TRANSMISI: JURNAL ILMIAH TEKNIK ELEKTRO, 27, (2), APRIL 2025
p-ISSN 1411-0814 e-ISSN 2407-6422

NEBULA: WEB-BASED INTERACTIVE DASHBOARD FOR MONITORING TUBERCULOSIS CASES IN SEMARANG CITY

Titis Rifamuthia[1], Ika Novita Dewi[1*)], Slamet Isworo[2], Sri Handayani[3] dan Sholikun[4]

[1]Faculty of Computer Science, Universitas Dian Nuswantoro, Indonesia
[2]Faculty of Health Science, Universitas Dian Nuswantoro, Indonesia
[3]Center of Excellence in Science and Technology, Universitas Dian Nuswantoro, Indonesia
[4]Semarang City Health Office, Indonesia

[*]*E-mail: ikadewi@dsn.dinus.ac.id*

Abstrak

Tuberkulosis (TB) masih menjadi masalah kesehatan masyarakat yang penting, terutama di negara-negara berkembang seperti Indonesia, di mana TB masih menjadi salah satu penyebab utama kematian. Di Kota Semarang, per Juli 2024, dilaporkan ada 1.624 kasus TB, yang menekankan perlunya sistem pemantauan yang kuat. Meskipun berbagai upaya telah dilakukan untuk mengurangi penularan TB, tantangan seperti sumber data yang terfragmentasi dan terbatasnya akses ke informasi real-time menghambat intervensi yang efektif. Penelitian ini menyajikan pengembangan NEBULA (Nafas Baru untuk Paru-paru), sebuah dasbor interaktif berbasis web yang dirancang untuk memvisualisasikan dan memantau penyebaran TBC di Kota Semarang. Dibangun dengan menggunakan Tableau, dasbor ini menggabungkan kemampuan penyaringan data real-time dan penelusuran data, sehingga pengguna dapat menganalisis kasus TB berdasarkan parameter seperti waktu, jenis kelamin, dan lokasi. Sistem ini bertujuan untuk membantu otoritas kesehatan setempat dan masyarakat umum dalam melacak kasus TB, mengidentifikasi area berisiko tinggi, dan memungkinkan pengambilan keputusan yang lebih tepat. Dengan menyediakan visualisasi data yang jelas dan dapat ditindaklanjuti, NEBULA meningkatkan pengawasan kesehatan masyarakat dan dapat menjadi model untuk mengelola penyakit menular di wilayah lain. Pengembangan di masa depan termasuk menggabungkan analisis prediktif untuk lebih meningkatkan manajemen TB dan strategi pencegahan.

Kata kunci: pemantauan tuberculosis, dasbor interaktif, visualisasi data, informatika kesehatan masyarakat, aplikasi berbasis web

Abstract

Tuberculosis (TB) remains a critical public health concern, particularly in developing nations like Indonesia, where it remains one of the leading causes of mortality. In Semarang City, as of July 2024, 1,624 TB cases were reported, emphasizing the need for robust monitoring systems. While various efforts have been made to reduce TB transmission, challenges such as fragmented data sources and limited access to real-time information hinder effective intervention. This study presents the development of NEBULA (New Breath for Lungs), a web-based interactive dashboard designed to visualize and monitor the spread of TB in Semarang City. Built using Tableau, the dashboard incorporates real-time data filtering and drill-down capabilities, allowing users to analyse TB cases based on parameters such as time, gender, and location. The system aims to assist local health authorities and the general public in tracking TB cases, identifying high-risk areas, and enabling more informed decision-making. By providing clear and actionable data visualization, NEBULA enhances public health surveillance and can serve as a model for managing infectious diseases in other regions. Future enhancements include incorporating predictive analytics to further improve TB management and prevention strategies.

Keywords: tuberculosis monitoring, interactive dashboard, data visualization, public health informatics, web-based application

1. Introduction

Tuberculosis (TB) is a significant public health threat, particularly in developing countries like Indonesia. Despite national efforts, such as the National Tuberculosis Control Strategy [1], TB remains one of the leading causes of death in the country. In Semarang City, the fight against TB is ongoing, with hundreds of new cases reported annually by local health authorities [2]. However, the current TB management system faces critical challenges that hinder effective intervention, particularly the lack of real-time, integrated data monitoring [3]. Existing systems often rely on fragmented data sources and manual processes, making it difficult for health officials to track the spread of TB efficiently or respond promptly to emerging hotspots [4]. In addition to government initiatives, community participation has played a vital role in TB prevention, with local volunteers, or cadres, actively involved in educating the public, identifying suspected cases, and assisting diagnosed patients [5]. Given the active involvement of the community, there is a pressing need for an efficient system that can monitor and manage TB spread data in real-time [4]. Previous studies on TB data distribution have explored various factors affecting its spread. However, spatial analysis of TB, such as examining population density and poverty levels, has yielded mixed results, showing no clear correlation between these factors and TB prevalence [6]. Current approaches to TB monitoring in Semarang are limited by disparate data collection methods, inconsistent data formats, and delays in data reporting, which prevent timely decision-making.

