International Seminar and Workshop on Public Health Action
https://swopha.dinus.ac.id

K-Means Clustering for Profiling BPJS Health Traders in Semarang by Ability To Pay (ATP) and Willingness To Pay (WTP)

Agnes Oktavi Maharani[1*], Eti Rimawati[2], Respati Wulandari[3],Edy Jaya Kusuma[4], Guruh Fajar Shidik[5]

[1]Faculty of Health, Universitas Dian Nuswantoro, p34202400011@mhs.dinus.ac.id*
[2]Faculty of Health, Universitas Dian Nuswantoro, eti.rimawati@dsn.dinus.ac.id
[3]Faculty of Health, Universitas Dian Nuswantoro, restiw@ymail.com
[4]Faculty of Health, Universitas Dian Nuswantoro, edi.jaya.kusuma@dsn.dinus.ac.id
[5]Faculty of Computer Science, Universitas Dian Nuswantoro, guruh.fajar@research.dinus.ac.id

Article Info
Article History
Submitted: 01/08/2025
Revised: 03/09/2025
Accepted: 26/08/2025
Keywords:
BPJS Health, Traders, RapidMiner, K-Means Clustering

Abstract

This study applies the K-Means clustering algorithm to profile market vendors in Semarang City based on their Ability To Pay (ATP) and Willingness To Pay (WTP) BPJS Health insurance contributions, addressing a notable decline in participation among Non-Salaried Workers (PBPU) in Indonesia's National Health Insurance (JKN-KIS). Utilizing quantitative data from 95 Bulu Market vendors, ATP and WTP were estimated via household income and expenditure analysis through linear regression models performed in R-Commander ($p < 0.05$). Clustering with k=4 yielded optimal segmentation (Davies-Bouldin Index = 0.079), identifying four distinct groups: Opportunistic (low ATP/WTP), Rational (high ATP/moderate-low WTP), Price-Sensitive (financially vulnerable with low ATP/WTP), and Prospective (high ATP and WTP). This novel integration of ATP and WTP in Indonesia's informal sector provides data-driven insights allowing BPJS policy makers to tailor intervention strategies, improving premium compliance within the JKN system. The methodology offers a replicable framework for health insurance segmentation in comparable middle-income Southeast Asian settings, supporting Universal Health Coverage goals.

eISSN 3063-2439

Correspondence Address:
Universitas Dian Nuswantoro
Jl. Imam Bonjol No.207, Pendrikan Kidul, Central Semarang District, Semarang City, Central Java 50131
E-mail: p34202400011@mhs.dinus.ac.id

Introduction

Healthcare services are a critical factor in a nation's development. According to the National Medium-Term Development Plan (RPJMN) in Presidential Regulation No. 18 of 2020, the vision and mission of the President for 2020–2024 is to realize an advanced, independent, sovereign, and distinctive Indonesia based on mutual cooperation (gotong royong)(Presiden Republik Indonesia Peraturan Presiden Republik Indonesia, 2020). The concept of gotong royong is the embodiment of the JKN system in Indonesia. JKN itself has a vision of achieving Universal Health Coverage (UHC) for all Indonesian citizens(Yuniati, 2018). Indonesia's Health Profile Data shows the coverage of JKN/KIS participants in 2023. The highest proportion of participants comes from the PBI (APBN) segment at 48.0%. However, the growth of participants in the non-PBI segment has decreased from the 2022 figure. By the end of 2023, the total coverage of JKN/KIS participants reached 267.3 million people, marking an increase compared to the number of participants in 2022(Kementrian Kesehatan Republik Indonesia, 2023).

Initial survey results from the Semarang Branch Office (KCU) of the BPJS Kesehatan (Social Security Agency for Health) showed a decline in JKN-KIS membership in the PBPU segment in Semarang City over the past three years. The breakdown is as follows: As of December 2019, the number of PBPU participants was 336,734, down to 305,602 in December 2020. The latest data, as of October 2021, shows that PBPU membership has dropped to 304,005. The deficit of the Health Social Security Agency (BPJS Health) has persisted since the inception of the JKN program in 2014 and continued until 2019. In 2019, the BPJS Health deficit reached Rp. 15.5 trillion. The causes of the BPJS Health deficit are estimated to include, JKN premiums that do not align with actuarial calculations, outstanding premiums from JKN participants, unlimited JKN benefit packages without *cost-sharing,* and high healthcare costs for catastrophic illnesses(Khoiri et al., 2025).

