JUSIKOM PRIMA (Jurnal Sistem Informasi dan Ilmu Komputer Prima)
Vol. 8 No. 1, August 2024 E-ISSN:2580-2879

Implementation of comparison of K-means algorithm with C4.5 algorithm to predict the feasibility of being a Catholic

Ricky Kristian Arifin Hulu[1], Alwis Ginting[2], Yonata Laia[3]*
[1,2,3]Universitas Prima Indonesia, Medan, Indonesia
rickykrist121100 @email.com

ABSTRACT

This research conducted a comparison between two methods, namely C4.5 and K-Means, with the aim of improving the work efficiency of the Catholic Church Secretariat in selecting data of prospective Catholic congregants. These methods were developed to assist secretariat employees in determining the eligibility of valid files and data of prospective Catholic congregants through a web-based application. The data used comes from the selection results of several files and data of prospective congregants that have been collected by the Catholic Church Secretariat. Data analysis was carried out using the K-Means and C4.5 algorithms to predict the feasibility of the prospective congregants' files. It is hoped that the results of this research can help the Catholic Church Secretariat to improve work efficiency, both in terms of time and effort, in the selection process of prospective Catholic congregants and increase accuracy in determining the eligibility of the data of each applicant at the Catholic Church Secretariat improve the efficiency of the selection process and enhance the accuracy in determining the suitability of potential members.

Keywords: Gender, Survey Radius, Domicile Status, Occupation, Amount of Income, Number of Loans, Tenor, Number of Dependents, Marital Status.

INTRODUCTION

As we know, many human activities can be transformed into applications and do not need to be done manually, such as employee data entry in companies, congregation data entry in churches, and many other tasks that can be computerized. The church is a group of Christ's followers who have transitioned from a sinful life to a righteous one. Throughout its history, the church has faced many conflicts leading to divisions. One of the most famous splits occurred during the Reformation, which gave birth to the Protestant group, eventually separating from the parent group known as Roman Catholicism. The Roman Catholic Church (RC) has many theological or doctrinal differences from Protestant groups, one of which is the concept of the Kingdom of God. The Catholic Church has a different process for accepting congregants compared to other churches. In Catholicism, there are several documents that must be fulfilled. All candidates register through their local community, which then submits the applications to the secretariat for data verification. The data will be thoroughly examined, including the origin of their parish,

reasons for moving, and several other requirements such as the main family card and baptism certificate. The problem in determining candidates to be accepted as congregants lies in the manual data collection process.

This study attempts to address this issue by comparing the C4.5 and K-Means methods. The goal is to provide a solution and assist the secretariat in determining whether candidates are eligible to become Catholic congregants. Data mining is a process or tool for extracting valuable information from large datasets by identifying patterns or interesting information through knowledge discovery analysis in databases using certain techniques [1]. Techniques in data mining include clustering and classification [2]. The clustering and classification algorithms currently widely used include K-Means and C4.5 [3]. The K-Means algorithm is used for iterative data grouping by dividing the dataset into a predefined number of clusters (k) [4]. Meanwhile, the C4.5 algorithm is used to form decision trees that can predict a decision by applying a series of decision rules [5].

The above description motivates the author to contribute to solving the problem of decision-making in the selection of Catholic congregants, especially in the Tanjung Anom church and its surroundings. This system will implement a comparison of the C4.5 method with K-Means in hopes of providing a solution to the secretariat in receiving data from prospective congregants.

LITERATURE REVIEW

In this section, the author will outline the theoretical basics related to the research topic, the modeling used, and the software used.

Theories related to the research topic an Analysis

Wanto and Windarto (2017:38) write that analysis is the ability to solve a subject's problems into constituent elements, look for internal relationships and between elements, and arrange the overall problem-solving formats that exist, so that in the end they produce an expected value. Other researchers such as Siregar *et al* (2020:95) also added that the first step for researchers in conducting analysis, especially in analyzing data from the data that has been collected, is to review the research proposal to examine the data presentation plan and the implementation of the analysis of statistics that have been originally determined.

Based on the definitions from experts that have been explained earlier, it can be concluded that

analysis is a thought process related to the systematic testing of an object to determine the parts involved in a series of activities such as decomposing, distinguishing, and selecting something to be classified and regrouped based on certain criteria.

