Design of big data technology prototype for classification of village status based on village development index involves k-means algorithm to support village ministry Pdtt work programs

Paska Marto Hasugian\textsuperscript{1*}, Harvei Desmon Hutahaean\textsuperscript{2}, Bosker Sinaga\textsuperscript{1}, Sriadhi\textsuperscript{2}, Saronom Silaban\textsuperscript{3}

\textsuperscript{1}Informatics Engineering Study Program, STMIK Pelita Nusantara, Medan, Indonesia
\textsuperscript{2}Computer Science and Informatics Education Study Program, Faculty of Engineering, Universitas Negeri Medan, Medan, 20221, Indonesia
\textsuperscript{3}Deparment of chemistry, Universitas Negeri Medan, Medan, 20221, Indonesia

*Email: paskamarto86@gmail.com

Abstract. Big data technology is the overall technology that can handle the processing associated with analyzing the data to explore the potential that is in it. Some of the uses of big data are based on the biggest data traffic sources, such as social media, financial transactions, public data, censorship data, and company data. The problem found in the accumulation of village status data that has not been utilized in the decision-making process to determine the status of the next village and the absence of an application to process the data stored in the data database so that a tool is developed to analyze the buildup. The results of this study are to produce a prototype design that is Unified Modeling Language (UML) design, Form Design, and Database Design that will be used to develop Big Data Technology in determining the classification of village status based on the Village Build index.

Keywords: Prototype, Bigdata, UML, Design

1. Introduction

Big Data is a term given to data sets that are very large and complex in size, making it impossible to process using conventional database management tools or processing applications [1-7]. Other data and according to Gartner IT big data is a Glossary Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation [8], according to these definitions it is necessary to develop a supporting device that can be used to process data that has been stored for a very long time and there is an accumulation of data grouping in Indonesian villages that have been recorded by the village ministry. In the big data application development process, several steps to completion are needed, namely software design.

Software design is focused on designing software requirements starting from UML (Unified Modeling Language) design, Interface Design, and database design. UML is used to model a system using object-oriented concepts [9-14]. And also to create a modeling language that can be used by
both humans and machines with several parts, namely use cases, activity diagrams, sequence diagrams, and class diagrams.

The use case in this design is a design that provides an overview of the role between the user and the system or more duties and responsibilities than the user. Activity diagrams are diagrams that function to describe or model user activities on the system as a whole from the initial formation to the analysis stage using algorithms [15-16]. Sequence Diagram is a diagram that is used to describe the behavior of objects in use cases by describing the lifetime of objects and messages sent and received between objects [17-18]. The class diagram is a diagram that shows the relationship between classes in which there are attributes and functions of an object or more dominant to the relationship between a.

The interface design developed is a form used in application development, in this case, it is related to the input, process, and output of the application to be developed, to provide communication between the user and the system [19]. The User Interface can receive information from the user and provide information to the user to help mobilize a problem tracking path until a solution is found. The design that must also be involved in a database that functions as a container for storing data related to the application to be developed as well as the connection between data and according to database theory is a set of data that has been arranged in such a way with certain related terms or rules making it easier for users to manage it also makes it easier to obtain information. [20-22] This research is a stage of developing big data applications in determining the status of the village in general. Some of the steps used are identifying problems and designing features that are loaded in the big data application.

2. Research Methods
In supporting the design of Big Data technology, this process is focused on designing to model software design into diagrams, with several activities, namely UML design with several designs, namely use cases, activity diagrams, sequence diagrams, and class diagrams. interface. And database design. With the following diagram:

![Figure 1. Steps of Making a Prototype](image-url)
The explanation of the work steps is described below:

a. Identification of problems
   At this stage, it ensures what is the need for system development which includes Input needs, Process Requirements, and Output needs.

b. UML (Unified Modeling Language) design,
   This stage is the stage of making the diagrams needed in system development according to the findings at the identification stage, the UML design used is use cases, activity diagrams, sequence diagrams, and class diagrams.

c. Database design
   This stage includes the database design required in database development, starting from the database for login, the database for the process, and the final result database.

d. Interface design
   This stage is the design that becomes an important part of the research to interact directly with the user.