The non-PBI segment of PBPU participants is the segment most "responsible" for their willingness to pay BPJS Health contributions (WTP). One of these is informal sector workers, whose unstable income and uncertain business continuity affect their ability to pay BPJS Health contributions (ATP)(Hardika & Purwanti, 2020). Market vendors are classified as informal workers, accounting for the second-largest percentage of the informal sector at 3.31 million people (18.98%), according to the Central Java Statistics Agency's Sakernas survey from August 2017 to 2019(BPS Jateng, 2019).

Semarang Bulu Market, the largest market in Semarang Utara, is located at Jl. Sugiyo Pranoto No. 2, Semarang, with a total of 741 traders. According to a preliminary study by Yunianti, in the working area of the Simpang Timbangan Community Health Center, which focused on traders, the results showed a lack of *Willingness To Pay* (WTP) for JKN contributions among traders. Most respondents had not registered for JKN membership, and one of the factors contributing to the low WTP was the low *Ability To Pay* (ATP) among traders(Yuniati, 2018).

K-Means is one of the most popular clustering algorithms in data analysis. This algorithm functions by dividing a dataset into *k* clusters based on similar characteristics(Aras & Sarjono, 2016). The *K-Means* algorithm is considered to have a high level of accuracy, making it suitable for generating *knowledge* about the characteristics of income, expenditure, ATP, and WTP of market traders regarding BPJS Health contributions in this study. Previous research conducted by Salvan et al., titled "Data Clustering of BPJS Health Service Usage Using the *K-Means* Algorithm," addressed the issue of equalizing BPJS Health service users using 1,000 datasets categorized by healthcare facility names. The study employed the K-Means clustering algorithm to group BPJS Health service usage into several *clusters*(Salvan N et al., 2022).

Despite increasing global adoption of clustering algorithms for health insurance data segmentation, few studies explicitly integrate both Ability To Pay and Willingness To Pay metrics, particularly in Indonesia's informal sector context. Recent research from India and Thailand post-2022 demonstrates clustering methods refining subsidy allocation and outreach (Singh et al., 2023; Charoenwong et al., 2024), yet applications remain limited in Southeast Asia. This study addresses this gap by profiling BPJS Health participants among Semarang market vendors through socio-economic payment capacities via K-Means clustering, contributing to sustainable health financing strategies and expanding the empirical base for JKN reforms consistent with regional trends.

Related Research

Related research conducted by other researchers using the classification algorithm method by Widyaningsih et al. grouped BPJS Health JKN/KIS participants in Cirebon Regency into several *sample* groups. The clustering results using the *Davies Bouildin Index* were used to determine the optimization of the *K-Means* algorithm*.* The clustering results showed that the largest group of BPJS Health participants was Cluster_0: Low Group with 86 participants, Cluster_1: Medium Group with 156 participants, and Cluster_2: High Group with 170 participants. The optimal value was determined as K=3 with a DBI value of 0.164(Widyaningsih & Pratama, 2023).

A study by Sari et al. titled "Application of the K-Means Algorithm for *Clustering* Poverty Data in Banten Province Using *RapidMiner"* grouped participants based on the number of poor people (thousands), average years of schooling, and adjusted per capita expenditure (thousands of rupiah per year). The results identified three clusters: moderate cluster level (C0), high cluster level (C1), and low cluster level (C2). This indicates that Tangerang Regency, Cilegon City, and Serang City are members of cluster 0, Pandeglang Regency, Lebak Regency, and Serang Regency are in cluster 1, and Tangerang City and South Tangerang City are in cluster 2(Ratna Sari et al., 2020).

Another study by Meiriza et al. grouped data from the BPJS Employment branch in Pekanbaru to classify non-wage programs based on their characteristics. Data distribution using *K-Means* resulted in Cluster 1 with 905 data points, Cluster 2 with 5 data points, and Cluster 3 with 90 data points. *K-Means* is the preferred method when computational time is a primary consideration. *K-Means* is faster and sufficiently accurate in grouping unlabeled data(Khatib Sulaiman et al., 2023).

