

Big Data Analysis of Covid-19 Spread Based on Distribution Map and Protocol Regulations with Business Intelligence

Anita Sindar Sinaga^{*1}, Arjon Samuel Sitio²

^{1,2}STMIK Pelita Nusantara

E-mail: ^{*1}haito_ita@yahoo.com, ²arjonsitio@yahoo.com

Abstract

Various parties analyze the total cases of the Corona virus every day to prevent a high level of spread. Big Data technology involves data that is diverse, rapidly changing, super large in size so that conventional technology cannot be handled effectively. Four stages of processing Big Data until the output it produces is information for users, namely acquired, accessed, analytics, and application. Business Intelligence serves to transform operational, transactional data into useful information within the company. Big Data Analytics analyzes information, IDE for current and past business decisions visualized in graphical form. The problem in this study is modeled in the Big Data architecture with the aim of comparing the trend of the spread of Covid-19 related to process regulation. Based on the distribution map data and the regulation of health protocol rules with various online data sources, structured and unstructured using Business Intelligence (BI) cleaning data, three patterns of spread of the corona virus were found, High Cases, Medium Cases and Low Cases from 34 provinces throughout Indonesia. West Java Province became the first order of the highest cases during May 2021 with an average case of 21% followed by DKI Jakarta 14%.

Keywords — Covid-19 Cases, Business Intelligence, Trend Data, Big Data Analytic

1. INTRODUCTION

The use of machine learning technology as part of Artificial Intelligence (AI) or artificial intelligence and other sciences is applied to study the figures for the spread of Covid-19. Business intelligence is a concept and method to improve the quality of business decision making based on data-based systems. A collection of raw data can be converted into information by being analyzed and structured based on the relationship between the data by knowing what data to collect and in what context it is desired [1]. Business intelligence (BI) is tasked with collecting, storing, and analyzing data from business activities to assist decision making. Traditional BI systems are efficient in extracting and analyzing data, but are rigid, slow, time consuming, and require expert knowledge for maintenance so researchers add modern features to shape the next generation of BI [2]. Big Data is a challenging concept in technology. Big Data refers to different quantities and volumes of data created by individuals, tools and machines, requiring large-scale innovation and technology to collect, provide and process analysis for large and large data and then stored in a single database to gain real insight [3]. Time business related to consumers. Big Data is used for processing data that exceeds the processing capacity of conventional databases, is too large, moves too fast, and does not match the structural capabilities of traditional database architectures, while Big Data Analytics is a process with an integrated system capable of handling big data [4]. The need for

software that accommodates various types of data can cause the data handling process to be slow and less effective. Four characteristics of Big Data, four Vs, data processing based on Volume, Velocity, Variety, Value [5]. Previous research on the application of big data as an effort to mitigate the covid-19 pandemic: contemplation of the application of new technology-based policies in Indonesia. In Big Data analysis there are several research methods used, sentiment analysis, Support Vector Machine, parametric forecasting techniques, prediction techniques in the health sector, Big Data is used to prevent epidemics and cure a disease. Big Data can also be generated from IoT such as sensors, mobile phones [6]. Research on the use of Big Data and protection of consumer privacy in the digital economy era, explains Big Data as a collection of data that is very large (volume), very fast changing/growing (velocity), comes in various forms/formats (variety), and has value. certain (value), comes from an accurate source (veracity) [7]. Tableau provides dashboards and scorecards, ad hoc analysis and queries, online analytical processing, data discovery, BI search, spreadsheet integration, and more. The introduction outlines the background of the problem being solved, the issues related to the problem being solved, the research reviews that have been done before by other researchers that are relevant to the research conducted [8].

2. RESEARCH METHOD

Big Data analysis research method using Business Intelligence on the spread of Covid-19, Figure 1.

