Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14611
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Business Intelligence: Enhancing Apache Superset Capabilities in PBB-P2 Receivables Monitoring

Sugeng Pranoto[1)*], Sri Wahyuni[2)], Muhammad Syahputra Novelan[3)]
[1)2)3)] Universitas Pembangunan Panca Budi, Medan
[1)]stu.genk@gmail.com, [2)]sriwahyuni@dosen.pancabudi.ac.id, [3)]putranovelan@dosen.pancabudi.ac.id

Submitted :March 1, 2025 | **Accepted** : April 17, 2025 | **Published** : April 17, 2025

Abstract: PBB-P2 Tax Revenue plays an essential role in regional finance, but managing receivables and analyzing taxpayer compliance levels still face many challenges. Business Intelligence (BI) technologies such as Apache Superset are often used for interactive data visualization. Still, they have limitations in advanced analysis, especially the application of machine learning algorithms such as K-Means for data clustering. This research aims to overcome the limitations of Apache Superset by developing an external application-based solution using the Java programming language and the SMILE library. This application is designed to cluster the level of taxpayer compliance in a batch process, with the results stored in the MySQL database. The clustered data is then visualized using Apache Superset. The results show that integrating these external applications can improve the efficiency of data analysis by utilizing more complex clustering algorithms. Visualization of clustering results also allows for more effective management of PBB-P2. This approach not only expands the capabilities of Apache Superset but also contributes to supporting data-driven tax revenue optimization strategies. This research opens up further opportunities for the integration of BI tools with machine learning algorithms in monitoring and managing complex data in the tax sector.

Keywords: Data Mining, Business Intelligence, K-Means, Java, Apache Superset

INTRODUCTION

Revenue from PBB-P2 Tax is one of the primary sources of regional revenue(Putera, Siahaan, Jabar, et al. 2024)(Sitorus, Pranoto, and Sutiono 2024) which requires effective and efficient management. However, challenges in PBB-P2 management, primarily related to the monitoring of receivables and analysis of taxpayer compliance levels, often require a technology-based approach. One approach that can be used is the use of Data Mining and Business Intelligence (BI) to support data-driven decision-making(Tavera Romero et al. 2021)(Macías and Borges 2024)(Bany Mohammed et al. 2024). Data mining is used in large-scale data mining to generate hidden information(Wahyuni and Marbun 2020)(Wahyuni 2018)(Wahyuni et al. 2019). The scale and complexity of data pose significant challenges to traditional data processing and analysis techniques(Greca, Shehi, and Nuhi 2023)(Weichbroth, Zurada, and Olszak 2024). Data mining can also solve a problem that must be optimized(Gallego et al. 2024)(Novelan et al. 2023).

Apache Superset, a widely used open-source BI platform, provides advanced and interactive data visualization capabilities(Superset 2024)(Wijerathne 2024). However, this platform has limitations in performing complex calculations, including applying machine learning algorithms. This becomes an obstacle when agencies need in-depth analysis that involves non-linear and complex data processing to develop more effective tax policy strategies. These weaknesses prompted the need to create external solutions to complement the capabilities of Apache Superset. One alternative is to build a supporting application using the Java programming language and the SMILE (Statistical Machine Intelligence and Learning Engine) library to cluster the level of taxpayer compliance based on the observation of the required variables. In clustering, K-Means can be used (Zhao, Deng, and Ngo 2018) because the process is faster and easier(Putra et al. 2021). This clustering process will be carried out in batches, where the data from the clustering will be stored in a MySQL database to be visualized using Apache Superset.

This model approach not only allows for deeper data exploration but also supports the use of machine learning algorithms in data-driven decision-making. Thus, this research contributes to improving the efficiency of PBB-P2

*name of corresponding author

management and expanding BI's technological capabilities in supporting strategic analysis and planning in the field of taxation.

This research improves Apache Superset's capabilities as a Business Intelligence tool for advanced data analysis while offering innovative solutions in the form of custom application integration that can efficiently run clustering calculations.