Methodology

This section explains the methods used, the dataset used, the process, and the evaluation used.

Dataset

The dataset used is a private dataset obtained from a journal article on *Ability To Pay* (ATP) and *Willingness To Pay* (WTP) for BPJS Health Insurance Contributions among Market Vendors in Semarang

City. The dataset consists of questionnaire data from a study by Maharani et al. in 2022, where the respondents were market vendors in Semarang City. The type of research used is descriptive research through a quantitative approach. The *Lemeshow* sample formula calculation resulted in a calculation of 95 samples. The sampling technique used purposive sampling based on considerations regarding characteristics appropriate to this study.

Ethical approvals for the study were obtained from Health Research Ethics Committee Faculty of Health Universitas Dian Nuswantoro, approval number 177/EA/KEPK-Fkes-UDINUS/III/2022. Data was collected via structured interviews using a validated questionnaire comprising demographic data, income, expenditure categories, and health insurance payment attitudes. *Ability To Pay* (ATP) and *Willingness To Pay* (WTP) estimations employed linear regression using R-Commander, with significance levels set at $p < 0.05$.

Data collection was conducted through direct interviews using open-ended questionnaires to obtain data on total household income and total expenditures, which were divided into expenditures for food, non-essential food items, and non-food items for the number of households (Ruta/Rumah Tangga) of traders at Bulu Market. The ATP and WTP of traders were analyzed using linear regression through the *R-Commander* statistical program to determine the significant variables and the coefficients of each expenditure variable. The real amount *of Ability To Pay* (ATP) for Fur Market traders had an average value of IDR 84.792,00 per person per month. Calculation of the estimated average *Willingness To Pay* (WTP) obtained was IDR 26.411.000(Maharani et al., 2023).

Preprocessing

The data *preprocessing* stage is a crucial step in preparing a dataset before it can be used for analysis or model building. This process involves several activities, including reading raw data from its source, filling in missing values through data imputation, applying filters to reduce the dataset by removing unnecessary columns such as name, address, village, district, regency, food expenditure details, non-essential food expenditure, non-food expenditure, and the impact of the pandemic. Next, the dataset is normalized to produce a binary representation of the data, enabling efficiency in analysis or *Machine Learning* model training. These steps aim to ensure the cleanliness, completeness, and readiness of the data without sacrificing essential information, which ultimately improves the quality of the analysis or the developed model(Sahamony et al., 2024).

Table 1. Dataset After *Preprocessing*

Id	Gender	Age	Family Members	Is the husband employed	Monthly income	Does the wife work	Husband/wife's income
1	1	55	1	2	IDR 0	1	IDR 2.000.000
2	1	35	2	1	IDR 1.170.600	1	IDR 2.800.000
3	1	45	0	2	IDR 0	1	IDR 2.200.000
4	2	45	1	1	IDR 1.800.000	1	IDR 2.000.000
5	1	40	2	1	IDR 2.700.000	2	IDR 0
...	...	...	...	...	...	...	...
95	2	37	1	1	IDR 2.000.000	1	IDR 2.000.000
Id	**Depend-ent children still in school**	**Income contribution from children**	**Total income of Ruta**	**Food expenses**	**Non-food expenses**	**WTP**	**ATP**
1	2	IDR 1.000.000	IDR 3.000.000	IDR 80.000	IDR 3.079.500	IDR 35.000	IDR 20.494
2	1	IDR 0	IDR 3.970.600	IDR375.000	IDR 3.821.500	IDR 35.000	IDR 149.100
3	2	IDR 150.000	IDR 2.350.000	IDR 180.000	IDR 2.084.500	IDR 30.000	IDR 265.500
4	1	IDR 0	IDR 3.800.000	IDR 620.000	IDR 3.667.800	IDR 30.000	IDR 132.200
5	1	IDR 500.000	IDR 3.200.000	IDR 80.000	IDR 3.037.100	IDR 20.000	IDR 162.900
...	...	...	...	...	...	...	...