Data is often siloed between different institutions, such as the Health Office, meteorological agencies, and population registries, without a unified system to consolidate and visualize the information. This fragmentation leads to an incomplete understanding of the disease's spread, making it difficult to identify high-risk areas, predict future outbreaks, or evaluate the effectiveness of interventions [7]. Furthermore, these studies faced limitations such as incomplete geographic data, low-resolution mapping, data collection biases, and privacy concerns. To address these challenges and improve the monitoring and management of TB data, data visualization and interactive dashboards [8] offer a promising solution. Data visualization simplifies complex information, making it more accessible for decision-makers, who can then respond swiftly and effectively [9][10]. Interactive dashboards have proven useful in other areas of public health, including monitoring child growth [11], health profiling [12], and preventing strokes [13]. An interactive dashboard allows users to dynamically explore the data, identifying patterns, trends, and correlations [14] in the distribution of TB in Semarang City. However, monitoring TB cases in Semarang remains difficult due to fragmented data sources and limited access to real-time information.

To address these challenges, this study presents the development of NEBULA (New Breath for Lungs), a web-based interactive dashboard designed to provide real-time, data-driven insights into TB cases in Semarang City. Built using Tableau [15], NEBULA integrates data from multiple sources, including the Semarang City Health Office, the Meteorology, Climatology, and Geophysics Agency, and the Central Statistics Agency, to create a unified platform for tracking and visualizing TB data. Tableau was selected for its intuitive user interface, its ability to seamlessly connect to diverse data sources, and its powerful interactive visualization capabilities [16].

The dashboard offers filtering and drill-down capabilities, allowing users to explore TB cases by time period, gender, and geographic location, down to the village and sub-district level. By centralizing and visualizing data, NEBULA enables health officials and policymakers to make informed, timely decisions, identify high-risk areas, and plan more effective interventions. NEBULA represents a critical advancement in the use of data visualization to enhance public health surveillance and TB management, helping Semarang move closer to the goal of becoming TB-free. NEBULA is available online at https://smg.city/nebula.

2. Research Methodology

The methodology for developing the web-based interactive dashboard for monitoring TB cases in Semarang City shown in Figure 1 consist of the following key steps: data collection, data preprocessing, dimensionality reduction, outlier analysis, and interactive dashboard development.

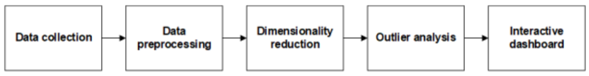

Figure 1. Research Flow

2.1. Data Collection

The data for monitoring TB cases in Semarang City was collected from multiple sources to ensure comprehensive and reliable inputs for analysis. This data collection process provides a rich, multi-dimensional dataset to analyse and visualize TB spread in Semarang City, considering both health-related factors and environmental and demographic influences. The sources and methods of data collection include Survey on Tuberculosis Vulnerability Mapping on 2023; Semarang City Health Office; Meteorology, Climatology, and Geophysics Agency; Central Statistics Agency; and Population and Civil Registration Agency.

Survey on Tuberculosis Vulnerability Mapping on 2023

Table 1. Sample of survey on Tuberculosis Vulnerability Mapping on 2023

Kelurahan	Pendidikan	Pekerjaan
Bambankerep	Tamat SMA	Karyawan
Bandarharjo	Tamat SMP	Pedagang
Banyumanik	Diploma	Tidak Diketahui
Barusari	Tamat SD	Tidak Diketahui
Bendungan	Tamat SMP	Buruh

A survey conducted by the Center of Excellence, Universitas Dian Nuswantoro Semarang, focused on identifying areas of vulnerability to TB infection within Semarang City. This survey provided geospatial data related to TB susceptibility across various districts and communities, helping to map out high-risk areas for targeted interventions. Some attributes used in this survey are *kecamatan, kelurahan, pendidikan, pekerjaan,* *pendapatan, perilaku, kondisi_tempat_tinggal, stigma, pengetahuan*, and *literasi_tb*. The result of Survey on Tuberculosis Vulnerability Mapping on 2023 shown in table 1.