Planning

Yoslando Gustin *et al* (2019:319) said that planning is the drawing, planning, and sketching or arrangement of several separate elements into a whole, and functioning as a system design can be realized in the form of diagrams, which are graphic tools used to show the sequences of processes in the system. Siregar *et al* (2018:179) also stated that design is the process of developing specifications for a new system based on the results of recommendations from the system analysis. In the opinion of experts, design is the process of drawing, planning, and sketching or arranging several elements that are put together into a whole and functional whole.

Thus, it can be said that planning is a pattern that is made to overcome problems faced in a subject or object, which provides an overview or sketch of the arrangement in a complete element. Lisa Ariyanti *et al* (2020:91) An information system is a combination of several people working together to achieve a specific goal. On the other hand, information systems also play a role in supporting decision-making. Erlin Elisa *et al* (2020: 50) also said that an information system is a system that meets the daily transaction processing needs and supports operations in an organization.

Thus, it can be concluded that an information system is a system that facilitates user performance and accelerates human work processes that previously took a long time.

Data umat

Ketut Adi Hardana (2016:153) Defines that the data of the people is the factual and actual reality that exists in the community, which provides an accurate picture of the situation and potential of the people in each place of worship.

Desktop-Based

In the Great Dictionary of Indonesian (KBBI), a desktop is defined as the entire display of the monitor screen consisting of icons that are arranged to facilitate the identification of application

programs. Meanwhile, in the Windows operating system, the desktop is the main display that functions similarly to a desk. On the desktop, *users* can place various shortcuts to open folders and various types of files, and can run the selected software (Aan Zaksa, 2020:1). Through this definition, it can be concluded that a desktop is the entire display of the monitor screen that contains application icons, which can be directly used to work by users.

Widodo, Nurhayati & Windasari (2016) in their study showed examples of the use of desktop-based applications that are very possible to be used as a census recording tool. The application built is a desktop-based application that uses the Delphi programming language. In their research, it was written that one of the drawbacks of using a desktop is security. Security of use only applies up to a certain login level, not the entire system.

Modelling Theory Used

1. UML (Unified Modeling Language)

According Munawar (2018:49) explained that "UML (*Unified Modeling Language*) is one of the most reliable tools in the world of object-oriented system development".

UML provides a huge number of diagrams that can be expanded to suit our needs. Diagrams are graphic representations of certain elements and their relationships. Diagrams are very important because they provide a visualization of a system or its parts. Graphical representations greatly facilitate the understanding of the system. UML provides a standard for designing a model of a system.

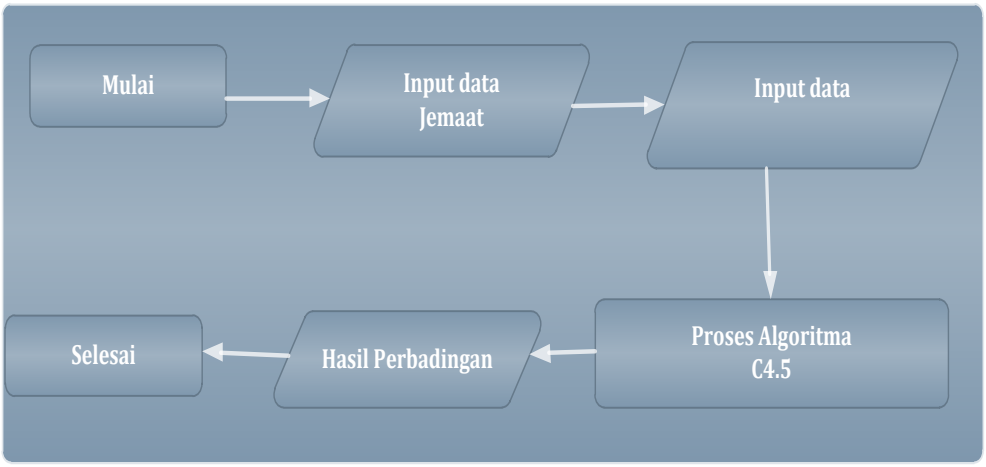
Using UML can be modeled for all types of software applications, where the application can run on any hardware, operating system, and network, and be written in any programming language. UML also uses *classes* and *operations* in its basic concepts, so it is better suited for writing software in object-oriented languages such as C++, Java, or Visual Basic.Net

METHOD

The dataset used is data on Catholic congregations at Simpang Melati Catholic Church from 2020 to 2024 as many as 100 congregations. This dataset is divided into two, namely congregational data from 2020-2023 as training data, and congregational data in 2024 as test data. The dataset will be divided based on the category of activity and contribution in the church,

namely active category congregation data and inactive category congregation data. Table 2 shows the number of Catholic congregations in Simpang Melati Catholic Church from 2020 to 2024.