95	1	IDR 0	IDR 4.000.000	IDR 410.000	IDR 3.949.000	IDR 35.000	IDR 51.000

The dataset contains numerical data used for clustering ATP and WTP for BPJS Health Contributions among traders.

Table 2. Types Of Market Vendor Datasets

Data Set	Description
Gender	Binomial
Age	Interger
Family member	Interger
Is the husband employed	Interger
Monthly income	Interger
Does the wife work?	Interger
Spouse's income	Interger
Children still in school	Interger
Contribution to income from children	Interger
Total income	Interger
Total food expenditure	Interger
Total non-food expenditure	Interger
WTP	Interger
ATP	Interger

Research Flow

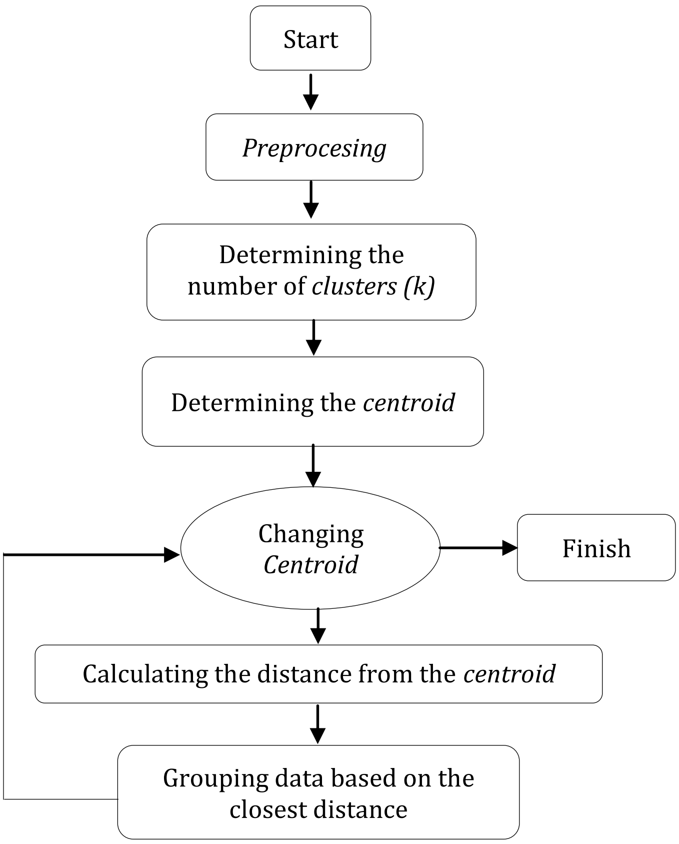

Figure 1. *K-Means* Algorithm Implementation Flow

Data Mining

Data mining is a process of implicit data filtering(Nasution et al., 2021). The purpose of this method is to determine the value of data by exploring new patterns from an available dataset(Marliska Hutabarat et al., 2019). Data mining is an analysis in concluding hypotheses through understanding data so that it is useful for researchers when results have been found and conclusions have been drawn, as well as for users as reference material.

Clustering

Clustering is a process of grouping data into several groups based on similarities in the data(Herlinda & Darwis, 2021). *Clustering* is the process of forming *groups/clusters* that were previously unknown to anyone through the dataset used(Priyatman et al., 2019).

K-Means

The *K-Means* algorithm is a grouping method to determine the size of groups through initial *centroid* values by repeating the process periodically(Indraputra & Fitriana, 2020).

The following are the steps in determining the *K-Means* algorithm:

1. Determine the value of K or the number of clusters in the dataset
2. Initial *centroid* values are randomly assigned. During the iteration phase, centroid values are calculated using the following formula:

$$\text{Vij} = \frac{1}{Ni}\sum_{k=0}^{Ni} X_{kj} \qquad (1)$$

Explanation:

Vij = Average *centroid of* the first cluster for variable j
Ni = Number of members in the 1st cluster
I,k = Index of the cluster
j = Index of variable
Xkj = K data value of the J variable for that cluster

3. Calculate the distance between the *centroid* and the object point using *Euclidean Distance* with the following formula:

$$\text{De} = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \qquad (2)$$

Explanation:

De = *Euclidean Distance*
i = Number of Objects
(x,y) = Object Coordinates
(s,t) = *Centroid* Coordinates

4. Grouping objects based on their distance to the nearest *centroid* through iteration until the *centroid* value is optimal.