Semarang City Health Office

Official TB case reports, demographic data, and time-series data related to TB occurrences were sourced from the Semarang City Health Office. This dataset includes historical and current TB case counts, allowing for tracking trends over time. Data retrieved from *Semarang City Health Office* consist of *pilar 4 stbm (tpsrt), pilar 2 stbm ctps, pilar 3 stbm pammrt, phbs 1 gizi seimbang, phbs aktivitas fisik, phbs cuci tangan, phbs tidak miras/narkoba, phbs air bersih, phbs jamban sehat, phbs sampah, phbs lantai kedap air, phbs tidak merokok, phbs asi eksklusif, phbs pemeriksaan kehamilan, kapasitas layanan puskesmas, vaksinasi bcg, pilar 1 stbm aman, pilar 1 stbm layak, pilar 1 stbm sharing, pilar 1 stbm belum layak, jumlah poliklinik, jumlah praktek dokter, pilar 3 tpammrt,* and *pilar 3 tplcrt*. The sample data of TB case counts from the Semarang City Health Office shown in table 2.

Table 2. TB case counts from the Semarang City Health Office

Kelurahan	Pilar 4 stbm	Pilar 2 stbm	Phbs Air Bersih	Phbs Jamban
Bambankerep	125	800	1364	1364
Bandarharjo	220	3500	4230	4230
Banyumanik	50	2870	4746	4744
Barusari	250	2250	2890	3932
Bendungan	238	395	3370	309

Meteorology, Climatology, and Geophysics Agency

Environmental data such as temperature, humidity, and air quality were obtained from the Meteorology and Climatology Agency. These variables are important in understanding how environmental factors might correlate with the spread of TB in specific areas. The environmental data used in this research are *Temperatur minimum, Temperatur maksimum, Temperatur rata-rata, Kelembapan rata-rata, Curah hujan, Lamanya penyinaran matahari, Kecepatan angin maksimum, Arah angin saat kecepatan maksimum, Kecepatan angin rata-rata, ID WMO, Nama Stasiun, Lintang, Bujur*, and Elevasi. The sample of environmental data shown in table 3.

Table 3. Environmental data

Bulan	Temperature minimum	Temperature maksimum	Temperature rata-rata
Januari	24,51612903	30,987	27,348
Februari	24,5	30,757	26,839
Maret	24,974	31,722	27,883
April	25,3	32,68	28,53
Mei	24,94	33,69	29,254

Central Statistics Agency

Socio-economic data, such as population density, poverty levels, and living conditions, was collected from the Central Statistics Agency that generated from semarangkota.bps.go.id/indicator/151/79.This information helps contextualize TB spread in relation to demographic and socio-economic factors Socio-economic data used are *curah hujan, kecepatan angin, kelembapan,* and *suhu udara.* The sample of socio-economic data shown in table 4.

Table 4. Socio-economic data

Bulan	Curah hujan	Kecepatan angin	kelembapan	Suhu
Januari	329.0	3.52	55.10	305.46
Februari	337.0	4.07	55.11	305.55
Maret	165.0	4.08	55.87	304.15
April	134.0	4.06	55.12	305.62
Mei	191.0	2.79	54.94	305.42

Population and Civil Registration Agency

Data on population demographics, including age distribution, gender, and population movement trends, was sourced from the Population and Civil Registration Agency. This data aids in refining the granularity of analysis by linking TB cases to specific population segments. Demographic data used in this research are *kecamatan, jumlah LK, jumlah PR*, and *total penduduk.* The sample of demographic data shown in table 5.

Table 5. Demographics data

Kecamatan	Jumlah LK	Jumlah PR	Total penduduk
Semarang Tengah	216	185	401
Semarang Utara	608	534	1.142
Semarang Timur	319	267	586
Gayamsari	384	359	743
Genuk	791	776	1.567

2.2. Data Preprocessing

After data collection, the next critical step is preprocessing to ensure the data is clean, consistent, and ready for analysis. The preprocessing workflow shown in figure 2 includes the following procedures: data categorization, handling missing value, data cleaning, data synthesizer, and one hot encoding.

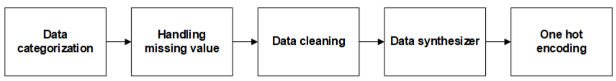

Figure 2. Data Preprocessing

Data Categorization

The raw data collected from various sources is categorized based on relevant features, such as demographic factors (age, gender, population density), environmental factors (temperature, humidity), and TB case statistics. This categorization helps streamline the analysis process by organizing the data into meaningful groups.

Handling Missing Values

Since data from multiple sources can sometimes have missing or incomplete entries, strategies for handling missing data are employed. Depending on the dataset, missing values may be addressed through imputation techniques (e.g., filling missing values using the mean, median, or interpolation) or by removing incomplete records if they are not critical to the analysis.

Data Cleaning

This step involves identifying and correcting any inconsistencies or inaccuracies in the data. This includes checking for duplicate entries, resolving inconsistencies in data formats, and ensuring all fields are correctly aligned with the intended categories. The goal is to eliminate any potential sources of error or bias that may skew the analysis.

Data Synthesis

For datasets with limited size or coverage, data synthesis is performed to generate additional records based on the patterns observed in the existing data. This synthetic data generation helps to improve the robustness of the analysis, especially in cases where the sample size is small or underrepresented. Techniques such as bootstrapping or random sampling with replication may be applied.