LITERATURE REVIEW

Previous Studies or Research

Table 1. Previous Studies

Author	Topic	Conclusion
Fatha et al., (2023)	The Role of Big Data in Business Intelligence as a Decision Support System (Systematic Literature Review)	Analyzing the role of Big Data in business intelligence (BI) as a decision support system.
Siahaan et al., (2024)	Application Of Business Intelligence in Decision Support in Providing Assistance to Business Actors in Deli Serdang Regency Using The Decision Tree Algoritm	Analyzing the role of Big Data in business intelligence (BI) as a decision support system
S. Pranoto & D. Nasution, (2024)	Business Intelligence Menggunakan Apache Superset untuk Sistem Pendukung Keputusan Kebijakan Penagihan Pajak Bumi dan Bangunan : Studi Kasus BPKPD Kota Tebing Tinggi	Application of Intelligence Using Apache Superset as a Decision Support for PBB-P2 Tax Collection Policy.
Iqbal et al., (2024)	Analysis of Student Achievement with K-Means on Socioeconomic, Behavioral, and Psychological Factors	Analyze students' academic achievement based on socio-economic, behavioral, and psychological factors using the K-Means clustering method.
Putra et al., (2021)	Application Of The K-Means Algorithm In Identifying Types Of Skin Disease	Use of K-Means algorithm for grouping skin disease types.

From previous studies, it is concluded that data maining and Business Intelligence (BI) provide a strong foundation for organizations to manage and analyze data effectively to support decision-making. Research that specifically examines the use of Apache Superset as a Business Intelligence (BI) tool and data maining using clustering algorithms in the context of monitoring PBB-P2 receivables at the local government level is still rare, so this study aims to fill this gap.

METHOD

Research Stages

His study uses a descriptive and experimental approach that focuses on developing and implementing application solutions [25] to support the monitoring of PBB-P2 receivables. The stages of the research are explained as follows:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14611
e-ISSN : 2541-2019
p-ISSN : 2541-044X

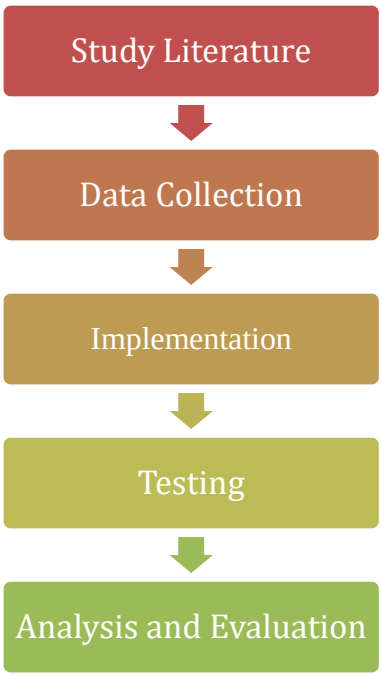

Figure 1. Waterfall Method\lllpbn;pp[p

Table 2. Explanation of the Research Flow

Stage	Method/Tool	Output
Study Literatute	Conventional and online literature	This stage seeks references and theories about PBB-P2, Apache Superset, Java Libraries using SMILE in clustering, and other supporting theories. In this study, it was carried out at BPKPD Tebing Tinggi City.
Data Collection	Collect PBB-P2 Receivables Data in Oracle Database using SQL Developer, and create a result table from clustering in MySQL with MySQL Administrator.	Generate Data View normalization of receivables and payment data. Datasets for clustering by having the fields of Arrears Frequency, Delinquency Rate, and Number of Days of Delay.
Implementation	Using the K-Means Algorithm, create an application using Java and the SMILE Library for Clustering. Designing Visualizations.	The result of this stage is an application that clusters the compliance level of PBB-P2 taxpayers with the K-Means algorithm. Datasets, Charts, and Dashboards on Apache Superset that function for visualization in monitoring PBB-P2 receivables.
System Testing	Test the app already built and run the visualization on the Apache superset.	The results of clustering stored in the MySQL database visualization of receivables data and clustering of PBB-P2 tax receivables
Analysis & Evaluation	Analyze and evaluate clustering results using the K-Means algorithm and clustering results with Apache Superset.	Knowing the centroid of each cluster at the level of compliance (compliant, less compliant, and non-compliant) and calculating the percentage of data for each cluster. Calculate the Silhouette Score to measure how well the data in each clustered cluster is performing. Visualization results in monitoring PBB-P2 receivables at the City, Sub-district, and Village levels.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14611
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Data Collection