RapidMiner

RapidMiner is an *open-source software tool* that is useful for data mining processes by first examining and analyzing patterns and attributes that will be used(Faid et al., 2019). It is designed to simplify complex data analysis, providing a *user-friendly* graphical interface, enabling users to create, manage, and visualize analytical workflows without requiring advanced programming skills(Ratna Sari et al., 2020).

Davies Bouildin Index

The *Davies-Bouildin Index* is one of the internal evaluation methods that can test the results of a cluster evaluation that will be carried out. The smaller the DBI value obtained through trials with non-negative values, the closer the value is to 0, the better the results will be in terms of the number of clusters determined.

Simulation and Result

Simulation

In this section, we will discuss the results of the *K-Means Clustering Algorithm* modeling in *RapidMiner*.

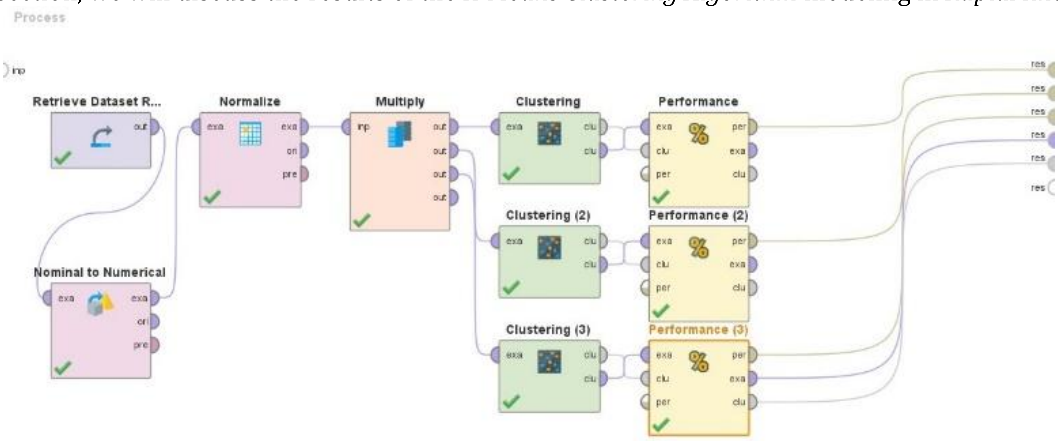

Figure 2. *Clustering* process using the *K-Means* algorithm in *RapidMiner*

Figure 1 shows the configuration of the *K-Means* design process in *RapidMiner*. The data is divided into three clusters named *Clustering, Clustering* (2), and *Clustering* (3). After the division in the Segmentation section, three *Cluster Distance Performance* tests are conducted, named *Performance, Performance* (2), and *Performance* (3), to evaluate the quality of the *clustering* results based on the distance between and within *clusters*. In the *Performance* section, the *main criterion* is selected as *the Davis-Bouldin Index* to identify the quality of data clustering and make adjustments if necessary to improve *clustering* results*.*

Results

Table 3. *Clustering Performance* Test Results For The *Max* Run/Iterations=10 Parameter Value

K Value	*Average distance within centroid*	*Davies Bouldin Index (DBI)*	*Cluster*
2	0.062	0.108	*Cluster* 0: 23 *Cluster* 1: 72
3	0.045	0.092	*Cluster* 0: 53 *Cluster* 1: 19 *Cluster* 2: 23
4	0.034	0.079	*Cluster* 0: 41 *Cluster* 1: 18 *Cluster* 2: 22 *Cluster* 3: 14

Table 3 shows the results of the *Performance Clustering* test using the *K-Means* algorithm performed on the three clusters. The evaluation results using the *Davies Bouldin Index* (DBI) at k=4 show a value of 0.079. The smaller the index value, the better the clustering results. This indicates that with four clusters, the clustering model is able to form *clusters* that are sufficiently compact and well separated from one another.

