BirEx JOURNAL
Budapest International Research in Exact Sciences
Medical, Biological, Argiculture, Engineering
Science and other related areas
e-ISSN : 2655-7827
p-ISSN : 2655-7835
http://www.bircu-journal.com/index.php/birex

Implementation of Data Mining Technique for Performance of WFH and WFO Agents Using the K-Means Method Case Study Study of PT. Infomedia Telkom Consumer Profiling Services

Jaja[1], Nandi Priatna[2], Tazkia Salsabila Ardan[3]
[1,2,3]*Faculty of Computer Science, Universitas Subang, Indonesia*
jaja@unsub.ac.id

Abstract: *Outbound Call Center PT. Infomedia, consumer profiling service PT. Telkom during the pandemic period divided its agents into 80% WFH agents (Work at Home) and 20% agents WFO (Work from Office). For the division of the working mechanism, it is necessary to measure its performance. In the discussion of this paper, we will discuss the measurement with the application of data mining using the K-Means method, so it is hoped that it will provide an overview, how the cluster of each WFH or WFO agent in terms of performance. The results of this discussion indicate that there is a significant difference between the performance of WFH and WFO Agents.*
Keywords: *data mining; outbound call; K-Means*

I. Introduction

A call center in terms of Techopedia's definition is a facility that handles incoming and/or outgoing calls on behalf of the organization. For example, the call center can handle customer service calls, complaints or other issues related to company products and services. Based on the type, Call Centers are divided into 2 (two) types, namely Inbound and Outbound Call Centers. One of the outbound call center services at PT. Infomedia is a consumer profiling service, where this service has a core business in updating the customer profile of PT. Telkom to provide convenience in providing information about loyalty programs, the latest products and Telkom promotions, and accelerating service improvements if there are problems.

Outbound Call Center Services at PT. Infomedia which is currently running is still 100% using agents as the spearhead of services, where during the Covid19 pandemic, the WFH and WFO mechanisms were implemented as a preventive step in reducing the spread of the Covid19 Virus. The proportion of total human resources is 80% WFH (Work from home) and 20% WFO (Work from Office).

The division of work mechanisms for WFH and WFO types needs to measure performance on the side of the agent, which to map the agent cluster based on its performance, so that it is expected to provide a clearer mapping, especially for stakeholders in providing rewarding, monitoring and more precise guidance.

The performance mapping of WFH & WFO agents will be carried out using data mining clustering of the K-Means method. The use of the K-means method is taken because besides being easy to adapt, it is also easy to implement in a relatively faster time. Therefore, in the discussion of this paper, the author will explore: "Implementation of Data Mining technique for performance of WFH and WFO agents using the K-Means method. Case study of PT. Infomedia Telkom Consumer Profiling services".

DOI: *https://doi.org/10.33258/birex.v3i2.1810*

II. Research Methods

2.1 Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) is a non-trivial process for identifying novel, valid, potentially useful, and ultimately understandable patterns in the data from Fayyad et al. (1996a). The term "pattern" refers to a subset of data expressed in several languages or models that are exploited to represent that subset. KDD aims to find patterns that (i) does not result in a straightforward calculation of a predetermined amount (i.e., non-trivial), (ii) can be applied to new data with a certain degree of certainty (i.e., are valid), (iii) have not known so far (i.e., novel), (iv) provides some benefit to the user or to further (i.e., potentially useful) tasks, and (v) leads to useful insights, immediately or after some post-processing (i.e., understandable) [1]. This KDD describes systematically looking for a new relationship in market basket analysis using several stages of data processing[2].The KDD process is an iterative and interactive sequence of main steps as in Figure 1 (F.Gulo, 2015):