In managing PBB-P2 receivables and payment data, BPKPD Tebing Tinggi City uses an application, Sistem Informasi Objek Pajak (SISMIOP), that uses Oracle databases. The first data needed is combining receivables with payments to produce data on unpaid receivables. This data is already available on the database with the name view HIS_PEMBAYARAN_LENGKAP with the following table structure :

Table 3. Structure View HIS_PEMBAYARAN_LENGKAP PBB-P2

Colum Name	Data Type	Description
NOP	CHAR(18)	Tax Object Number
NM_WP	VARCHAR2(30)	Taxpayer Name
NM_KECAMATAN	VARCHAR2(30)	Name of District
NM_KELURAHAN	VARCHAR2(30)	Name of the Village
TOTAL_TUNGGAKAN	NUMBER	Total Arrears
FREQ_TUNGGAKAN	NUMBER	Frequency of Arrears
RASIO_TUNGGAKAN	NUMBER	Arrears Ratio
TOTAL_HARI_TERLAMBAT	NUMBER	Total late payment days

The next step is to create a clustering dataset with Delinquency Frequency, Delinquency Ratio, and Number of Late Days. This dataset will later be processed in clustering using Java applications and Libarry SMILE, which has the K-Means algorithm function and is stored in the MySQL database. This dataset is named DS_KMEANS and is built by creating a View in the Oracle database. The View structure that will be built is as follows:

Table 4. Structure View DS_KMEANS

Colum Name	Data Type	Description
NOP	CHAR(18)	Tax Object Number
NM_WP	VARCHAR2(30)	Taxpayer Name
NM_KECAMATAN	VARCHAR2(30)	Name of District
NM_KELURAHAN	VARCHAR2(30)	Name of the Village
THN_PAJAK_SPPT	CHAR(4)	Tax Year
PBB_YG_DIBAYAR_DIBAYAR_SPPT	NUMBER	PBB-P2 receivables
TGL_JATUH_TEMPO_SPPT	DATE	PBB-P2 Due Date
JML_SPPT_YG_DIBAYAR	NUMBER	Total Payment of PBB-P2 Receivables
TGL_PEMBAYARAN	DATE	Payment Date

In addition to the Data View and Dataset mentioned above, a table is also needed that will later be used to accommodate the clustering results by Java applications. This data is also used as a data visualization source in Apache Superset to monitor the clustering of PBB-P2 receivables in the level of taxpayer compliance, created using a MySQL database named DS_PBB. The table structure is as follows:

Table 5. Structure View DS_PBB

Colum Name	Data Type	Description
NOP	CHAR(18)	Tax Object Number
NM_WP	VARCHAR2(30)	Taxpayer Name
NM_KECAMATAN	VARCHAR2(30)	Name of District
NM_KELURAHAN	VARCHAR2(30)	Name of the Village
TOTAL_TUNGGAKAN	NUMBER	Total Arrears
FREQ_TUNGGAKAN	NUMBER	Frequency of Arrears
RASIO_TUNGGAKAN	NUMBER	Arrears Ratio
TOTAL_HARI_TERLAMBAT	NUMBER	Total late payment days
KATEGORI	INTEGER	Clustered product categories

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14611
e-ISSN : 2541-2019
p-ISSN : 2541-044X

System Design

The System Design to be built can be described in the System Architecture as follows:

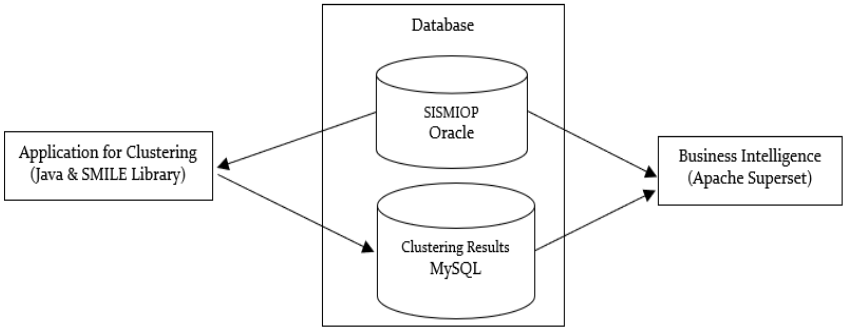

Figure 2. System Architecture

From the System Architecture, it can be traced that the system is an application built using Java. The SMILE Library will cluster data with the K-Means algorithm taken from the SISMIOP database. The clustering results will be stored in the MySQL database, and Apache Superset will subsequently perform visualizations sourced from data, either from clustering results or receivables data in the SISMIOP database.

To create an application that functions to cluster taxpayer compliance levels with the K-Means algorithm using the Java programming language and the SMILE Library. The Spring Molecular function will run this application in a Batch Process. The main program code listing is as follows:

The next step is to create a dataset on Apache Superset. After making the dataset, we can create a Chart and put it into the Dashboard on Apache Superset.

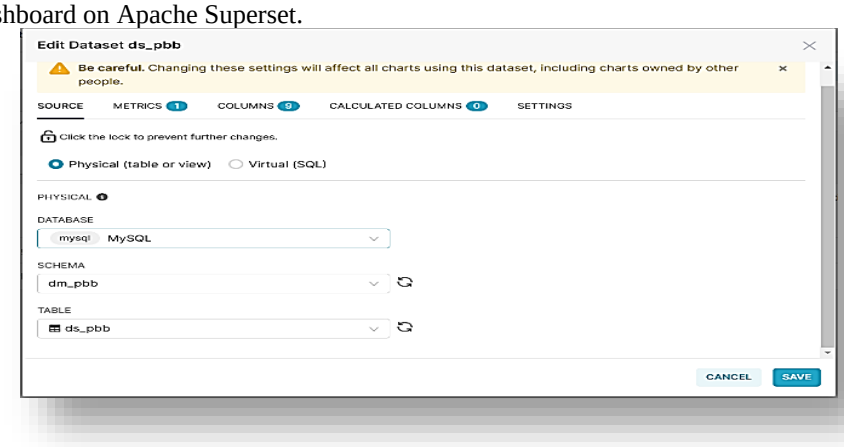

Figure 3. Apache Superset Dataset;

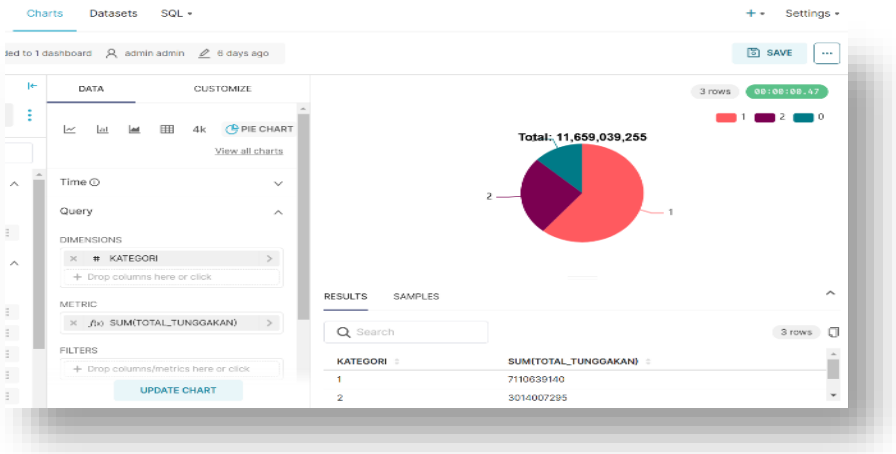

Figure 4. Apache Superset Chart

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14611
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