When the *K-Means* algorithm is applied with k=3, the DBI evaluation yields a value of 0.092, which is lower than with four clusters, indicating an improvement in clustering quality*.* With fewer clusters, the identified patterns become more detailed, although some clusters may not appear fully separated.

Finally, when applied with k=2, the DBI value becomes 0.108, indicating a decrease in clustering quality. Clusters become smaller and more specific, but the distance between clusters becomes less significant. The results suggest that some clusters may not provide useful information for further analysis.

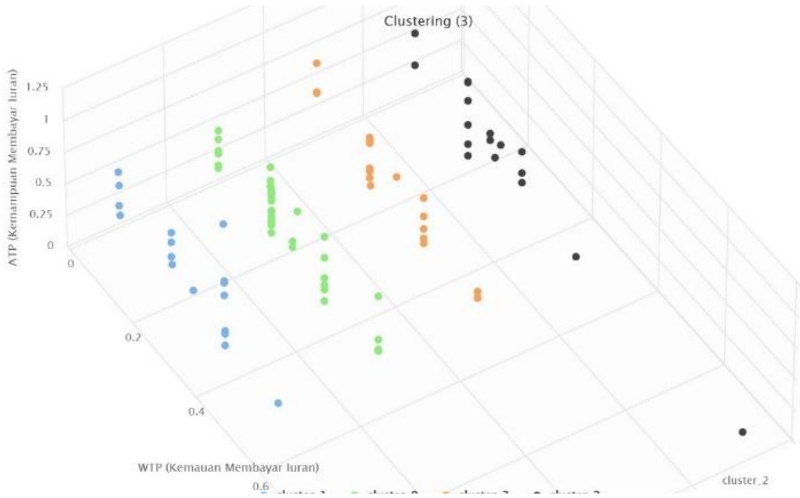

Figure 3. Display of *K-Means* clustering results with a 3D *scatter* diagram. Display of data distribution in *K-Means*

Figure 3 shows the 3D *Scatter* plot, X = *Ability To Pay* (ATP), Y = *Willingness To Pay* (WTP), Z = Total Income (IDR), which identifies four clusters, represented by different colors with separate *cluster* distributions.

The results can be seen in Table 4.

Table 4. Characteristics And Performance Metrics Of Clusters Among Traders

Cluster & Persona	Cluster Size	Compactness (*Avg. Dist.*)	Average ATP (Reputation Score)	Average WTP (Reputation Score)	Implications
Cluster_0 (Opportunists)	41 *Items* (43.2%)	0.019 (Very Cohesive)	0.21	0.15	Highest default risk
Cluster_1 (Rational)	18 *Items* (18.9%)	0.044 (Moderate)	0.82	0	High service quality demands
Cluster & Persona	Cluster Size	Compactness (*Avg. Dist.*)	Average ATP (Reputation Score)	Average WTP (Reputation Score)	Implications
Cluster_2 (Price-Sensitive)	22 *Items* (23.1%)	0.059 (Least Divergent)	0.25	0	Primary target of the subsidy program
Cluster_3 (Prospective)	14 *items* (14.7%)	0.018 (Very Compact)	0.85	0.90	Highest income potential

Table 4 shows the characteristics and performance metrics of each *cluster*. The Opportunistic Segment (Cluster_0) is the largest group in the respondent population (43.2%). Quantitatively, they are very homogeneous with a compactness of 0.019, which means that their characteristics are very similar. Their main profile is low ability to pay (ATP) and very low willingness to pay (WTP). The behavior of this segment tends to be opportunistic, meaning they will participate or pay premiums only if there is an urgent need or if the cost is very minimal. Implications for BPJS Health due to its large size and weak financial profile, this segment is the primary source of potential premium arrears. An effective approach is not aggressive collection but rather a supportive and preventive strategy. The BPJS Health Graduated Payment Plan (REHAB) is well-suited for this *cluster*. Education strategies should focus on the ease of payment methods, financial consequences, and services available if membership becomes inactive.