One-Hot Encoding

For categorical variables, one-hot encoding is applied to convert non-numeric data (e.g., gender, district names) into a binary format, which is more suitable for computational analysis. Each categorical feature is transformed into a set of binary variables (0 or 1), making the data compatible with various machine learning models and analysis techniques.

2.3. Dimensionality Reduction

Dimensionality reduction is applied by using Principal Component Analysis (PCA). PCA is used to identify patterns in data and express the data in terms of uncorrelated principal components [17]. By organizing information into these principal components, the dimensionality of the dataset is reduced without losing much of the critical information. In the context of TB monitoring, PCA helps reduce the complexity of the dataset by focusing on the most significant variables that contribute to the spread and distribution of TB cases, such as environmental factors (temperature, humidity), socio-economic data (poverty levels, population density), and demographic information. By reducing the number of features [18], the dashboard becomes more responsive and focused, allowing users to explore the key drivers of TB spread with fewer variables, which in turn improves the interpretability of the visualizations.

2.4. Outlier Analysis

After dimensionality reduction, the next step is to apply K-means clustering for outlier detection. K-means clustering is used to group data points into distinct clusters based on their similarity [19]. By clustering the data, normal patterns and relationships between TB cases can be identified. Outlier detection is crucial for identifying abnormal patterns in the TB dataset, such as sudden spikes in cases, incorrect data entries, or areas where the reported TB cases significantly deviate from expected trends.

Several factors influence the performance of K-means, including the choice of k (the number of clusters), the initialization of centroids, and the presence of non-spherical or overlapping clusters [20]. To mitigate the impact of these factors, we employed the elbow method to determine the optimal k and used multiple random initializations to avoid local optima [21]. Despite its limitations, K-means remains a widely used and effective algorithm for outlier detection in various domains due to its simplicity and computational efficiency.

2.5. Interactive Dashboard Development

The interactive dashboard visualization process is carried out in two stages, namely data visualization using Tableau and creating an interactive website display. This process ensures that the data is not only visualized effectively but is also made accessible through a user-friendly web platform that allows for real-time interaction and monitoring.

Data Visualization using Tableau

Tableau is utilized to transform the processed data into meaningful visualizations. The dashboard also provides filter and drill-down features that allow users to view data based on various parameters, such as subdistrict and village areas, gender, or specific time periods. This stage focuses on making the data easily interpretable for decision-makers and health officials, who can then make informed interventions.

Creating the Interactive Website Display

Once the data visualizations are built in Tableau, the next stage involves embedding these visualizations into a web-based interface. The interactive dashboard is hosted on a website called NEBULA, which is accessible to the general public and health authorities. The website integrates

Tableau's visualizations and adds interactive features that allow users to engage with the data dynamically. The goal of the website is to provide an intuitive, easy-to-use platform for tracking TB trends, identifying high-risk areas, and planning interventions.

3. Result and Discussion

This section presents the outcomes of the dashboard development and key insights from the data.

3.1. Result of data source integration

Data from various sources were integrated to create a comprehensive dataset for monitoring and visualizing the spread of TB in Semarang City. To integrate data from multiple sources (Survey on Tuberculosis Vulnerability Mapping, Semarang City Health Office, Meteorology, Climatology, and Geophysics Agency, Central Statistics Agency, and Population and Civil Registration Agency), the process involves several steps to ensure that diverse datasets can be aligned, merged, and standardized for analysis. The new dataset consists of 134 attributes and has 982 records.

3.2. Result of data preprocessing

The preprocessing stage carried out in the research is data categorization, Handling Missing Value, Data Cleaning, Data Synthesizer, and One Hot Encoding. The data categorization stage is carried out to group variables into categories based on the attributes contained in the dataset, at this stage categorizing attributes based on categorical variables and numerical variables from the dataset. This result shown in table 6.

Table 6. Data categorization

Data type	Attributes
Categorical	Kecamatan, kelurahan, Pendidikan, pekerjaan, pendapatan, perilaku, kondisi_tempat_tinggal, stigma, pengetahuan, literasi_tb, uuid
Numerical	luas wilayah per kelurahan, jumlah penduduk per kelurahan, kepadatan penduduk, pilar 4 stbm (tpsrt), pilar 2 stbm ctps, pilar 3 stbm pammrt, phbs 1 gizi seimbang, phbs aktivitas fisik, phbs cuci tangan, phbs tidak miras/narkoba, phbs air bersih, phbs jamban sehat, phbs sampah, phbs lantai kedap air, phbs tidak merokok, phbs asi eksklusif, phbs pemeriksaan kehamilan, kapasitas layanan puskesmas, vaksinasi bcg, pilar 1 stbm aman, pilar 1 stbm layak, pilar 1 stbm sharing, pilar 1 stbm belum layak, jumlah poliklinik, jumlah praktek dokter, pilar 3 tpammrt, pilar 3 tplcrt, temperature, humidity, wind

Identifying the missing value check in the TB vulnerability distribution dataset, it is found that 10 data that are missing data and 1 unknown data. These missing data is handled by filling in the missing values with the value "Unknown". The next step is to perform a data synthesizer as a data privacy protector and add the column of Universally Unique Identifier (UUID) to the Data Frame with a unique UUID value for each column. Sample of the UUID shown in Table 2.