The first test was to run the clustering application from Java using PBB-P2 Tax data at the Tebing Tinggi City BPKPD from 2020 to 2024. The application runs well; running one clustering process with storage to the reservoir database takes approximately 80 seconds. The following is the result displayed on the execution results screen. This research successfully developed and implemented a Business Intelligence-based PBB-P2 receivables monitoring system using Apache Superset. This system visualizes accounts receivable data in real-time, interactively, and informatively, thus helping the monitoring and decision-making process related to regional receivables. The following are the main results of the research:

```
Cluster Centroids:
Cluster 0: [0.4579643676133357, 0.14478673487387547, 185.87412242017993]
Cluster 1: [4.616702355460386, 0.9219316251938272, 3875.7828398434617]
Cluster 2: [2.4164501521787543, 0.5111008833791107, 1908.483037636404]
Silhouette Score: 0.7504464864055562
```

Figure 5. Program Execution Results

The next is a visualization test by Apache Superset from the clustered data. The visualization results went well and can be seen as follows:

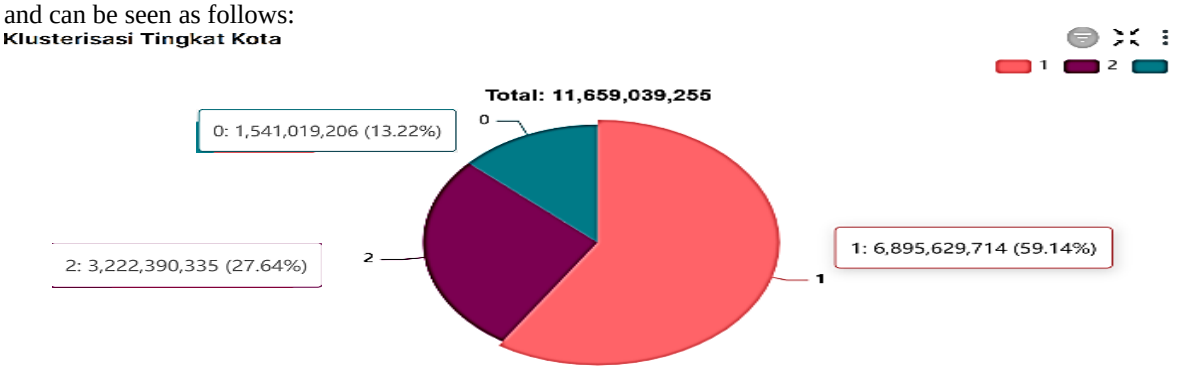

Figure 6. City-Level Clustering

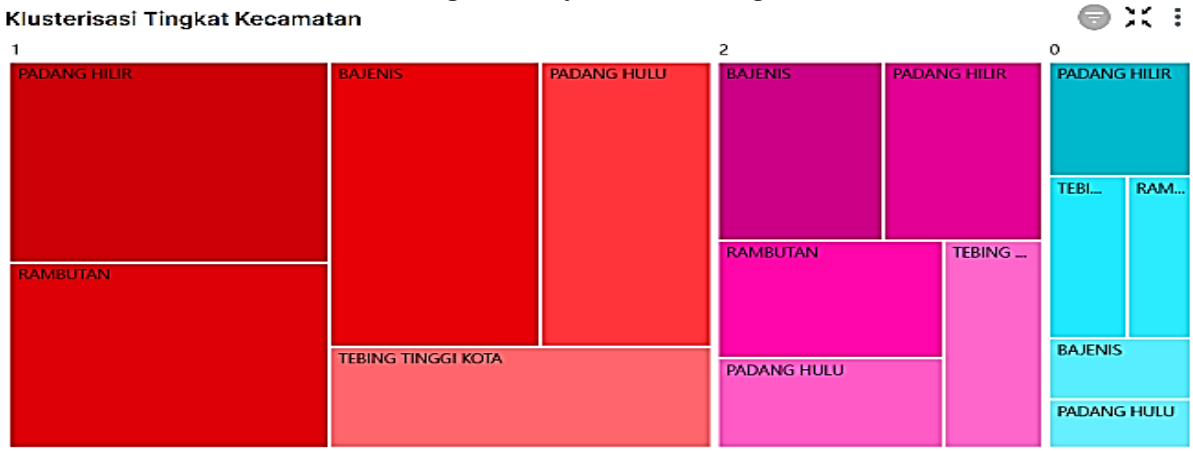

Figure 7. Sub-district Level Clustering

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 2, April 2025
DOI : https://doi.org/10.33395/sinkron.v9i2.14611
e-ISSN : 2541-2019
p-ISSN : 2541-044X