The Rational Segment (Cluster_1) consists of financially capable individuals, as indicated by high ATP values. However, this is not matched by a corresponding willingness to pay (WTP), which remains at a medium-to-low level. This indicates that this group is critical and makes decisions based on *value for money*. These merchants will not default on payments due to inability to pay, but rather because they perceive the quality of services as not commensurate with the premiums they pay. To maintain collection rates from this segment, BPJS Health must prioritize improving the quality and transparency of services at healthcare facilities (Faskes). Prompt complaint handling, ease of access, and evidence-based communication about program success will be highly effective in convincing this segment and maintaining their compliance.

The Price-Sensitive Segment (Cluster_2) represents the most financially vulnerable group, with the lowest ATP and WTP values. Mathematically, this cluster is also the most heterogeneous (diverse), indicating that although they all share financial vulnerability, their other backgrounds (such as demographics or geography) may vary greatly. This group is the primary representation of participants in the Premium Assistance Program (PBI) or self-paying participants in the lowest income decile. Their payment compliance is almost entirely dependent on government subsidies. Therefore, they are not the target of collection strategies but the primary focus of social protection programs. Data on this segment is crucial for policy advocacy with the government regarding the importance of sustaining the PBI program.

The Prospective Segment (Cluster_3), despite being the smallest segment (14.7%), is the most ideal from a financial and compliance perspective. They are highly homogeneous and characterized by the highest ATP and WTP values. This is the profile of participants who are aware of the importance of health insurance, have financial stability, and demonstrate high compliance levels. The Prospective Segment serves as the pillar supporting stable premium cash flow. The primary strategy for this segment is retention and maximizing payment ease, making them the top candidates for automatic debit programs. In the era of Standard Inpatient Care (KRIS), this segment is also the most potential to accept *co-payment* schemes or additional benefits if such policies are implemented.

This profiling approach aligns with international studies employing K-Means for health insurance participant segmentation in middle-income countries such as India (Saxena et al., 2023) and the Philippines (De la Cruz et al., 2024), highlighting financially vulnerable groups that require tailored subsidy measures. However, our study's limited sample size (n=95) and purposive sampling restrict the generalizability beyond Semarang. Furthermore, potential biases inherent in self-reported income and expenditure data may affect precision in ATP and WTP estimates. Future research should expand with larger randomized samples across multiple regions and explore non-financial factors influencing participant behavior, including cultural and service quality perceptions.

Conclusion

Data processing using *RapidMiner* software with the *K-Means* method can provide optimal merchant segmentation with a DBI value of 0.079 at k=4. The government or BPJS Health can use the clustering results to contribute to identifying different segments of market merchants. Strategies can be adjusted based on the characteristics of each cluster (high WTP, low ATP, income, food expenditure, non-food expenditure, etc.), such as socialization strategies, billing, or incentive programs, to be tailored to each segment for greater effectiveness in increasing active participation in BPJS Health.

Understanding behavior involves understanding how WTP, ATP, income, food expenditure, and non-food expenditure are interrelated and shape the behavior of different groups of traders. Overall, this visualization provides a clear picture of the characteristics of market traders in the context of *Ability To Pay* (ATP) and *Willingness To Pay* (WTP) BPJS Health contributions, which is highly useful for decision-making and policy formulation.

Piloting the REHAB graduated payment program for Cluster_0 using mobile app platforms to simplify premium payments and enhance compliance. Annual updates to the segmentation using ongoing data collection are advised, supporting dynamic policy adaptation aligned with Sustainable Development Goal 3. Cross-sector collaboration across ministries and BPJS Health remains essential to align subsidy policies with actual community payment abilities and improve health insurance sustainability.

BPJS Health needs to strengthen its information system regularly by conducting similar segmentation and integrating it into management dashboards for real-time decision-making. There is a need for cross-sectoral government collaboration (through relevant ministries such as the Ministry of Finance, Ministry of Social Affairs, Ministry of Health, and BPJS Health) to ensure that contribution policies and subsidies align with the actual economic capacity of the community. Conduct further research to understand the non-financial factors influencing participation decisions and utilization of health services across each segment.

Acknowledgement

The authors would like to thank the committee of the 5th International Seminar and Workshop on Public Health Action (ISWOPHA 2025) for the opportunity and support in this publication process.

