pinjam,” 2009.

[33] Arinta D Y 2014 Pengaruh Karakteristik Individu, Karakteristik Usaha, Krakeristik Kredit Terhadap Kemampuan Debitur Membayar Kredit pada BPR Jatim Cabang Probolinggo *J. Ilm. Mhs. FEB* **2** 1 1–16

**Acknowledgments**

The authors expressed the appreciation and thank you to the Research and Publishing Center of UIN Sunan Gunung Djati Bandung, which has provided funding support for the publication of this article.