The data of the active category congregation has variables such as Congregation ID, baptismal certificate, letter of change of origin, ID card, and parent family card. As for the data of congregations in the inactive category, there are variables such as congregation ID, baptismal certificate, letter of change of origin, ID card, parent family card, attendance at church activities, and financial contributions.



(a)

Table 3.1 Catholic Congregation Data

Figure 3.1 In this research, there is a form to fill in the values that will be processed by the c4.5 method and which will be classified by the k-mean method which can be seen in the form above.

The results of the C4.5 Method process. if Gender = Male and Survey Radius = 1-5 KM and Ones Domicile = Contract and Occupation = Self-Employed and Large Income = <=2.5 Million and Loan Amount = <=2.5 Million and Tenor = 12 Months and Number of Dependents = <=2 People and Marital Status = Marriage then `Class = Current.`

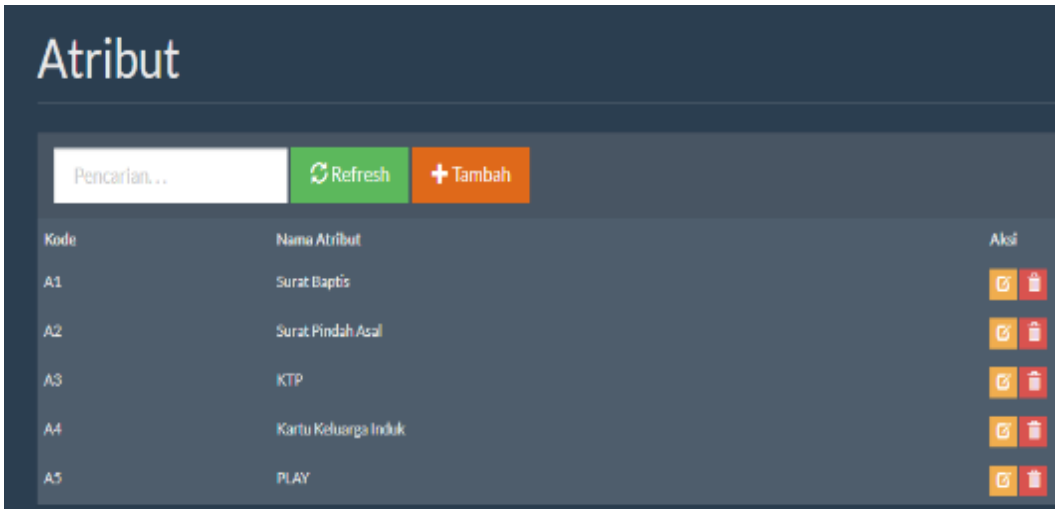
1. Test Design

Currently, Simpang Melati Catholic already has an assessment system to determine customers who are worthy of being accepted but it is still done manually not based on the system. This prediction system is indispensable to obtaining information about the standards for submitting prospective believers in each environment. Based on the information from this prediction, improvements can be made to the new community data selection system. To solve this problem, a prediction system can be designed by applying the C4.5 method comparison. Broadly speaking, the working procedure of the system to be created can be described as seen in Figure 3.1

RESULT

Attribute

Attributes are data that will be used as material in the selection of each customer. In this study, to make it easier for researchers to create a system, it is first necessary to determine the attributes that will be used in data selection. Attributes to be used include:



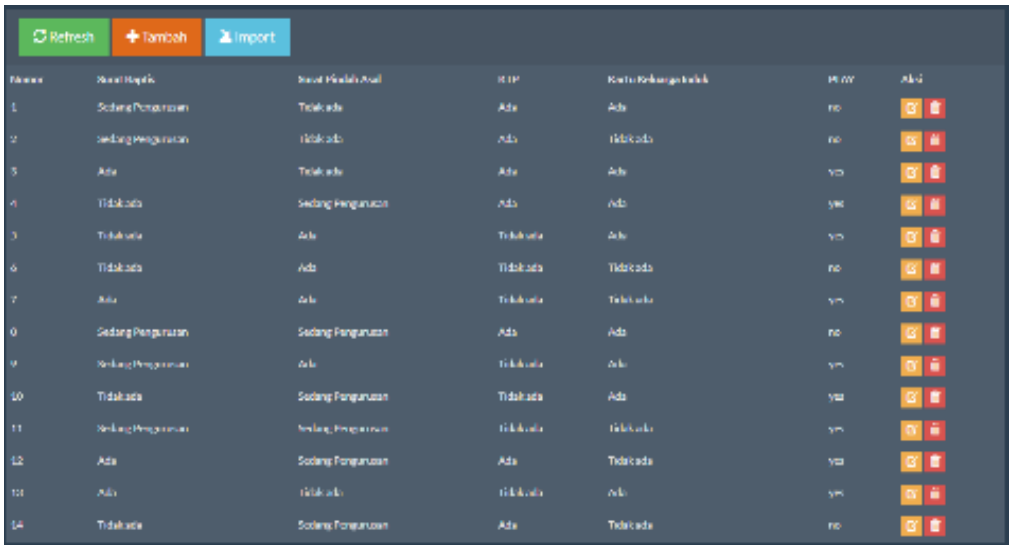


Figure 3.2. Attributes and datasets

In this study, a process is made with attribute and dataset criteria where the list of data used in the program is shown in figure a while figure b displays the dataset displayed on the system.

C4.5 Calculation

This study has succeeded in making calculations based on the method used, namely the c4.5 and k-means methods which can be seen in the results of the program below which are processed by the system with the method used:

```
===Root Branch Calculation ===
Baptismal Letters:
       Medium Management(5/14): 0.971
       Ada(4/14): 0
       None(5/14): 0.971
       GAIN: 0.247
```

```
        SPLIT INFO: 1.577

        GAIN RATIO: 0.156

Letter of Transfer of Origin:

        None(4/14): 1

        Medium Management(6/14): 0.918

        Ada(4/14): 0.811

        GAIN: 0.029

        SPLIT INFO: 1.557

        GAIN RATIO: 0.019

KTP:

        Ada(7/14): 0.985

        None(7/14): 0.592

        GAIN: 0.152

        SPLIT INFO: 1

        GAIN RATIO: 0.152

Parent Family Card:

        Ada(8/14): 0.811

        None(6/14): 1

        GAIN: 0.048

        SPLIT INFO: 0.985

        GAIN RATIO: 0.049

Best attribute: Baptism Letter (0.156)
```

```
===Calculation of Baptism Branch (Under Management)===

Letter of Transfer of Origin:

       None(2/5): 0

       Medium Management(2/5): 1

       Ada(1/5): 0

       GAIN: 0.571

       SPLIT INFO: 1.522

       GAIN RATIO: 0.375

KTP:

       Ada(3/5): 0

       None(2/5): 0

       GAIN: 0.971

       SPLIT INFO: 0.971

       GAIN RATIO: 1

Parent Family Card:

       Ada(3/5): 0.918

       None(2/5): 1

       GAIN: 0.02

       SPLIT INFO: 0.971

       GAIN RATIO: 0.021

Best attribute: ID card (1)
```

```
===KTP Branch Results (Exist):no===


===KTP Branch Results(None):yes===


===Results of  the Baptist Letter(Exist):yes===


===Calculation of Baptismal Branch (None)===
Letter of Transfer of Origin:
       Medium Management(3/5): 0.918
       Ada(2/5): 1
       GAIN: 0.02
       SPLIT INFO: 0.971
       GAIN RATIO: 0.021
KTP:
       Ada(2/5): 1
       None(3/5): 0.918
       GAIN: 0.02
       SPLIT INFO: 0.971
       GAIN RATIO: 0.021
Parent Family Card:
       Ada(3/5): 0
```

```
        None(2/5): 0

        GAIN: 0.971

        SPLIT INFO: 0.971

        GAIN RATIO: 1

Best attribute: Parent Family Card (1)


===Results of the Parent Family Card Branch(Exist):yes===


===Hasil Cabang Kartu Keluarga Induk(Tidak ada):no===
```

Example of K-Means Method Calculation

The following is an example of the calculation of the k-means method with the necessary steps:

1. Specifies the desired number of clusters (k). For example, let's say we want to create 3 clusters.
2. Randomly select the starting point as the center of the cluster. For example, let's say we have the following dataset with x and y coordinates:

   Point A(2, 10) Point B(2, 5) Point C(8, 4) Point D(5, 8) Point E(7, 5) Point F(6, 4)

   We will select three starting points as the cluster centers, for example: Cluster Center 1: (2, 10) Cluster Center 2: (5, 8) Cluster Center 3: (6, 4)