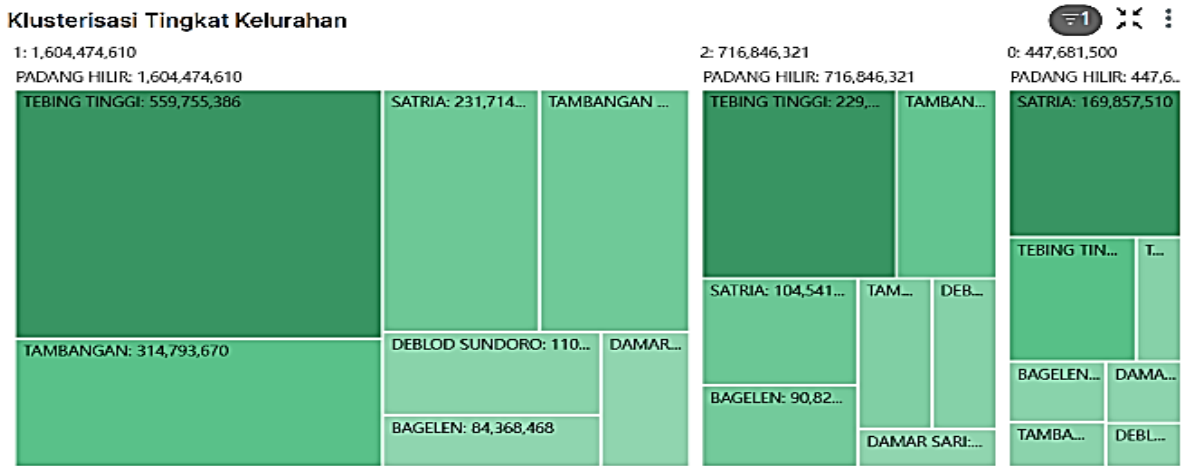

Figure 8. Village Level Clustering

This research produces an Apache Superset-based Business Intelligence (BI) system that significantly improves the ability of local government agencies to monitor PBB-P2 receivables efficiently, informatively, and based on real-time data. The main results obtained from this research are as follows:
Receivables Summary Dashboard
Displays total receivables overall and per sub-district
KPI indicators: total active receivables, collectible receivables, overdue receivables
Visual elements: Bar Chart, KPI Metric Card, Interactive Map
Purpose: Provide a quick overview for decision makers
Historical Trends Dashboard
Displays a graph of receivables trends for the last 5 years
Equipped with a Line Chart and Donut Chart per year
Findings: There is the highest spike in receivables in 2022 in 4 specific sub-districts
Purpose: Evaluate collection success from year to year
Payment Realization Dashboard
Displays comparison between target and payment realization
Using Stacked Bar Chart and Bullet Chart
Findings: 5 sub-districts have payment realization < 60% of target
Objective: Direct the focus of collection and intensification policies
Priority Taxpayer Dashboard
Table of taxpayers with the highest receivable value (Top 50)
Filter features: by region, type of tax object, payment status
Purpose: Provide a list of field collection priorities

DISCUSION

The results showed that the implementation of business intelligence through the use of Apache Superset had a significant impact on increasing the effectiveness of monitoring PBB-P2 receivables. With the interactive dashboard and easy-to-understand data visualization, the data analysis process that was previously time-consuming and done manually can now be done quickly and efficiently. This supports the theory that Business Intelligence (BI) can improve the quality of data-driven decision making, especially in the government sector, which is still dominated by conventional reports.

From the results of the execution of the clustering program, it is explained that for the compliant category in cluster 0 with centroid [0.4579643676133357, 0.14478673487387547, 185.87412242017993], while the non-compliant category is in cluster 1 with centroid [4.616702355460386, 0.9219316251938272, 3875.7828398434617] and group 2 (less compliant) at [2.4164501521787543, 0.5111008833791107, 1908.483037636404]. In addition, the results of the Silhouette Score measurement show a value of 0.7504464864055562, close to the value of 1, which means that the quality of the cluster is good. To run one clustering process takes approximately 80 seconds; if this application is implemented in a batch process, it should be with a time interval of 15 minutes.