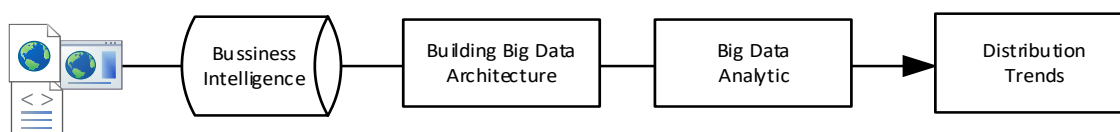


Figure 1. Research Flow

2.1. Data Cleaning

The stages of preparing relevant data from online media data and offline data, relating to the spread of Covid-19. Data has the potential to be processed as useful information. The variety of data increases, the data flow is unpredictable, changes frequently and varies greatly. Variables of data collected are events per day (percentage) doing protocol health, male and female (gender), type of public place (location), regulatory rules (characteristics) [9]. Cleansing data using Business Intelligence techniques, namely the process of extracting, transforming, managing, and analyzing business data to support decision making. Form a large database so that it is easy to identify [10]. Business Intelligence techniques are used to consolidate data, analyze, store and access large amounts of data to assist in decision making, such as software for database querying and reporting, tools for multidimensional data analysis, and data mining. Business Intelligence Stages [11]:

- 1) Identify the business problem that needs to be solved with the data warehouse and determine the data needed to solve the problem.
- 2) Identify the location of the required data and retrieve it from the storage source.
- 3) Changing the data obtained from these various sources into a consistent data.
- 4) Retrieve the modified data into a centralized location.
- 5) Create a data warehouse with existing data in a centralized location.
- 6) Install a product or application that can provide access to the data contained in the cube.

Data is growing at a rapid rate, generated from multiple sources in enormous volumes. The stages of Big Data implementation are determining the strategy used for Big Data,

identifying Big Data sources, accessing, managing and storing data using a database warehouse, cloud, data lake or Hadoop, data analysis [12].

2.2. Big Data Analytics

Data is growing at a rapid rate, generated from multiple sources in enormous volumes. The stages of Big Data implementation are determining the strategy used for Big Data, identifying Big Data sources, accessing, managing and storing data using a database warehouse, cloud, data lake or Hadoop, data analysis. Displays the results of analyzing and organizing big data by finding patterns, trends and useful information [13]. Many statistics and machine learning algorithms can be used for calculations on data sets that are too large to be stored in memory. Data analysis and machine learning with for multidimensional data analysis, statistics and the Machine Learning toolbox provides feature selection, stepwise regression, Principal Component Analysis (PCA), regularization, and other dimensionality reduction methods that allow identifying variables or features [14]. Using K-Nearest Neighbors, K-Means, K-Medoid, hierarchical clustering, Gaussian mixed model, and hidden Markov model. Tableau tools are very user friendly, capable of processing multiple data sources, combining various data sources, big data, spreadsheets, clouds, and various other data types, with a mobile friendly dashboard [15].

3. RESEARCH RESULTS AND DISCUSSION

Big Data is used to optimize structured and unstructured data sources in making decisions to break the Corona virus chain. Import online data in the form of text file formats, spreadsheets, and other file formats accessed from the web. In the form of various and large files, data pre-processing is then carried out for data cleaning and data grouping. Import data in the form of text files by interactively selecting data, using the MatLab Import Tool feature.

Research data on the distribution of Covid 19 during May 2021 sourced from: <https://bing.com/covid/?vert=vaccineTracker&form=WSHCOV>, <https://covid19.go.id/>, <https://health.detik.com>, <https://covid19.sumutprov.go.id/>. Cleaning and clustering data require pre-processing techniques to ensure accurate, efficient or meaningful analysis. Data cleaning refers to methods for finding, deleting, and replacing bad or lost data, correcting data as well as whether it is feasible to go to the next stage, Figure 2.



Figure 2. Source Data

Detect local extremes that occur from sudden changes to help find significant trends in data or data models. Softening and detrending methods are used to remove the disturbances that arise and perform a temporary scaling of the data boundaries. Data processing that begins with data cleansing using BI is input into architectural model. The graph of the increase in cases per day during May 2021 shows that West Java Province is the highest with a total of

31,114 cases. In Big Data, three models of the spread of Covid-19 cases were obtained during May 2021, the results of the analysis were grouped into 3 clusters based on the total number of sufferers per day, Table 1.