Table 7. Data privacy protector

Kelurahan	UUID
Bambankerep	d99ddc02-6d73-42b7-a804-35433b53e2d0
Bandaharjo	b99836cf-5ea3-43ae-bf8e-0306cd892b59
Banyumanik	3869e1c3-fe87-4109-ad2e-90dcbcdb5eed

One Hot Encoding is done by applying Pandas Library to convert categorical data into numeric data format. For example, we have a feature of *Pendapatan* with categorical values. *Pendapatan* has two distinct categories: *Rendah* and *Menengah*. For One-Hot Encoding, we create new columns for each category. Each column represents one category, and we mark a 1 if that row belongs to that category, and 0 otherwise.

3.3. Result of dimensionality reduction

the initial variables consisted of the 134 attributes described in Section 3.1. PCA was employed to transform these initial variables into a new set of uncorrelated variables called principal components. These components are ordered by the amount of variance they explain, with the first component capturing the most variance.

Dimensionality reduction performed by PCA algorithm resulted in the highest variance of 0.05. This means that each of the selected principal components captures 5% of the total variability in the data as shown in figure 3. From this result, it can be concluded that there are 20 important variable components that will be applied to the K-Means algorithm.

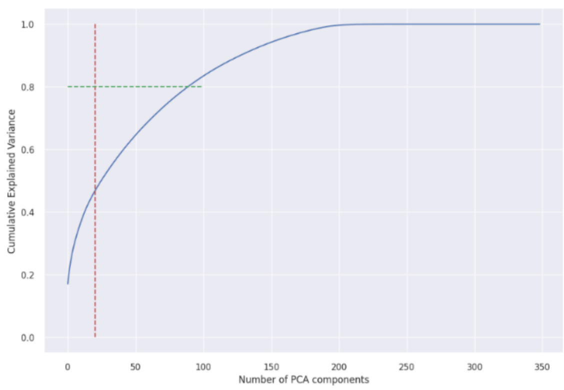

Figure 3. Cumulative explained variance curve

Figure 3 shows the cumulative explained variance as a function of the number of PCA components, with the goal of retaining a sufficient portion of the dataset's variance while reducing its dimensionality. The red dashed line marks the optimal number of components, based on a green horizontal line indicating a 90% variance threshold, ensuring minimal information loss. The analysis reveals that the highest variance captured by any single component is 0.05, with the cumulative variance rising steeply at first before plateauing, indicating that a reduced number of components can summarize the dataset effectively. This dimensionality reduction improves the efficiency and clarity of the subsequent data analysis and visualization processes in the interactive dashboard.

3.4. Result of outlier analysis

The results of applying the K-Means clustering algorithm to detect outliers in the tuberculosis dataset shown in figure 4. The elbow method is used to determine the optimal number of clusters, with the elbow point identified at k = 15, where the distortion score is approximately 0.322. The distortion score, which measures the sum of squared distances between data points and their cluster centroids, decreases as the number of clusters increases, but the rate of improvement diminishes after k = 15, indicating that this is the ideal number of clusters. The plot also shows the computation time for each k value, highlighting the balance between efficient clustering and computational performance. Using k = 15 allows for effective outlier detection, as data points far from cluster centroids can be flagged as potential anomalies, which is important for identifying unusual cases in the dataset.