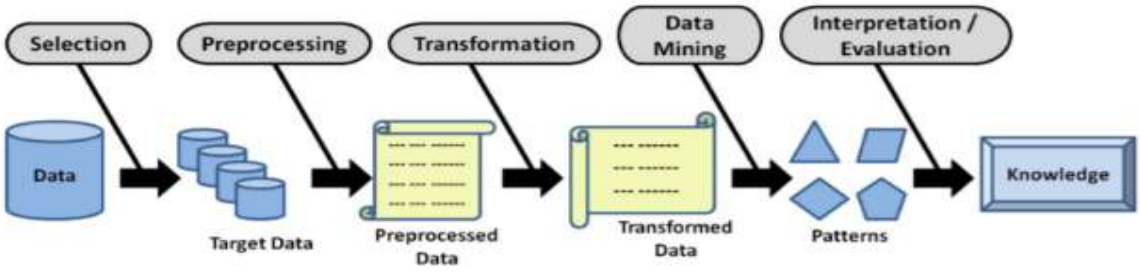

Figure 1. *Process Knowledge Discovery in Databases*

2.2 Data Mining

Data Mining is the process of employing one or more machine learning which is useful for analyzing and extracting knowledge automatically (J.Eska, 2018). Data mining is based on several techniques. Techniques are also based on different tools and algorithms (L. Mushunje, 2019). Data mining uses a discovery-based approach that is when pattern matching is carried out and other algorithms are used in determining the key relations in the analyzed data. Data mining or data mining has the meaning of searching for valuable business information from a very large database (Khormarudin, A.N. 2016).

By analogy, data mining should more accurately be called "knowledge mining from data", which is unfortunately rather old. In the short term, however, knowledge mining may not reflect an emphasis on mining large amounts of data (S.Agarwal, 2014).

Various kinds of defining data mining (L. Muflikhah, 2018) include:

- Decomposition (which is not simple) from a set of data into information that has potential implicitly (not real/clear) that was not previously known.
- Excavation and analysis, using automatic or semi-automatic devices, of large amounts of data to find meaningful patterns.
- Data Mining is a part of KDD.

As a series of processes, data mining can be divided into several process stages as illustrated in Figure 2. These stages are interactive, users are directly involved or through a knowledge base (M.S. Mustafa, M.R (2018).

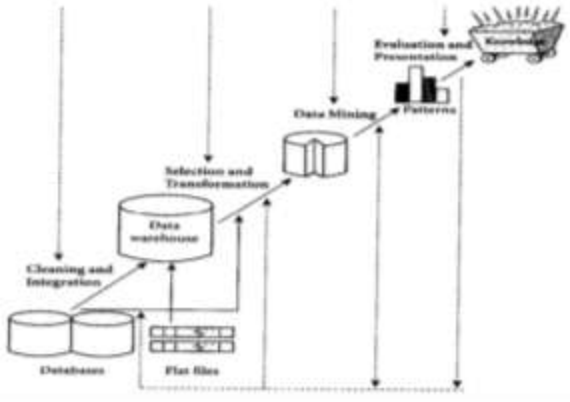

Figure 2. *Data Mining Stages*

The stages of data mining are as follows:

a. Data Cleaning
   The process of removing data noise, inconsistent data or irrelevant data.
b. Data Integration
   Data integration is the process of combining data from various databases into one new database.
c. Data Selection
   When data from databases are extracted only those suitable for analysis.
d. Data Transformation
   The activity of changing data is then combined into a format suitable for processing in data mining.
e. The Mining Process
   The process of finding new valuable knowledge from the data that has been obtained.

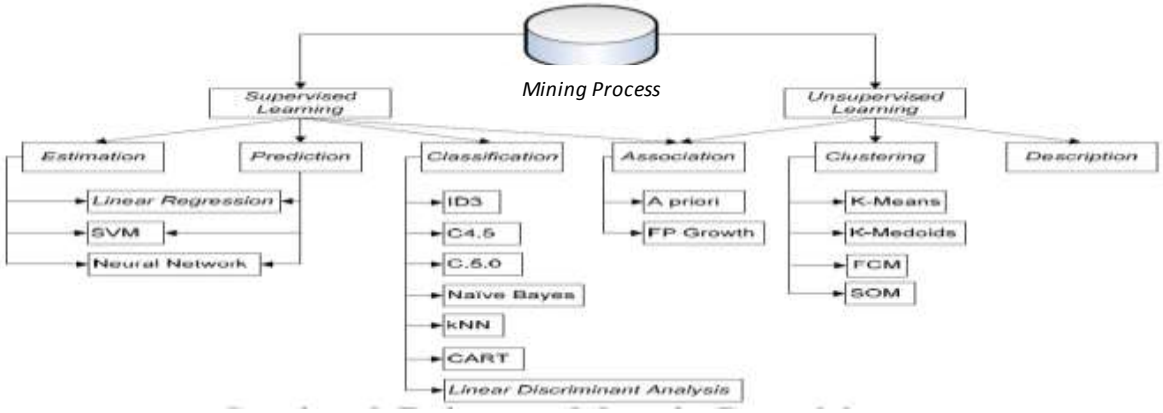

Figure 3. *Data Mining Methods*

2.3 Clustering

Clustering is a method for grouping data that have similarities and then labeled as desired (K. dan P. A. J. Dina Sunia, 2019). The purpose of data clustering can be divided into two, namely grouping for understanding and grouping for use (F. I. Sri Rahayu, 2014).