3. Results and Discussion

3.1 Use case diagram
The use case below illustrates graphically the form of interaction between the user and the system. And what functions are contained in the system? Use cases have several forming elements that are used to denote the intent of the system. There is an element or actor symbol that is used to define the user or users who use the system. And there is the use case symbol itself which is used to describe what actions are contained in the system.

![Figure 2. Use Case Diagram](image)
Some of the descriptions of the use case in the figure are the login process for self-verification as a system user, data input is a process used to input village status data which will be grouped by uploading data or manual data input, for cluster center input is the facility used for choosing the central point, the K-means process is the design or facility used to perform calculations, while the logout is the facility to exit the system.

3.2 Activity Diagram
Activity diagrams are diagrams that function to describe or model the activities that users perform on the system as a whole starting from login.

a. Activity Diagram Login
To log in, the user must use a username and password. Enter the username and password in the login form at the beginning of opening the system with a picture.

![Activity Diagram Login](image)

**Figure 3. Activity Diagram Login**

b. Data Upload Activity Diagram
To enter the data to be clustered into the system, the user can upload the upload form, the uploaded file will then be saved directly to the database by the system.
c. Activity Diagram Input Cluster

To determine the initial cluster center, the system displays the data that will be in the cluster, from which data can then be selected which data will be used as the cluster center.

d. K-means Process Activity Diagram

The k-means process is carried out only by pressing the process button provided in the system, this is done because the formula of the K-means clustering algorithm has been placed in the program. In this case, the user only remains.
3.3. **Sequence Diagram**

The sequence diagram is one of the diagrams that are included in the UML. Sequence Diagrams are diagrams used to describe the behavior of objects in use cases by describing the lifetime of objects and messages sent and received between objects.

a. **Login Sequence Diagram**

The login process is the process of matching the data entered by the user with the data that is already available in the database. The username and password data entered by the user is checked against the data in the database.

![Login Sequence Diagram](image-url)
b. Sequence Diagram Upload Data
Upload data is the process of entering data to be clustered into the system. This process is done via the upload form. Then the insert function is performed to enter the uploaded data to the database.

![Sequence Diagram Upload Data](image)

**Figure 8.** Sequence Diagram Upload Data

c. Input Cluster Sequence Diagram
The sequence diagram for input data cluster is a diagram that describes the work process of inputting cluster data with several main controls, namely the admin input cluster data, the main menu displays the input form and cluster value and provides information that the upload is successful; the interface is a connection between input, create, query and execute and close that serve with connection databases and cluster tables.

![Input Cluster Sequence Diagram](image)

**Figure 9.** Input Cluster Sequence Diagram

d. Sequence Diagram Logout
Sequence logout diagram is a diagram depicting the work process of the logout function when the application was developed with a general description that is used is the admin main menu and interface.
3.4 Class Diagram
The class diagram is a model used to describe how the table structure in the Class Diagram database on a system is designed. The following is a class diagram that will be made for this system. With the name of the K-Means database and adding 3 tables to accommodate the data needed to carry out village grouping processing.

![Class Diagram](image)

**Figure 11.** Class Diagram

3.5 Design Interface
Interface design for big data development with the following display:

1. **Design Login Form**
   The login design form displays when the application is first to run. In the username and password dialog box, the admin must enter the user id and password. And if the admin does not enter the correct user id and password, then the admin cannot enter the application. In the login screen, to enter the next page, the admin can press the login button. And then, the admin will go to the main menu page.
Figure 12. Design Login Form

2. Main Menu Design
The main form menu appears if the login is successful, and in the main menu form there are several sub-menus such as 1) home the main view; 2) data input; 3) cluster input; 4) process and 5) logout as a button to exit the system.

Figure 13. Main Menu Design

3. Village Data Input Design
Village Data Input Design is an interface form design that is specifically for inputting cluster data used, some of the related features are village data, namely the name of the village, district, district and province and the IDM value regarding the IKS, IKE and IKL values with the help button is save for data storage, edit for data changes and delete for data deletion control.

Figure 14. Village Data Input Interface Design
3. K-Means Process Design
The K-means process design form is prepared for the village classification process, by uploading previously processed village data.