3.5. Dashboard visualization by Tableau

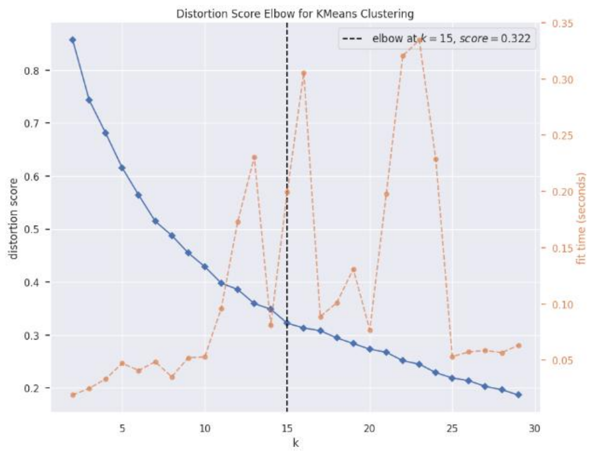

Figure 4. Optimal cluster selection

Figure 3 shows the cumulative explained variance as a function of the number of PCA. The use of interactive visualizations is critical for conveying complex insights effectively, and Tableau offers an ideal platform for creating such visualizations due to its versatility and user-friendly interface. The pre-processed dataset, which has undergone various stages of processing such as categorization, missing value handling, data cleaning, dimensionality reduction using PCA, and outlier detection with K-means clustering, was prepared in a format compatible with Tableau. This dataset, structured in a tabular form, contains essential information such as demographic variables (e.g., age, gender), temporal variables (e.g., month and year), and geographical identifiers (e.g., sub-districts, villages) linked to the spread of TB.

The dataset was imported into Tableau by connecting directly to an Excel file containing the cleaned and structured data. Tableau's data preview feature was used to ensure that all fields, such as dates, geographical locations, and case counts, were correctly interpreted. Basic exploratory data analysis was conducted within Tableau to confirm the integrity of the dataset and to ensure that key metrics, such as the number of TB cases over time and across regions, were accurately represented.

Tableau's powerful filtering capabilities were employed to create an interactive dashboard that allows users to explore the data dynamically. Filters for key variables, such as time periods, gender, and geographical regions, were added to the dashboard. This interactivity allows users to drill down into specific sub-districts or explore TB trends over custom time intervals. By enabling users to interact with the data, the dashboard becomes more than a static visualization; it becomes a tool for decision-making and in-depth analysis.

After individual visualizations were created, they were compiled into a comprehensive dashboard. This dashboard was carefully designed to ensure that the most important information was readily accessible while also providing tools for in-depth exploration. Multiple visualizations, such as maps, line graphs, and bar charts, were arranged in a cohesive layout, enabling cross-filtering between charts. For example, selecting a specific sub-district on the map would automatically update the line graph to show the TB trend for that particular region.

3.6. Web based interactive dashboard

Figure 5. Landing page of the website

Once the dashboard was finalized, it was published to Tableau Public, making it accessible via a web-based platform. A URL link to the dashboard was generated, allowing health officials, researchers, and the general public to access the tool. The web-based dashboard can be embedded in websites and shared with stakeholders, ensuring that the valuable insights gleaned from the data are available to those responsible for managing and mitigating TB in Semarang City. The website called

Nebula, can be accessed from https://smg.city/nebula. The landing page of the website shown in figure 5.

The dashboard includes a slicer feature, such as shown in Figure 6, which enhances its interactivity by allowing users to filter data dynamically. This feature enables users to easily select specific areas or criteria from one or more columns within the dataset, such as region, time period, or demographic variables. As a result, the data visualization is automatically adjusted based on the user's selections, offering a tailored view of the information. This functionality improves the usability of the dashboard, providing decision-makers with a focused and detailed understanding of the tuberculosis spread in different regions or population groups.

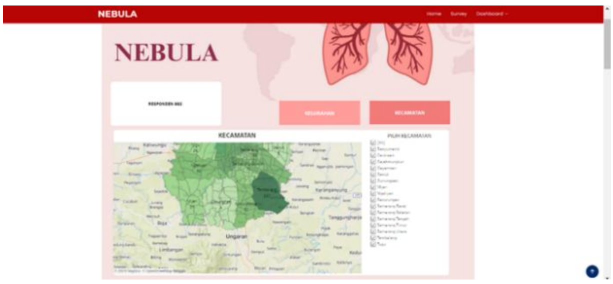

Figure 6. Slicer feature to filter data dynamically

The number of TB vulnerability indices in the sub-districts of Semarang City, such shown in figure 7, refers to the assessment of different factors that contribute to the risk of tuberculosis (TB) within each area. These indices may include various data points such as population density, poverty rates, access to healthcare services, environmental factors, and historical TB case data. The vulnerability index is crucial for identifying which sub-districts are at higher risk of TB outbreaks, allowing for targeted interventions and resource allocation.

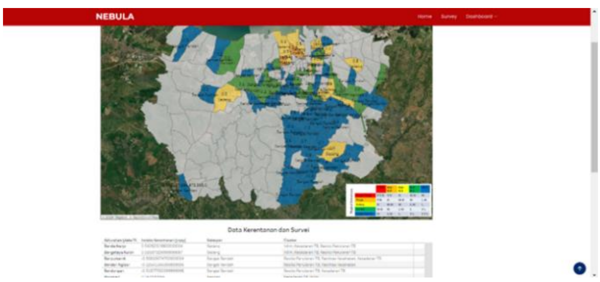

Figure 7. The number of TB vulnerability indices in Semarang City sub-districts

By using the integrated data sources, such as data from the Semarang City Health Office, Meteorology and Climatology Agency, and other local agencies, a comprehensive index can be calculated for each sub-district. This information is visualized in the dashboard, enabling health officials and policymakers to quickly identify areas with the highest vulnerability and tailor TB prevention efforts accordingly. The dashboard's slicer feature can also filter the vulnerability indices for specific sub-districts, allowing users to compare risk levels across different regions within the city.