Table 1. Covid-19 Spread Cluster May 2021

Date	Clusters	Cases	Date	Clusters	Cases	Date	Clusters	Cases
8 Mei	2	11623	27 Mei	3	9227	2 Mei	3	10585
9 Mei	2	7209	28 Mei	3	10697	3 Mei	3	9595
10 Mei	2	11623	29 Mei	3	10400	19 Mei	3	9371
11 Mei	2	7039	30 Mei	3	11386	20 Mei	3	11036
12 Mei	2	8612	31 Mei	3	10538	21 Mei	3	10606
13 Mei	1	6650	1 Mei	2	8389	22 Mei	3	10291
14 Mei	1	5133	4 Mei	2	8674	23 Mei	3	9647
15 Mei	1	4833	5 Mei	2	12103	24 Mei	3	9081
16 Mei	1	6528	6 Mei	2	8690	25 Mei	3	9626
17 Mei	1	8594	7 Mei	2	12111	26 Mei	3	9446
18 Mei	1	8191						

Big data consists of layers, Data Ingestion sheets, Data Collector sheets, Data Processing sheets, Data Storage sheets, Data Query sheets, Data Visualization sheets. Data Preparation phase, requires an analytical sandbox, working with data and performing analytics, carrying out an extract, load and transform (ELT) or extracting, transforming and load (ETL) process to prepare data to the sandbox. Data visualization with Tableau focuses on business intelligence, making it easier for analysts to perform data analysis by conducting data drills and roll ups. Drill down and roll up are operated to view global or detailed data along the dimensional hierarchical levels. Roll up to view data globally or summaries. Drill down guides users to obtain more detailed data. Drill down is used to answer questions on a particular case. The Discovery phase, seeks and investigates facts, problems (identification of the problem base), develops rules of knowledge, learn the format of origin of the necessary data from those available for successful project analysis decisions. The Discovery phase assesses the resources available to support technology, time, and data projects. Building a Planning Model, such as determining methods, techniques, and workflows. Stages of development form model, developing datasets testing stage, training stage, and production purposes (generating new data from existing data). In addition, in this phase the team builds and executes a model based on the work done in the Model form planning phase. The Communicate Results phase determines the success or failure of the results based on the criteria developed in the initial phase by identifying key findings, measuring size of business requirement, developing explanation summarizes and follows up on findings to stakeholders. on operational processes, showing final reports, directives, codes, and technical documents.