In Data Mining there is a method called Data Clustering which is unsupervised. In the process of grouping data in data clustering, there are two types, namely hierarchical data clustering and non-hierarchical data clustering.

The non-hierarchical clustering method begins by determining the number of clusters desired. After the number of clusters is known, then the clustering process can be carried out without following the hierarchical process. This method is called K-Means Clustering (A. Maulana, 2018).

2.4 Algoritma K-means Clustering

K-means is one of the simplest unsupervised learning algorithms used to solve various grouping problems (A. V. D. Sano, 2016). K-Means is a distance-based clustering method that divides data into several clusters and algorithms, and this method can only be used in numeric attributes. The K-Means algorithm includes partitioning clustering that separates data into k separate sub-regions. The K-Means algorithm is known for its convenience and ability to cluster large data and outliers quickly. In the K-Means algorithm, each data must belong to a certain cluster and it can be possible for each data that belongs to a certain cluster at one stage of the process, at the next stage it moves to another cluster. (Y. Darmi, 2016).

K-Means is a way to sort data into several groups so that the group is homogeneous among its members or in groups that form the smallest data variation .

Stages of the K-Means Clustering Algorithm (G. Gustientiedina, 2019):

1. Determine **k**, namely the number of clusters to be formed
2. Determining the initial **k** cluster center point (centroid) which is done randomly from the available objects as many as k clusters, to calculate the next centroid of the ith cluster, as follows:

   $$v = \frac{\sum_{i=1}^{n} X_i}{n} : i = 1,2,3 \ldots n \quad \ldots\ldots\ldots\ldots (1)$$
3. Calculate the distance from each object to each centroid of each existing cluster using the Euclidean Distance formula, as follows:

   $$d(x,y) = ||x-y|| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} : i = 1,2,3, \ldots n. (2)$$
4. Moves data from each object to the nearest centroid. The allocation of objects into each cluster during iteration is generally carried out using the hard K-Means, where each object is explicitly stated as a member of the cluster by measuring its proximity to the cluster center point.
5. Perform iteration and then determine the position of the new centroid using the equation.
6. Repeat step three if the new centroid positions are not the same.

III. Discussion

The data source for data mining used is from operational databases, so it is necessary to withdraw data first so as not to interfere with operations. The following is the Sql command for withdrawing data:

```
SELECT
  month(lup) as tgl,
  update_by,
  count(DISTINCT ncli) as jml_data
FROM
  trans_profiling_verifikasi
WHERE
  DATE( lup ) BETWEEN '20200401'
  AND '20201130'
GROUP BY
update_by, tgl
```

Figure 4. *Unique Data Retrieval SQL Command by ncli (Client Number)*

This distinct command is meant not to be counted for double submit data in ncli. The next step is to perform data cleansing.

1. Data Cleansing
   Inconsistent data that will be eliminated is eliminating data:
   a. update_by is null, removes null ncli
   b. ISNUMERIC (ncli) = 0, remove non-numeric ncli (invalid data)
   c. Jml_data < 1760, eliminate agents whose HK (working days) is less than 22 days per month, i.e. with a minimum achievement per day: 80.
2. Data Integration
   Data integration is performed to combine the data agent (update_by) with the WFH and WFO programs. As for the orders:

```
SELECT
  month(trans_profiling_verifikasi.lup) as tgl,
  trans_profiling_verifikasi.update_by,
  count(DISTINCT trans_profiling_verifikasi.ncli) as jml_data,
  sys_user.kategori,
  sys_user.tl
FROM
  sys_user
LEFT JOIN trans_profiling_verifikasi
ON sys_user.agentid = trans_profiling_verifikasi.update_by
WHERE
  DATE( lup ) BETWEEN '20200401'
  AND '20201130'
GROUP BY
trans_profiling_verifikasi.update_by, tgl
```