3. Calculate the distance between each data point and the existing cluster center. In this case, we will use the Euclidean distance. Distance between point A and Cluster Center 1 = sqrt((2-2)^2 + (10-10)^2) = 0 Distance between point A and Cluster Center 2 = sqrt((2-5)^2 + (10-8)^2) = 3.6055 Distance between point A and Cluster Center 3 = sqrt((2-6)^2 + (10-4)^2) = 6.3246 . Distance between point F and Cluster Center 1 = sqrt((6-2)^2 + (4-

10)^2) = 6.3246 Distance between point F and Cluster Center 2 = sqrt((6-5)^2 + (4-8)^2) = 4.1231 Distance between point F and Cluster Center 3 = sqrt((6-6)^2 + (4-4)^2) = 0

4. Each data point will be grouped into the nearest cluster based on the previously calculated distance. For example, the result is as follows:

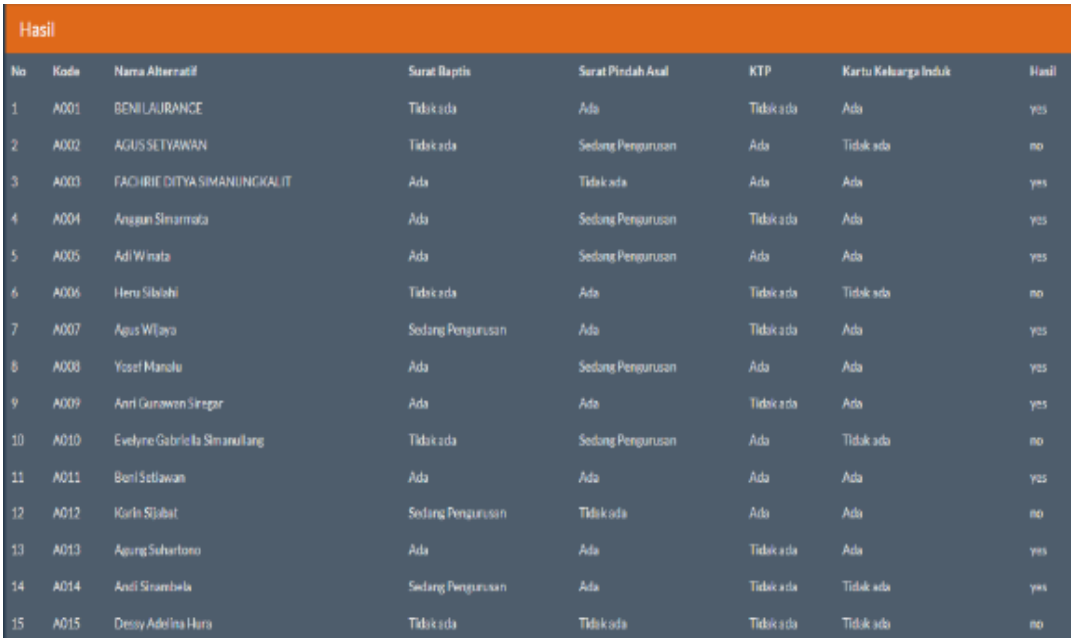
   Cluster 1: Point A, Point B Cluster 2: Point D, Point E Cluster 3: Point C, Point F

5. Recalculates the new cluster center based on the average coordinates of each data point in the same cluster. For example:

   Pusat Cluster 1: ((2+2)/2, (10+5)/2) = (2, 7.5) Pusat Cluster 2: ((5+7)/2, (8+5)/2) = (6, 6.5) Pusat Cluster 3: ((8+6)/2, (4+4)/2) = (7, 4)

6. Repeat steps 3 through 5 until there is no change in the cluster grouping or has reached the specified maximum iteration.

The results of the system obtained by the k-mean method based on the results of the classification are based on the results of the classification obtained can be seen in the figure below.