The results of the Apache Superset visualization can describe the clustering of taxpayer compliance levels, both at the City, District, and Village levels. For the City level, it can be explained that cluster 0 (compliant) has receivables of Rp. 1,541,019,026, and cluster 1 (non-compliant) amounted to Rp. 6,895,629,714,-, while for cluster

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

2 (less compliant) amounted to Rp. 3,222,390,335,-. For the visualization of sub-district level receivables data that are not compliant in Padang Hilir District and for the sub-district level that is not compliant in Padang Hilir Sub-district is Tebing Tinggi Village.

CONCLUSION

The use of the Apache Superset dashboard is proven to provide information that is faster, more accurate, and easier to understand, both by technical staff and policy makers. This facilitates more targeted and data-driven decision-making. In addition, this system also encourages increased operational efficiency, reduces the workload of manual reporting, and increases the transparency of regional revenue management. Using Business Intelligence (BI) in tax management, especially the Rural and Urban Land and Building Tax (PBB-P2), is a strategic step in supporting the efficiency and effectiveness of data-based decision-making. However, the limitations of BI tools such as Apache Superset in handling advanced data analysis, including the application of machine learning algorithms, are challenges that need to be overcome. This study concludes that developing supporting applications using the Java programming language and the SMILE library for clustering taxpayer compliance levels can complement these weaknesses. This integration leverages an efficient batch process to group taxpayer data automatically, then saves the processing results into a MySQL database for further visualization using Apache Superset. This approach not only provides more in-depth analysis capabilities but also supports more effective data-based policy planning. With measurable implementation and cross-disciplinary collaboration, the developed solution can improve the efficiency of local tax management, reduce receivables levels, and drive significant improvements in taxpayer compliance. This research opens up opportunities for further exploration of integrating BI with machine learning technology to manage regional resources more sophisticatedly and based on modern technology.

REFERENCES

Bany Mohammed, Ashraf, Manaf Al-Okaily, Dhia Qaism, and Mohammad Khalaf Al-Majali. 2024. “Towards an Understanding of Business Intelligence and Analytics Usage: Evidence from the Banking Industry.” *International Journal of Information Management Data Insights* 4(1): 100215. doi:10.1016/j.jjimei.2024.100215.

Fatha, Mafda Khoirotul, Seftin Fitri Ana Wati, Bhagas Satrya Dewa, and Krisna Eko Prasetyo. 2023. “Peran Big Data Pada Intelijen Bisnis Sebagai Sistem Pendukung Keputusan (Systematic Literature Review).” In *Prosiding Seminar Nasional Teknologi Dan Sistem Informasi*, Surabaya, 318–26. doi:https://doi.org/10.33005/sitasi.v3i1.612.

Gallego, Victor, Jessica Lingan, Alfons Freixes, Angel A. Juan, and Celia Osorio. 2024. “Applying Machine Learning in Marketing: An Analysis Using the NMF and k-Means Algorithms.” *Information (Switzerland)* 15(7): 1–16. doi:10.3390/info15070368.

Greca, Silvana, Ingrid Shehi, and Jonuz Nuhi. 2023. “Analyzing Climate Changes Impacts Using Big Data Hadoop.” *CEUR Workshop Proceedings* 3402: 21–27.

Iqbal, Muhammad, Sardo Pardingotan Sipayung, Alex Rikki Sinaga, and Paska Marto Hasugian. 2024. “Analysis of Student Achievement with K-Means on Socioeconomic , Behavioral , and Psychological Factors.” *Jurnal Info Sains : Informatika dan Sains* 14(04): 715–28. doi:10.54209/infosains.v14i04.

Macías, José A., and Clemente R. Borges. 2024. “Monitoring and Forecasting Usability Indicators: A Business Intelligence Approach for Leveraging User-Centered Evaluation Data.” *Science of Computer Programming* 234. doi:10.1016/j.scico.2023.103077.