Figure 5. *The Sql Command Integrates Agent Performance Data with Categories*

3. Data Selection
   Data selection at this stage the authors omit the TL (team leader) field data, because during this year there has been no team change.
4. Data Transformation
   Data transformation is performed to convert the available table formats into a processable format for data mining. The following changes the data format:

Table 1. Format Data Before Transformation

Date	Update_by	Jml_data	category	Tl
5	AF6540	1000	WFH	1
5	NI3506	1000	WFO	2
5	BDGCTS_04 1	1000	WFH	1
5	RA2708	1000	WFO	2
5	SP2089	1000	WFH	1
5	AI0292	1000	WFO	2
5	AR180293	1000	WFH	1
5	DE7748	1000	WFO	2
5	AF6540	1000	WFH	1
...	...	...	...	...

Table 2. Format Data after Transformation

Update_by	Category	April	...	November
(Id)	(label)	(reguler)	(reguler)	(reguler)
AF6540	WFH	1000	1000	1000
NI3506	WFH	1000	1000	1000
BDGCTS_041	WFH	1000	1000	1000
RA2708	WFH	1000	1000	1000
SP2089	WFH	1000	1000	1000
AI0292	WFH	1000	1000	1000
AR180293	WFH	1000	1000	1000
DE7748	WFH	1000	1000	1000
DR2891	WFH	1000	1000	1000
BDG_003	WFH	1000	1000	1000
DH1297	WFH	1000	1000	1000
DP0395	WFH	1000	1000	1000
DA1096	WFH	1000	1000	1000
ME2205	WFO	1000	1000	1000
...	....	....	....	....

5. The Mining Process

- Import result transformed data

Figure 6. *Import Result Transformed Data*

- Process Design

Figure 7. *Kmeans Clustering Design Process with Rapidminer*

- Result

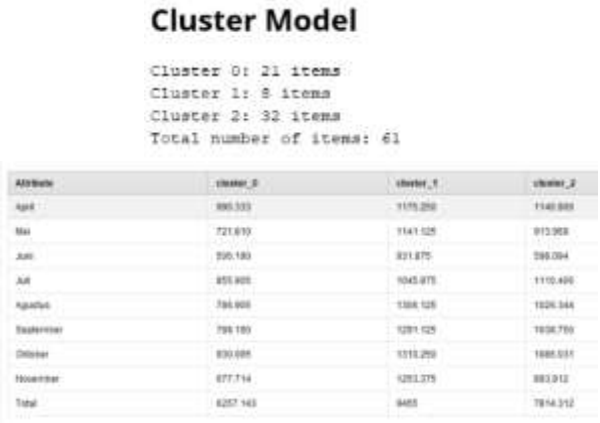

Figure 8. *Resume Clustering Result*

6. Evaluate Patterns

Figure 9*. Clustering Total Performance Agent*

In the clustering pattern, the total performance agent as a whole can be seen that cluster_2 is the largest cluster, while cluster_0 is the second largest cluster.

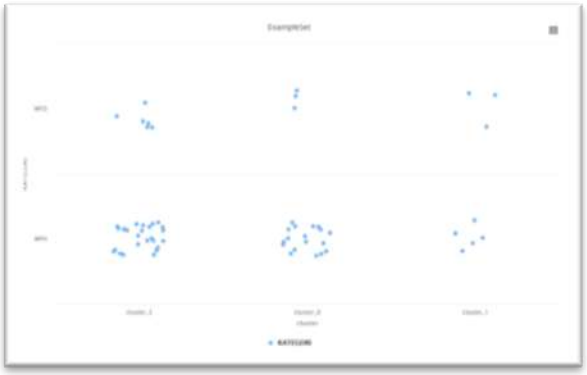

Figure 10. Clustering Agent Category

In this clustering agent category, it illustrates that the proportion of WFH & WFO agents is more in cluster_2, but WFO agents have the same proportion on cluster_0 and cluster_1.

7. Knowledge Presentation

This knowledge presentation is carried out to map the knowledge resulting from data

mining processing, the following points can be described from table 3:

- The performance agent at cluster_1 (best performance) has a ratio of WFH and WFO agents of 10.2%: 25%. This means that WFO is more effective in improving agent performance.
- Performance agent on cluster_0 (lowest performance) is more on WFH agent than WFO agent with a ratio of 36.7%: 25%.
- Meanwhile, for agents with cluster_2 (moderate performance) the ratio of WFH and WFO is: 53.1%: 50%.