Hasil

No	Kode	Nama Alternatif	Surat Baptis	Surat Pindah Asal	KTP	Kartu Keluarga Induk	Hasil
1	A001	BENI LAURANCE	Tidak ada	Ada	Tidak ada	Ada	yes
2	A002	AGUS SETYAWAN	Tidak ada	Sedang Pengurusan	Ada	Tidak ada	no
3	A003	FACHRIE DITYA SIMANUNGKALIT	Ada	Tidak ada	Ada	Ada	yes
4	A004	Anggun Simarmata	Ada	Sedang Pengurusan	Tidak ada	Ada	yes
5	A005	Adi Wirata	Ada	Sedang Pengurusan	Ada	Ada	yes
6	A006	Heru Sibalahi	Tidak ada	Ada	Tidak ada	Tidak ada	no
7	A007	Agus Wijaya	Sedang Pengurusan	Ada	Tidak ada	Ada	yes
8	A008	Yosef Manalu	Ada	Sedang Pengurusan	Ada	Ada	yes
9	A009	Anri Gunawan Siregar	Ada	Ada	Tidak ada	Ada	yes
10	A010	Evelyne Gabriella Simanullang	Tidak ada	Sedang Pengurusan	Ada	Tidak ada	no
11	A011	Beni Setiawan	Ada	Ada	Ada	Ada	yes
12	A012	Karin Sijabat	Sedang Pengurusan	Tidak ada	Ada	Ada	no
13	A013	Agung Suhartono	Ada	Ada	Tidak ada	Ada	yes
14	A014	Andi Sirambela	Sedang Pengurusan	Ada	Tidak ada	Tidak ada	yes
15	A015	Dessy Adelina Hura	Tidak ada	Tidak ada	Tidak ada	Tidak ada	no

Table 3.1 Centroid center

Name	Gender	Radius Survey	Status Domisili	Work	Large Income	Loan Amount	Tenor	Number of Dependents	Marital Status
Problematic	1.62	2.76	1.9	1.66	2.92	3.74	2.38	2.14	1.46
Lancar	1.42	2.54	2.14	1.38	3.58	2.8	2.7	1.9	1.58

The table above displays the centroid center data which will be used as a benchmark in the calculation of the next stage.

Table 3.2 Distance to Centroid Center

Name	Gender	Radius Survey	Status Domisili	Work	Large Income	Loan Amount	Number of Dependents	Marital Status	Problematic	Lancar	Group
1	2	1	1	2	2	3	3	1	2.5802	2.9941	Problematic
2	1	2	1	1	5	3	1	2	3.2523	2.8009	fluent
3	1	4	1	2	1	1	1	1	4.1782	4.2338	Problematic
4	2	1	2	1	4	5	2	2	2.6263	2.9266	Problematic
5	2	4	3	1	3	3	2	1	2.5956	2.401	Fluent

Name	Gender	Radius Survey	Status Domisili	Work	Large Income	Loan Amount	Number of Dependents	Marital Status	Problematic	Lancar	Group
6	2	4	2	2	5	2	1	2	3.5577	3.0667	Fluent
7	2	4	3	1	5	5	2	1	3.1012	3.3234	Problematic
8	1	2	3	2	4	5	1	1	2.5879	2.8714	Problematic
9	1	1	1	1	3	4	1	2	2.6033	2.6163	Problematic
10	2	4	1	2	3	3	2	1	2.3016	2.7865	Problematic
11	2	1	1	1	3	1	2	2	3.7706	3.2871	Fluent
12	1	4	1	1	4	1	3	2	3.9366	3.2007	Fluent
13	1	1	2	2	4	5	3	1	3.1523	3.3474	Problematic
14	2	2	3	1	3	5	1	2	2.4366	2.7865	Problematic
15	2	2	1	2	5	3	2	2	2.6414	2.2461	fluent
16	2	4	2	1	5	1	1	1	4.1686	3.4532	Fluent
17	2	3	2	1	1	4	2	1	2.6939	3.2931	Problematic
18	1	2	2	2	5	5	2	2	2.7308	2.8991	Problematic
19	1	4	2	2	5	4	3	1	2.7888	2.7937	Problematic
20	1	4	2	1	5	3	1	2	3.3848	2.6842	Fluent
21	2	1	2	2	4	5	2	2	3.0095	3.163	Problematic
22	2	3	1	2	1	3	1	1	2.6939	3.1882	Problematic
23	2	3	3	2	2	3	1	2	2.1581	2.3843	Problematic

Because the new group (31, 32, 31, 31, 32, 32, 31, 31, 31, 31, 32, 32, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 32, 32, 32, 31, 31, 32, 31, 32, 31, 32, 32, 31, 31, 32, 32, 31, 31, 31, 31, 32, 32, 31, 32, 31, 32, 31, 32, 32, 32, 31, 32, 31, 31, 32, 31, 32, 31, 31, 32, 32, 32, 31, 31, 32, 32, 32, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 31, 32, 32, 32, 32, 32, 32, 31, 31, 31, 32) <> the previous group (), then the iteration is continued.