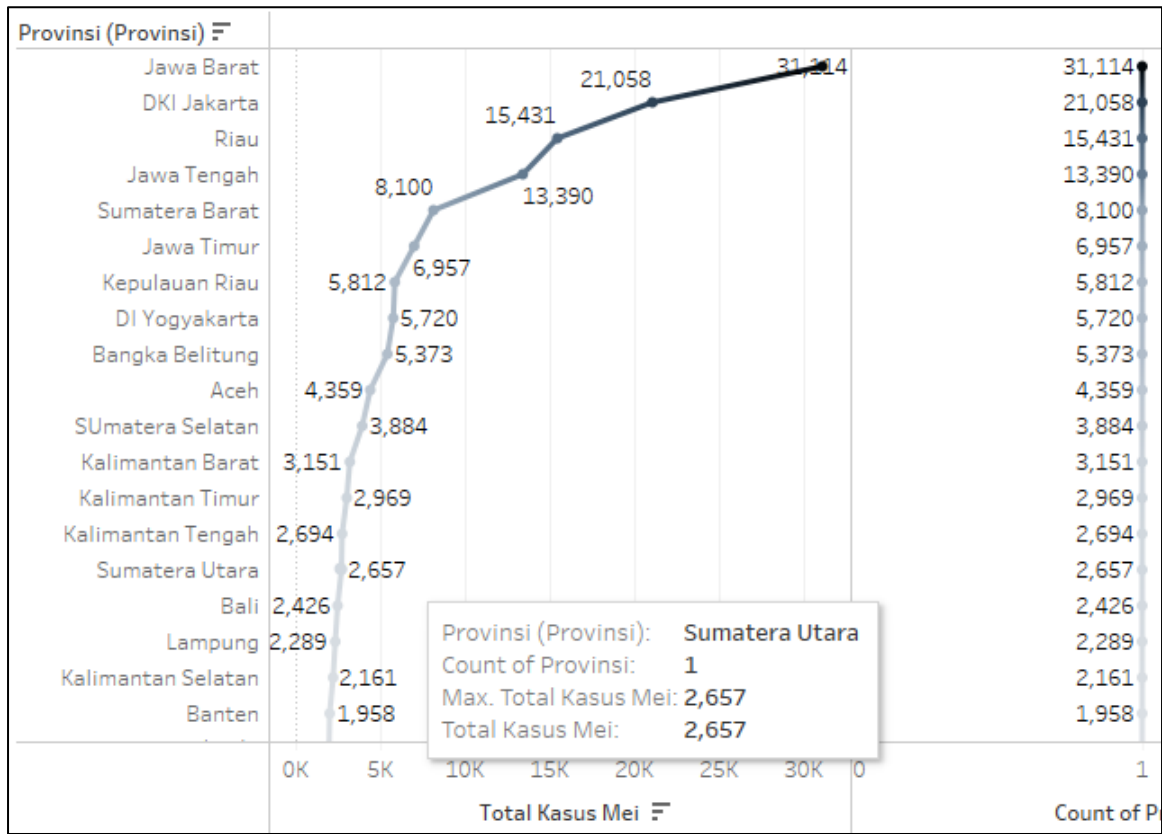


Figure 3. Rating spike in cases of the pandemic in May 2021

Stages of implementing big data analytics:

- 1) Data mining. There are two things that are focused on in big data analytics, namely data mining and data extraction. stages of collecting data from various web links into a data shelter.
- 2) Data collection. in Big data there is no complete button so that data can continue to grow data is entered in various formats. Not only increasing because of new data, data extraction must continue to be carried out to collect data changes that occur from each person.
- 3) Data store. Storing data, especially large data, certainly cannot be arbitrary. a new shelter is needed for the results of the data processing. Not only that, good storage as well as large storage space.
- 4) Data cleaning. The data obtained from the big data analytics process is obtained entirely through the internet. From 100% of the data that has been obtained, there may be 30%-40% of data that are inaccurate and not needed by the company. Therefore, it takes data cleaning alias data cleaning to filter which data is needed or not. From this, data analysts don't have to bother analyzing and figuring out which data to use. Through this step, the data analyst will immediately get the data according to the company's wishes because it has been sorted automatically.
- 5) Data analysis. Data analysis. a series of activities to analyze large amounts of data. Analyzing the data, the data analyst will enter into the patterns and habits of the audience and find out which ones are most needed by the client. Analysis is the process of asking specific questions and finding the right answer. Qubole and statwing are claimed to be very powerful analytical tools for this process.

- 6) Data consumption. Data is used for different purposes and needs by companies, governments, agencies and even organizations.

The total cases that emerged, data were analyzed which showed compliance with health care procedures and those that ignored health care rules, the data source for the results of the COVID-19 Pandemic Community Behavior Survey, (7-14 September 2020), Figure 4.

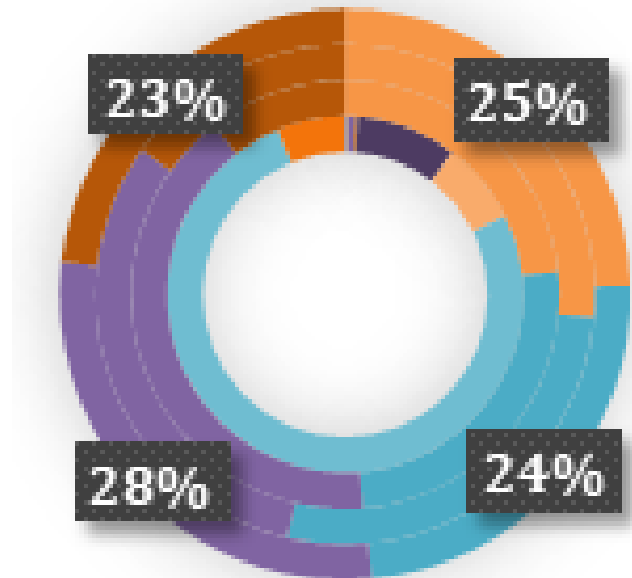


Figure 4. Percentage of Behavioral Analysis Implementing Health Protocol

The online survey data on health care behavior argues that there is no sanction as a reason for the community not to implement the health protocol. Community behavior towards the implementation of health protocol rules obtained at work 2.08%, malls/plazas/shopping places 1.69%. traditional markets/traders K517.32%, places of worship 5.78%, Public Services 1.40%. The percentage of compliance with the recommended habits during the pandemic is 25% distancing, washing hands / using hand sanitizer 24%, wearing masks 28%, checking thermoguns 24%. Average total cases per month of May, grouped into 3 clusters, clusters 1 shows Low Cases, Cluster 2 shows Medium Cases, Cluster 3 shows High Cases, Figure 5.