References

Aras, Z., & Sarjono. (2016). Analisis Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan Bedah Rumah Menggunakan Metode Clustering K-Means (Studi Kasus: Kantor Kecamatan Bahar Utara). *Jurnal Manajemen Sistem Informasi*, *1*, 159–170.

BPS Jateng. (2019). Keadaan Ketenagakerjaan Provinsi Jawa Tengah Agustus 2019. *Berita Resmi Statistik*, 1–11.

Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, *8*(1), 11–16. https://doi.org/10.34148/teknika.v8i1.95

Hardika, C. P., & Purwanti, E. Y. (2020). Analisis Willingness To Pay Terhadap Iuran BPJS Kesehatan Pada Pekerja Sektor Informal di Kota Semarang. *Diponegoro Journal Of Economics*, *9*(3). http://ejournal-s1.undip.ac.id/index.php/jme

Herlinda, V., & Darwis, D. (2021). Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means. *Darwis, Dartono*, *2*(2), 94–99. http://jim.teknokrat.ac.id/index.php/JTSI

Indraputra, R. A., & Fitriana, R. (2020). *K-Means Clustering Data Covid-19*.

Kementrian Kesehatan Republik Indonesia. (2023). *Profil Kesehatan Indonesia 2023*.

Khatib Sulaiman, J., Meiriza, A., & Ali, E. (2023). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan. *Indonesian Journal of Computer Science*.

Khoiri, A., Sandra, C., Ngakan, D., Wahyu, G., & Herawati, Y. T. (2025). *Analisis Willingness to Pay (WTP) Peserta BPJS Kesehatan dengan Pendekatan Contingent Valuation Method* (Vol. 6, Issue 2).

Maharani, A. O., Nurmandhani, R., Kusuma, A. P., & Iqbal, M. (2023). Ability To Pay dan Willingness To Pay Iuran BPJS Kesehatan Pada Pedagang Pasar Bulu Semarang di Era Pandemi. *Visikes Jurnal Kesehatan*.

Marliska Hutabarat, S., Sindar, A., & Pelita Nusantara Jl Iskandar Muda No, S. (2019). Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, *2*(2).

Nasution, R. W., Suhada, I., Kirana, I. G., & Sari, I. P. (2021). Penerapan Data Mining untuk Pengempokan Minat Konsumen Terhadap Pengguna Jasa Pengiriman Pada PT. Jalur Nugraha Ekakurir (JNE) Pematangsiantar. *Resolusi Rekayasa Tek. Inform. Dan Inf*, 274–281.

Presiden Republik Indonesia Peraturan Presiden Republik Indonesia (2020).

Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*.

Ratna Sari, Y., Sudewa, A., Ayu Lestari, D., & Ika Jaya, T. (2020a). Penerapan Algoritma K-Means untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering System and Science)*, *5*(2), 2502–2714.

Ratna Sari, Y., Sudewa, A., Ayu Lestari, D., & Ika Jaya, T. (2020b). *Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan RapidMiner* (Vol. 5, Issue 2).

Sahamony, N. F., Terttiaavini, T., & Rianto, H. (2024). Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(2), 413–422. https://doi.org/10.57152/malcom.v4i2.1210

Salvan N, S., Hadikristanto, W., & Edora, E. (2022). Klasterisasi Data Penggunaan Layanan BPJS Kesehatan Menggunakan Algoritma K-Means. *Prosiding Sains Dan Teknologi*, 121–127.

Saxena, A., Taluja, R., Sararswat, M., Shalini, N., Krishna, O., & Kumar, P. (2023). Data Mining Methods for Internet of Things: A Survey. *6th International Conference on Contemporary Computing and Informatics (IC3I)*, 2610–2618.

Widyaningsih, T. A., & Pratama, D. (2023). *Penerapan Metode K-Means Clustering Dalam Mengelompokkan Jumlah Peserta BPJS Kesehatan JKN/KIS DI Kabupaten Cirebon*. *18*, 17–27.

Yuniati, AD. (2018). Willingness To Pay Timbangan kabupaten Ogan Ilir Tahun 2018. *Jurnal Kesehatan Masyarakat*.