DISCUSSIONS

In this study, we implemented and compared two algorithms, namely C4.5 and K-Means, to evaluate the eligibility of potential parishioners at the Catholic Church of Simpang Melati. This discussion section delves deeper into the results obtained, data interpretation, implications of the findings, and recommendations for future research.

Results of C4.5

The C4.5 algorithm was used to construct a decision tree based on the available attributes. This decision tree assists in determining the eligibility of potential parishioners based on various criteria such as baptism certificates, transfer letters, identity cards, and family cards of the parents. Based on the calculations using C4.5, the attribute "Baptism Certificate" was identified as the best attribute with the highest gain ratio of 1 The decision tree revealed that if a potential parishioner had a baptism certificate, the probability of being eligible was significantly higher. Other significant attributes included the transfer letter and the family card. The tree structure provides a clear and interpretable model for assessing eligibility, which is crucial for the church's administrative processes.

Results of K-Means Clustering

The K-Means clustering algorithm was employed to identify patterns and group the data into clusters based on similarity. The clustering analysis resulted in the formation of several distinct clusters that represent different profiles of potential parishioners. The problematic data cluster centers were identified at: 1.5283, 2.566, 1.9811, 1.5283, 2.8491, 4.3585, 2.4717, 1.8491, and 1.3962. In contrast, the clusters representing smooth data were centered at: 1.5106, 2.7447, 2.0638, 1.5106, 3.7021, 2.0426, 2.617, 2.2128, and 1.6596. These clusters help in understanding the underlying patterns in the data and provide insights into the different categories of parishioners. For instance, clusters with higher values in attributes such as income and occupation indicate a distinct group with specific needs and expectations.

Comparison Between C4.5 and K-Means

The comparison between the C4.5 decision tree and the K-Means clustering results highlights the

strengths and limitations of each method. The decision tree offers a clear and interpretable model for eligibility determination, making it practical for administrative decision-making. On the other hand, the clustering analysis provides a broader view of the data, uncovering hidden patterns and groupings that might not be apparent through a decision tree alone.

For example, while the decision tree can pinpoint the exact criteria for eligibility, the clustering method reveals the overall distribution and segmentation of the parishioner data. This dual approach ensures a comprehensive understanding of the data, allowing for both precise eligibility assessment and a strategic overview of the parishioner demographics.

Implications of the Findings

The findings of this study have significant implications for the Catholic Church of Simpang Melati. By leveraging the C4.5 algorithm, the church can streamline its eligibility assessment process, ensuring that all necessary criteria are met efficiently. The decision tree model can be implemented as a tool for quick decision-making, reducing the administrative burden on church officials. The insights gained from the K-Means clustering analysis can be used to tailor church programs and services to meet the needs of different parishioner groups. For instance, understanding the distinct clusters can help the church design targeted outreach programs, improve community engagement, and allocate resources more effectively.

Recommendations for Future Research

Future research can build upon this study by exploring the integration of additional data attributes, such as attendance records and participation in church activities, to further refine the eligibility criteria. Additionally, incorporating advanced machine learning techniques, such as random forests or neural networks, could enhance the accuracy and robustness of the models. Another avenue for future research is the application of these algorithms in other administrative areas within the church, such as event planning and volunteer management. Expanding the scope of data analysis can provide deeper insights and support more informed decision-making across various aspects of church operations.

In conclusion, the combination of the C4.5 and K-Means algorithms provides a powerful approach for evaluating and understanding parishioner data. The decision tree model offers a

practical tool for eligibility assessment, while the clustering analysis uncovers valuable patterns that inform strategic planning. By continuing to refine and expand these methods, the Catholic Church of Simpang Melati can enhance its administrative processes and better serve its community.

CONCLUSION

In this study, two methods, C4.5 and K-Means, were used and compared with the aim of improving the work efficiency of the Catholic Church Secretariat in selecting data of prospective Catholic congregants. The results show that if data collection is conducted using the C4.5 method, conditions such as Name, Baptism Certificate, ID Card, Family Card, Domicile Status, and Marital Status will result in Class = Current. Additionally, the clustering results using the K-Means method identified several problematic cluster centers with values 1.5283, 2.566, 1.9811, 1.5283, 2.8491, 4.3585, 2.4717, 1.8491, 1.3962, while the data classified as smooth showed values of 1.5106, 2.7447, 2.0638, 1.5106, 3.7021, 2.0426, 2.617, 2.2128, 1.6596