3.7. Limitation of the study

The integration of pre-processed tuberculosis (TB) data into an interactive Tableau dashboard marks a significant step forward in enhancing the monitoring and management of TB cases in Semarang City. The dashboard's ability to provide real-time insights, offer drill-down filtering capabilities, and visually display complex data empowers health authorities and community leaders to make more informed decisions. By enabling stakeholders to analyze trends, identify high-risk areas, and evaluate the effectiveness of interventions, the dashboard plays a pivotal role in the ongoing efforts to reduce TB incidence in the region. Moreover, its user-friendly interface ensures that a wide range of users, including public health officials, researchers, and community members, can engage with the data to support TB prevention strategies.

However, this study has several limitations that should be acknowledged. First, the data used in this study was limited to data from 2023 and early 2024 from Semarang City only. This limited the scope of the analysis and may not be representative of longer-term trends or other geographic regions. Second, the PCA analysis was performed using a specific set of variables, and different variable selections may yield different results. Third, the usability of the dashboard was not formally evaluated with end-users. While the dashboard was designed with usability in mind, further testing with healthcare professionals and community members is necessary to ensure its effectiveness and identify areas for improvement. Finally, the predictive capabilities of NEBULA are currently limited. The system primarily provides descriptive insights, and future work is needed to incorporate predictive modeling to forecast TB outbreaks and optimize resource allocation. We also did not provide rationale on negative and positive impact.

4. Conclusion

The integration of pre-processed tuberculosis (TB) data into an interactive Tableau dashboard marks a significant step forward in enhancing the monitoring and management of TB cases in Semarang City. The dashboard's ability to provide real-time insights, offer drill-down filtering capabilities, and visually display complex data empowers health authorities and community leaders to make more informed decisions. By enabling stakeholders to analyze trends, identify high-risk areas, and evaluate the effectiveness of interventions, the dashboard plays a pivotal role in the ongoing efforts to reduce TB incidence in the region. Moreover, its user-friendly interface ensures that a wide range of users, including public health officials, researchers, and community members, can engage with the data to support TB prevention strategies.

Looking forward, several opportunities for further development remain. Future work includes the integration of predictive analytics, which would allow the system to project future trends and optimize resource allocation. The potential to expand the dashboard's scope to other infectious diseases and adapt it for mobile platforms would further enhance its utility and accessibility. Ensuring robust data security and privacy protections will also be a critical focus moving forward. With these advancements, the interactive dashboard can serve as a model for public health initiatives, leveraging real-time data visualization and analytics to improve decision-making and public health outcomes on a broader scale.

5. Acknowledgments

We would like to express our sincere gratitude to the Center of Excellence, Universitas Dian Nuswantoro (Udinus), for their invaluable support and collaboration throughout this research. Their expertise and resources, particularly in conducting the Survey on Tuberculosis Vulnerability Mapping in Semarang City, have played a crucial role in the success of this project.

References

[1]. The Ministry of Health Republic of Indonesia, "Revised National Strategy of Tuberculosis Care and Prevention in Indonesia 2020-2024 and Interim Plan for 2025-2026," 2023.

[2]. Kementerian Kesehatan RI, "Laporan Kinerja Direktorat Pencegahan Dan Pengendalian Penyakit Menular Tahun 2022," *Kementeri. Kesehat. Republik Indones.*, pp. 1–119, 2022.

[3]. Francisca Chibugo Udegbe, Ejike Innocent Nwankwo, Geneva Tamunobarafiri Igwama, and Janet Aderonke Olaboye, "Real-Time data integration in diagnostic devices for predictive modeling of infectious disease outbreaks," *Comput. Sci. IT Res. J.*, vol. 4, no. 3, pp. 525–545, 2023, doi: 10.51594/csitrj.v4i3.1502.

[4]. R. Saadatzadeh, "The Future of Public Health : Integrating Artificial Intelligence in Disease Surveillance and Prevention," *Foresight Public Heal.*, vol. 1, no. 2, pp. 1–15, 2024.

[5]. S. Ulfah, S. K. Nasution, and F. A. Siregar, "Performance of Tuberculosis Program Officers in the Implementation of Contact Investigations in Medan City," *J. Soc. Res.*, vol. 2, no. 4, pp. 1232–1245, 2023, doi: 10.55324/josr.v2i4.792.