Novelan, Muhammad Syahputra, Syahril Efendi, Poltak Sihombing, and Herman Mawengkang. 2023. “Optimation Cavacity Vehicle Routing Problem with K-Nearest Neighbor in Classification of Goods Ditribution Route.” In *2023 International Conference of Computer Science and Information Technology (ICOSNIKOM)*, IEEE, 1–6.

Putera, Andysah, Utama Siahaan, Muhammad Irsyad, Alviona Marsya, and M Dico Triyadi. 2024. “Application Of Business Intelligence in Decision Support in Providing Assistance to Business Actors in Deli Serdang Regency Using The Decision Tree Algoritm.” 1(3): 210–14. doi:10.30596/jitcse.

Putera, Andysah, Utama Siahaan, Ami Abdul Jabar, Sugeng Pranoto, and Sulis Sutiono. 2024. “Analysis of Property Tax Bill Classification Using the C4 . 5 Algorithm.” *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)* 1(3): 181–85. doi:10.30596/jitcse.

Putra, Purwa Hasan, Zulfahmi Syahputra, and Muhammad Syahputra Novelan. 2021. “Application Of The K-Means Algorithm In Identifying.” *Jurnal Infokum* 9(2): 281–86.

Putra, Purwa Hasan, Zulfahmi Syahputra, Muhammad Syahputra Novelan, Panca Budi, Sumatera Utara, and Article Info. 2021. “Application Of The K-Means Algorithm In Identifying Types Of Skin Disease.” *Jurnal Infokum* 9(2): 281–86.

S.Pranoto and D.Nasution. 2024. “Business Intelligence Menggunakan Apache Superset Untuk Sistem Pendukung

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Keputusan Kebijakan Penagihan Pajak Bumi Dan Bangunan : Studi Kasus BPKPD Kota Tebing Tinggi." *Indonesian Journal of Education* 2(3): 154–60.

Sitorus, Zulham, Sugeng Pranoto, and Sulis Sutiono. 2024. "Comparison of Accuracy between Naïve Bayes and Decision Tree Methods for Property Tax (PBB-P2) Compliance in Tebing Tinggi City." *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)* 1(2): 121–28. doi:10.61306/jitcse.v1i2.

Superset, Apache. 2024. "Apache Software Foundation." *URL: https://superset. apache. org/[accessed 2022-03-10]*.

Tavera Romero, Carlos Andrés, Jesús Hamilton Ortiz, Osamah Ibrahim Khalaf, and Andrea Ríos Prado. 2021. "Business Intelligence: Business Evolution after Industry 4.0." *Sustainability (Switzerland)* 13(18): 1–12. doi:10.3390/su131810026.

Wahyuni, Sri. 2018. "Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree." *Journal of Physics: Conference Series* 970(1). doi:10.1088/1742-6596/970/1/012030.

Wahyuni, Sri, and Murni Marbun. 2020. "Implementation of Data Mining in Predicting the Study Period of Student Using the Naïve Bayes Algorithm." *IOP Conference Series: Materials Science and Engineering* 769(1). doi:10.1088/1757-899X/769/1/012039.

Wahyuni, Sri, Muhammad Zarlis, Solikhun, Deny Jollyta, M. Safii, and Indri Sulistianingsih. 2019. "Implementation of MD Heuristic Method for Classifying Numerical Data in Data Preprocessing." *Journal of Physics: Conference Series* 1255(1). doi:10.1088/1742-6596/1255/1/012060.

Weichbroth, Paweł, Jozef Zurada, and Celina M. Olszak. 2024. "Exploring the Benefits, Challenges, and Opportunities of Collaborative Business Intelligence." *Proceedings of the Annual Hawaii International Conference on System Sciences* 1: 278–87.

Wijerathne, Isanka. 2024. "Empowering Educational Researchers with a Privacy-Centric Data Platform : Design , Implementation , and Implications." *Conference: The 32nd International Conference on Computers in Education (ICCE 2024)* (December).

Zhao, Wan Lei, Cheng Hao Deng, and Chong Wah Ngo. 2018. "K-Means: A Revisit." *Neurocomputing* 291: 195–206. doi:10.1016/j.neucom.2018.02.072.

*name of corresponding author