Table 3. Percentage of WFH & WFO Clusters

Cluster	WFH	% WFH	WFO	% WFO	Info
Cluster_0	18	36,7%	3	25%	Low
Cluster_1	5	10,2%	3	25%	High
Cluster_2	26	53,1%	6	50%	Medium

Based on the results of the clustering, coaching, refreshment and rewarding programs can be carried out as described in the following table:

Tabel 4. Program Increase Performance Agent

Cluster	Coaching	Refreshment	Rewarding
Cluster_0			
Cluster_1			
Cluster_2			

IV. Conclusion

The results of clustering using K-Means, there are interesting conclusions. Namely, WFO agents are better at maintaining performance, while WFH agents have a lower percentage of agents in the best performance cluster. In terms of follow-up, management can perform different treatments for each predetermined cluster, such as cluster_0 for coaching, cluster_2 for refreshment programs, and cluster_1 for rewards programs.

Monitoring on the WFH agent needs to be improved, especially for the cluster_0 agent, while for the WFO agent on cluster_0 it is necessary to monitor on the spot by TL (Team Leader).

References

F. Gullo, "From patterns in data to knowledge discovery: What data mining can do," *Phys. Procedia*, vol. 62, pp. 18–22, 2015, doi: 10.1016/j.phpro.2015.02.005.

Alfiqra and F. Y. Alfizi, "Penerapan Market Basket Analysis Menggunakan Proses Kdd (Knowledge Discovery In Database) Sebagai Strategi Penjualan Produk Swalayan (Studi Kasus : Swalayan X) Alfiqra," pp. 509–516, 2018.

J. Eska, "Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5," vol. 2, 2018, doi: 10.31227/osf.io/x6svc.

L. Mushunje, "Fraud Detection and Fraudulent Risks Management in the Insurance Sector

Using Selected Data Mining Tools," *Am. J. Data Min. Knowl. Discov.*, vol. 4, no. 2, p. 70, 2019, doi: 10.11648/j.ajdmkd.20190402.13.

Agus Nur Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: https://ilmukomputer.org/category/datamining/.

S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.

L. Muflikhah, D. Ratnawati, and R. Putri, "Data Mining - Lailil Muflikhah, Dian Eka Ratnawati, Rekyan Regasari MP - Google Buku." p. 128, 2018, [Online]. Available: https://books.google.co.id/books?id=V_NqDwAAQBAJ&printsec=frontcover&dq=data+mining&hl=id&sa=X&ved=0ahUKEwi6jIWMzJbiAhVCcq0KHYVKD-QQ6AEIMDAB#v=onepage&q=data mining&f=false.

M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.

K. dan P. A. J. Dina Sunia, "Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means," *J. Ilm. Mhs. Tek. Inform.*, vol. Vol 1 No 2, no. 2016, pp. 121–134, 2019.

F. I. Sri Rahayu, Dodon T. Nugrahadi, "Clustering Penentuan Potensi Kejahatan Daerah Di Kota Banjarbaru Dengan Metode K-Means," *Kumpul. J. Ilmu Komput.*, vol. 01, no. 01, pp. 33–45, 2014.

A. Maulana and A. A. Fajrin, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, p. 27, 2018, doi: 10.20527/klik.v5i1.100.

A. V. D. Sano and H. Nindito, "Application of K-Means Algorithm for Cluster Analysis on Poverty of Provinces in Indonesia," *ComTech Comput. Math. Eng. Appl.*, vol. 7, no. 2, p. 141, 2016, doi: 10.21512/comtech.v7i2.2254.

Y. Darmi and A. Setiawan, "Penerapan metode clustering k-means dalam pengelompokan penjualan produk," *J. Media Infotama Univ. Muhammadiyah Bengkulu*, vol. 12, no. 2, pp. 148–157, 2016.

S. S. Nagari and L. Inayati, "Implementation of Clustering Using K-Means Method To Determine Nutritional Status," *J. Biometrika dan Kependud.*, vol. 9, no. 1, p. 62, 2020, doi: 10.20473/jbk.v9i1.2020.62-68.

M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

,