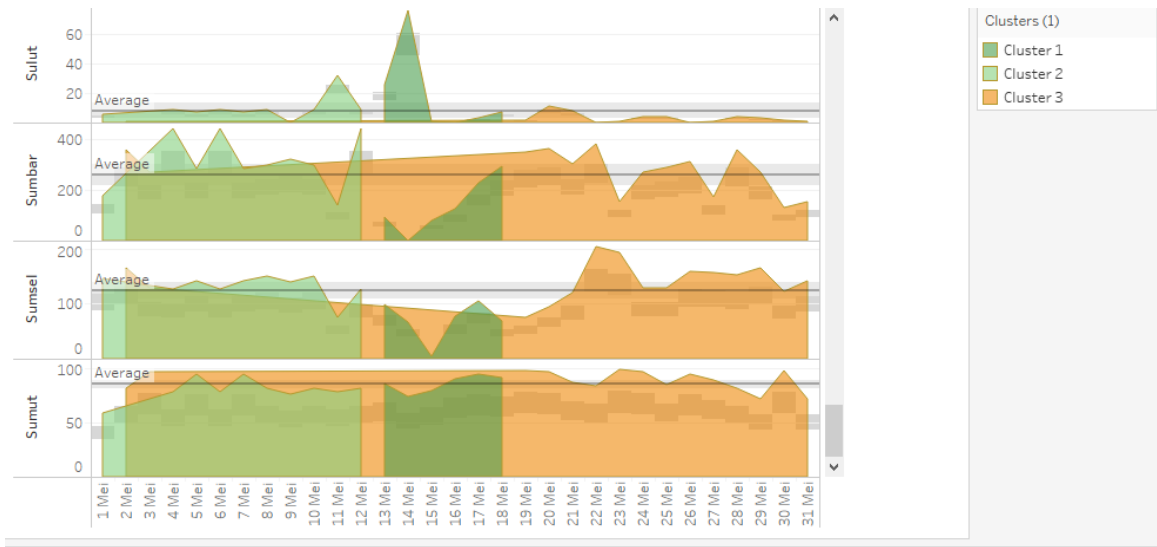


Figure 5. Trend of Spread of Covid-19 May 2021

The percentage of people's awareness that they do not implement the regulations on social distancing procedures, wash hands/use hand sanitizer, wear masks, the highest thermogun inspection in Traditional Markets/K5 Traders is 17.32%, while the lowest is for public services 1.4%. Of the procedures that are required, the procedures for wearing masks are more adhered to in public services 94.83%, the lowest distance distancing procedures are carried out in Malls/Plazas/Shopping Places 2.08%. Habits of health protocols between men and women have significantly different percentages according to the location of the activity, Figure 6.

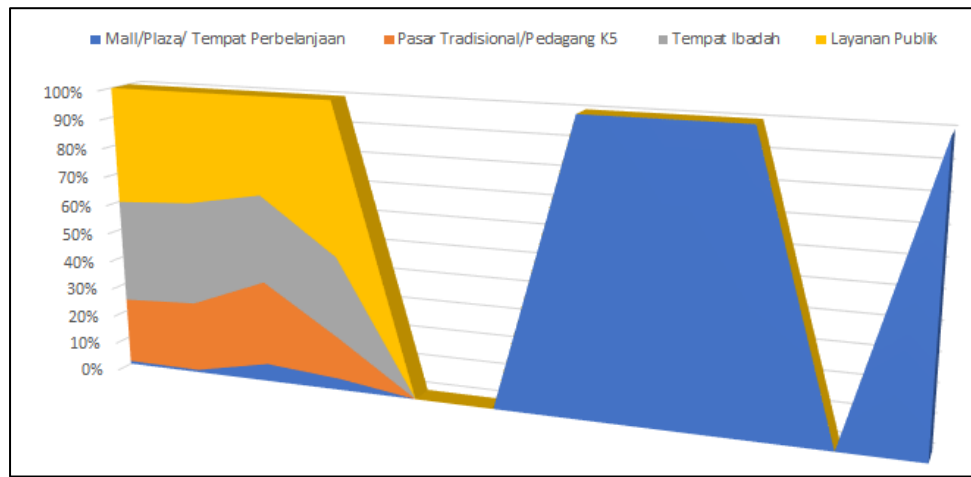


Figure 6. Visualization of Compliance with Health Protocols

4. CONCLUSION

Big Data analysis is sourced from a large amount of data in the form of structured and unstructured data with an increasing volume, such as Covid-19 distribution data, compliance data carrying out procedures in public places is modeled to form a Big Data architecture generating trends or patterns to help make decisions. Organizing data by drill down and roll up

is used to accommodate data from various perspectives, the types of data are processed using algorithms and data representations

5. SUGGESTED

Thank you to LPPM STMIK Pelita Nusantara for the assignment of implementing the 2021 research program.