REFERENCES

S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analysis of K-Medoids Clustering Algorithm in Grouping the Spread of Covid-19 in Indonesia," Jti (Jurnal Teknol. Information), vol. 4, no. 1, pp. 166 –173, 2020, n.d.;

A. P. Windarto, U. Indriani, M. R. Raharjo, and L. S. Dewi, "Part 1: Combination of Clustering and Classification Methods (Covid-19 Pandemic Cases in Indonesia)," J. Media Inform. Budidarma, vol. 4, no. 3, p. 855, 2020, doi: 10.30865/mib.v4i3.2312, n.d.;

I. G. I. Sudipa, I Nyoman Alit Arsana, and Made Leo Radhitya, "Determination of Students' Understanding of Social Distancing Using the C4.5 Algorithm," SINTECH (Science Inf. Technol. J., vol. 3, no. 1, pp. 1 –7, 2020, doi: 10.31598/sintechjournal.v3i1.562, n.d.;

Astuti FD, Guntara M. Performance Analysis of K-NN and C4.5 Algorithms on Poor Data Classification. J Technol Engineering Inf. 2018; 2(2):135. doi:10.30872/jurti.v2i2.1865, n.d.;

S. Turnip and P. Siltionga, "Analysis of Disease Spread Patterns Using C4.5 Algorithm," J. Tek. Inform. Unika St. Thomas, vol. 03, no. 479, pp. 3 –7, 2018, n.d.;

A. T. R. Saragih, A. S. Sembiring, and M. Sayuthi, "Application of K-Means Clustering Method

for the Selection Process of Prospective Participants of the MTQ Competition," Pelita Inform., vol. 17, no. April, pp. 117–122, 2018, [Online]. Available: https://ejurnal.stmik-budidarma.ac.id/index.php/pelita/article/download/776/704, n.d.;

N. Dwitri, J. A. Tampubolon, S. Prayoga, F. Ilmi Zer, and D. Hartama, "Application of K-Means Algorithm in Determining the Level of Online Learning Satisfaction During the Covid-19 Pandemic in Indonesia," Jti (Jurnal Teknol. Information), vol. 4, no. 1, pp. 101-105, 2020, n.d.;

A. Susilo et al., "Coronavirus Disease 2019: A Review of the Latest Literature," J. Internal Medicine Indones., vol. 7, no. 1, p. 45, 2020, doi: 10.7454/jpdi.v7i1.415, n.d.;

R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," vol. 10, no. 3, pp. 275-282, 2020 , n.d.;

D. Alverina, A. R. Chrismanto, and R. G. Santosa, "Perbandingan Algoritma C4.5 Dan Cart Dalam Memprediksi Kategori Indeks Prestasi Mahasiswa," Jurnal Teknologi dan Sistem Komputer, vol. 6, no. 2, April 2018. doi: 10.14710/jtsiskom.6.2.2018.76-83, n.d.;

Silitong Parasian. Irene Sri Morina. 2018. "Implementation of K-Means Clustering on Patient Data Of National Social and Healthcare Security by Using Java"., International Journal of Computer Science Engineering (IJCSE), ISSN: 2319-7323 Vol. 7 No.1 Jan-Feb 2018, n.d.;

UNIKOM_PWK_FEBRI GUINENSA PUTRA_10614010_BAB II TINJAUAN PUSTAKA, n.d.;

ANALISIS SENTIMEN MASYARAKAT TERHADAP, n.d.; BAB II Tinjauan Teori , n.d.; Eka-Citta 27, n.d.;

IMPLEMENTASI ALGORITMA C4.5 DAN K-MEANS PADA DIAGNOSIS PENYAKIT GINJAL KRONIS, n.d.; Makalah DInding Indra Sakti Ramadhan, n.d.;

Modul Kekayaan Intelektual Tingkat Dasar Bidang Hak Cipta (Edisi 2020), n.d.; Alamat et al., 2022; Bagus Gede Sarasvananda et al., 2021; Gozales et al., 2024; Immanuel et al., 2019; Nareswari Manuaba & Sukihana, 2020a, 2020b, 2020c;

Penerapan Kecerdasan Buatan et al., n.d.; Prasatya et al., 2020a, 2020b; Qomaruddin et al., 2021; Ramadhan et al., 2024; Rosita et al., 2018; Supriyadi et al., n.d.; Tria et al., 2018; Widodo, n.d.)