![Figure 15. K-Means Process Design](image)

4. Conclusion
From the results and discussion, a design has been produced that can support the Big Data application development process to produce a decision in determining the classification of village status based on IDM (Build Village Index). The research resulted in a number of supporting UMLs, namely Use Case, Form Design used in Development, Database Design, and Hypo Diagram Design.

Reference
[1] M. Mardani, G. Mateos, and G. B. Giannakis, “Big Data,” in Cooperative and Graph Signal Processing: Principles and Applications, 2018.
[2] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, “Data mining with big data,” IEEE Trans. Knowl. Data Eng., 2014.
[3] B. N. Silva, M. Diyan, and K. Han, “Big Data Analytics,” in SpringerBriefs in Computer Science, 2019.
[4] B. Schultz, “Big data in big companies,” Baylor Bus. Rev., 2013.
[5] R. Sowmya and K. R. Sunneetha, “Data Mining with Big Data,” in Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017, 2017.
[6] Big Data, “BIG DATA : retos y oportunidades para el turismo,” Agència Valencia. del Tur., 2015.
[7] S. Sagiroglu and D. Sinanc, “Big data: A review,” in Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013, 2013.
[8] M. D. L. L. G. P. Bob Gill, “Gartner Reprint,” Gartner, 2015.
[9] S. Chren, B. Buhnova, M. MacAk, L. Daubner, and B. Rossi, “Mistakes in UML diagrams: Analysis of student projects in a software engineering course,” in Proceedings - 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering Education and Training, ICSE-SEET 2019, 2019.
[10] R. Fauzan, D. Siahaan, S. Rochimah, and E. Triandini, “Use case diagram similarity measurement: A new approach,” in Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019, 2019.
[11] V. Yasin, O. S. Sitompul, M. Zarlis, and P. Sihombing, “Big data measurement model in achieving maximum accuracy using the model Hierarchy of Grid Partition (HGP) method,” in 2019 3rd International Conference on Electrical, Telecommunication and Computer Engineering, ELTICOM 2019 - Proceedings, 2019.
[12] Z. A. Hamza and M. Hammad, “Generating UML use case models from software requirements using natural language processing,” in 2019 8th International Conference on Modeling Simulation and Applied Optimization, ICMSAO 2019, 2019.

[13] D. Arcelli, V. Cortellessa, and D. Di Pompeo, “Automating Performance Antipattern Detection and Software Refactoring in UML Models,” in SANER 2019 - Proceedings of the 2019 IEEE 26th International Conference on Software Analysis, Evolution, and Reengineering, 2019.

[14] S. Latif, A. Rehman, and N. A. Zafar, “Blockchain and IoT based formal model of smart waste management system using TLA+,” in Proceedings - 2019 International Conference on Frontiers of Information Technology, FIT 2019, 2019.

[15] M. T. Rizky Muhammad; Irma Kartika Wairooy, S.Kom., “UML Diagram : Activity Diagram,” https://socs.binus.ac.id/, 2019.

[16] C. Britton and J. Doake, “Activity diagrams,” in A Student Guide to Object-Oriented Development, 2005.

[17] B. Selić and S. Gérard, “An Introduction to UML Profiles,” in Modeling and Analysis of Real-Time and Embedded Systems with UML and MARTE, 2014.

[18] Rajagukguk J, Kaewkhao J, Djamal M, Hidayat R, Ruangtaweep Y. Structural and optical characteristics of Eu3+ ions in sodium-lead-zinc-lithium-borate glass system. Journal of Molecular Structure. 2016 Oct 5;1121:180-7.

[19] Rajagukguk J, Hidayat R, Djamal M, Ruangtaweep Y, Horprathum M, Kaewkhao J. Structural and optical properties of Nd3+ doped Na2O-PbO-ZnO-Li2O-B2O3 glasses system. InKey Engineering Materials 2016 (Vol. 675, pp. 424-429). Trans Tech Publications Ltd.

[20] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” Sensors. 2012.

[21] B. Smit, P. A. J. Hilbers, K. Esselink, L. A. M. Rupert, N. M. Van Os, and A. G. Schlijper, “Computer simulations of a water/oil interface in the presence of micelles,” Nature, 1990.

[22] E. L. Hutchins, J. D. Hollan, and D. A. Norman, “Direct Manipulation Interfaces,” Human-Computer Interact., 1985.