6. REFERENCES

- [1] B. Maryanto, "Big Data dan Pemanfaatannya," *Media Inform.*, vol. 16, no. 2, pp. 14–19, 2017, [Online]. Available: https://jurnal.likmi.ac.id/Jurnal/7_2017/0717_02_BudiMaryanto.pdf.
- [2] M. Z. Kastouni and A. Ait Lahcen, "Big data analytics in telecommunications: Governance, architecture and use cases," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2020, doi: 10.1016/j.jksuci.2020.11.024.
- [3] A. Pujiyanto, A. Mulyati, and R. Novaria, "PEMANFAATAN BIG DATA DAN PERLINDUNGAN PRIVASI," vol. 15, no. 2, pp. 127–137, 2018.
- [4] P. F. Kurnia and Suharjito, "Business Intelligence Model to Analyze Social Media Information," *Procedia Comput. Sci.*, vol. 135, pp. 5–14, 2018, doi: 10.1016/j.procs.2018.08.144.
- [5] A. P. Narendra, "Big Data, Data Analyst, and Improving the Competence of Librarian," *Rec. Libr. J.*, vol. 1, no. 2, p. 83, 2016, doi: 10.20473/rlj.v1i2.1162.
- [6] S. Mubaroq and I. M. Insyiroh, "TEKNOLOGI KECERDASAN BUATAN , BIG DATA ANALYSIS , DAN INTERNET OF THINGS : POTENSI DAN PERANNYA DALAM PENANGANAN COVID-19 DI INDONESIA (ARTIFICIAL INTELLIGENCE TECHNOLOGY , BIG DATA ANALYSIS , AND INTERNET OF THINGS : THE POTENTIAL AND THE ROLE TO FIGHT ," vol. 2902, pp. 109–114, 2020.
- [7] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [8] K. K. Halim, S. Halim, and Felecia, "Business intelligence for designing restaurant marketing strategy: A case study," *Procedia Comput. Sci.*, vol. 161, pp. 615–622, 2019, doi: 10.1016/j.procs.2019.11.164.
- [9] A. Q. Syarli, Rosmawati Tamin, "Perancangan Business Intelligence System Pada Gudang Farmasi Dinas Kesehatan Kabupaten Mamasa," *JUTEKS (Jurnal Keteknik dan Sains)*, vol. 1, no. 1, pp. 7–14, 2018.
- [10] S. C. Huang, S. McIntosh, S. Sobolevsky, and P. C. K. Hung, "Big Data Analytics and Business Intelligence in Industry," *Inf. Syst. Front.*, vol. 19, no. 6, pp. 1229–1232, 2017, doi: 10.1007/s10796-017-9804-9.
- [11] M. Alnoukari, "From Business Intelligence to Big Data," no. January, pp. 44–62, 2020, doi: 10.4018/978-1-7998-5781-5.ch003.

- [12] A. Yunistira and H. Fudholi, "Analisis Penerapan Model Business Intelligence pada Aplikasi Payment Point Online Banking dalam Meningkatkan Strategi Pemasaran (Studi Kasus: Aplikasi ApotikKuota) Analysis of Business Intelligence Model Implementation in Payment Point Online Banking to I," vol. 7, pp. 1–10, [Online]. Available: <http://journal.ipb.ac.id/index.php/jika>.
- [13] S. J. Alsunaidi *et al.*, "Applications of big data analytics to control covid-19 pandemic," *Sensors*, vol. 21, no. 7, 2021, doi: 10.3390/s21072282.
- [14] D. Horgan *et al.*, "Digitalisation and COVID-19: The Perfect Storm," *Biomed. Hub*, vol. 5, no. 3, pp. 1–23, 2020, doi: 10.1159/000511232.
- [15] A. S. G. A S Sinaga, "University Accreditation using Data Warehouse," *J. Phys. Conf. Ser.*, vol. 755, no. 1, 2016, doi: 10.1088/1742-6596/755/1/011001.