[6]. H. Helmy, M. T. Kamaluddin, I. Iskandar, and Suheryanto, "Investigating Spatial Patterns of Pulmonary Tuberculosis and Main Related Factors in Bandar Lampung, Indonesia Using Geographically Weighted Poisson Regression," *Trop. Med. Infect. Dis.*, vol. 7, no. 9, 2022, doi: 10.3390/tropicalmed7090212.

[7]. Ibrahim Adedeji Adeniran, Christianah Pelumi Efunniyi, Olajide Soji Osundare, and Angela Omozele Abhulimen, "Data-driven decision-making in healthcare: Improving patient outcomes through predictive modeling," *Int. J. Sch. Res. Multidiscip. Stud.*, vol. 5, no. 1, pp. 059–067, 2024, doi: 10.56781/ijsrms.2024.5.1.0040.

[8]. R. Rabiei and S. Almasi, "Requirements and challenges of hospital dashboards: a systematic literature review," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, pp. 1–11, 2022, doi: 10.1186/s12911-022-02037-8.

[9]. H. Shao, R. Martinez-Maldonado, V. Echeverria, L. Yan, and D. Gašević, "Data Storytelling in Data Visualisation: Does it Enhance the Eficiency and Efectiveness of Information Retrieval and Insights Comprehension?," *Conf. Hum. Factors Comput. Syst. - Proc.*, 2024, doi: 10.1145/3613904.3643022.

[10]. Q. Hasanah, I. Y. R. Pratiwi, and Z. Amalia, "Analysis distribution and segmentation of micro, small, and medium enterprises (MSMEs) in Kediri Residency Area: implications for local economic development," *MATRIX J. Manaj. Teknol. dan Inform.*, vol. 13, no. 3, pp. 130–141, 2023, doi: 10.31940/matrix.v13i3.130-141.

[11]. F. Yusuf Alfiansyah and D. Arisandi, "Perancangan Dashboard Monitoring Status Gizi Balita di Puskesmas Sukanagalih," *J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 23, no. 3, pp. 1579–1589, 2023.

[12]. J. B. Angela and A. Irsyad, "Implementasi Visualisasi Data Berbasis Web Pada Exploratory Data Analysis Profil Kesehatan Kota Samarinda," *Kreat. Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 9–16, 2023.

[13]. I. N. Rizki, M. L. Puspita, D. Prayoga, M. Q. Huda, and T. Selatan, "Implementasi Exploratory Data Analysis untuk Analisis dan Visualisasi Data Penderita Stroke Kalimantan Selatan Menggunakan Platform Tableau," *J. Inform. dan Tek. elektro Terap.*, vol. 12, no. 1, 2024.

[14]. M. Zhuang, D. Concannon, and E. Manley, "A Framework for Evaluating Dashboards in Healthcare," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 4, pp. 1715–1731, 2022, doi: 10.1109/TVCG.2022.3147154.

[15]. L. Beard and N. Aghassibakes, "Resource review," *Journal of the Medical Library Association*, vol. 109, no. 1. pp. 159–161, 2021, doi: 10.5195/jmla.2021.1135.

[16]. A. Mishra, "The Role of Data Visualization Tools in Real- Time Reporting : Comparing Tableau , Power BI , and Qlik Sense," *Int. J. Sci. Technol.*, vol. 11, no. 3, pp. 1–8, 2020.

[17]. B. M. S. Hasan and A. M. Abdulazeez, "A Review of Principal Component Analysis Algorithm for Dimensionality Reduction," *J. Soft Comput. Data Min.*, vol. 2, no. 1, pp. 20–30, 2021, doi: 10.30880/jscdm.2021.02.01.003.

[18]. H. W. Gichuhi, M. Magumba, M. Kumar, and R. W. Mayega, "A machine learning approach to explore individual risk factors for tuberculosis treatment non-adherence in Mukono district," *PLOS Glob. Public Heal.*, vol. 3, no. 7, pp. 1–20, 2023, doi: 10.1371/journal.pgph.0001466.

[19]. E. U. Oti, M. O. Olusola, F. C. Eze, and S. U. Enogwe, "Comprehensive Review of K-Means Clustering Algorithms," *Int. J. Adv. Sci. Res. Eng.*, vol. 07, no. 08, pp. 64–69, 2021, doi: 10.31695/ijasre.2021.34050.

[20]. K. P. Sinaga, I. Hussain, and M. S. Yang, "Entropy K-Means Clustering with Feature Reduction under Unknown Number of Clusters," *IEEE Access*, vol. 9, pp. 67736–67751, 2021, doi: 10.1109/ACCESS.2021.3077622.

[21]. A. Larasati, R. Maren, and R. Wulandari, "Utilizing Elbow Method for Text Clustering Optimization in Analyzing Social Media Marketing Content of Indonesian e-Commerce," *J. Tek. Ind.*, vol. 23, no. 2, pp. 111–120, 2021, doi: 10.9744/jti.23.2.111-120.