RECORD : Journal of Loyality and Community Development
E-ISSN : 3032-7946, P-ISSN : 3032-7954
Volume 1 No 1 January – April (2024)

Analyzes Health Data By Applying Data Science For Students in The Undergraduate Study Program in Public Health Faculty of Health Hang Tuah University Pekanbaru

Eka Sabna[1], Zupri Henra Hartomi[2], Yayang Sahira[3]

Universitas Hang Tuah Pekanbaru, Indonesia[123]

ABSTRACT

Health services are one of the rapidly growing public service sectors currently, resulting in large piles of patient medical record data. This pile of data can provide valuable knowledge if processed in the right way. Data Science is a series of processes for exploring hidden knowledge patterns in large data sets. Data Science can be applied to discover knowledge patterns from patient profiles and health history data. The knowledge gained can be used for analysis and decision making, including to predict the type of disease, determine the pattern of disease spread, and see the effectiveness of treatment. So far, students from the Hang Tuah University Pekanbaru Public Health Study Program have been carrying out the data analysis process using statistics. Community Service Activities aim to enable students to use Data Science as an alternative in analyzing health data. So students can use Data Science to help analyze health data in their research. Data Science techniques discussed include Basic Concepts and Data Science Algorithms. The output of this PkM activity is increasing partner skills, publication in mass media and scientific publications.

Keywords: *Data Science, Health Data, PkM, Data Analysis, Students*

Corresponding Author:
Eka Sabna
es3jelita@yahoo.com

Received: February 28, 2024
Revised: March 15, 2024
Accepted: April 01, 2024
Published: April 25, 2024

1. INTRODUCTION

We can also say that Data Science is a series of calculation and reasoning processes to explore more value in the form of information that has not been known manually from a database so far (Han, 2011). Data Science is the process of extracting hidden knowledge from data automatically or semi-automatically. This series of processes is carried out by extracting various patterns from data which are then manipulated, in order to obtain more valuable information and recognize patterns by extraction in a database.

The use of Information Technology itself has been used in many fields, one of which is the health sector. In fact, in 2012, the Minister of Health (Habibah et al., 2023) launched a national patient safety movement. So from that moment on, hospital information systems began to develop that could make things easier for the public and the hospital itself. The Data Mining approach is also one of the technological developments in the computer field(Ha et al., 2012) (Daniel T, 2005). Analyzing health data using data mining technology is an important step towards improving the quality of health services. By understanding the potential of data mining, clinics and healthcare facilities can take full advantage of this technology to optimize healthcare services. The following is some research and outreach on the application of data mining, namely research conducted by Nurul Rofiqo (2018) aimed at utilizing the Clushtering Algorithm in grouping population numbers (Rofiqo et al., 2018), educational activities for Amanah Clinic employees about the benefits of data mining in medical data analysis (Abdillah et al., 2023), the application of data mining using the multiple linear regression method is used to predict the growth in the number of HIV

patients so that the provision of required treatment can be adjusted to the number of available HIV patients (Zunaidi et al., 2020). The application of Data Science increases operational efficiency such as drug inventory management, patient scheduling, and administration management. Data Science also improves clinical decision making, monitors patient care outcomes, identifies risks, and improves quality of care.

The Public Health Study Program is one of the Study Programs at Hang Tuah University Pekanbaru. Public Health Study Program students do not know much about the basic concepts of Data Analysis using Data Science concepts. The solution to this problem is through community service activities (Pengabdian kepada Masyarakat-PkM) with training in Health Data Analysis using a Data Science approach. This training activity aims to provide basic knowledge and skills about the main concepts and roles in Data Science, the stages of the Data Science process in solving Data Science cases specifically in the health sector. The benefit of this training is that students' understanding of the Public Health Study Program will increase.

2. METHODS

Community Service Activities Training for students will be carried out from July to August 2023. The number of students who take part in this activity are students from the 4th semester of the Public Health Study Program. The method for implementing this PkM activity is divided into several stages. The stages carried out in implementing this PkM are as follows:

Figure 1. Methods for Implementing PKM Activities

The stages in Figure 1 are details of PkM activities.

1. Licensing
   Licensing process is carried out together with the Public Health Study Program.
2. Determining Time and Place
   Coordination of the PKM Team with the Study Program. The team conducted surveys, observations and held discussions as well as set the activity agenda, coordinating regarding time and place.
3. Preparation.
   Preparation of Materials and Participants for PKM activities
   a. Prepare topics or materials that will be presented in PKM activities. The determination of materials and training participants is carried out by the PKM team to achieve the objectives of PKM.
   b. Participant
      Student Participants from the Public Health Study Program

4. Implementation.

Training for students, the material will be delivered following the steps in Data Mining. The steps are as follows:

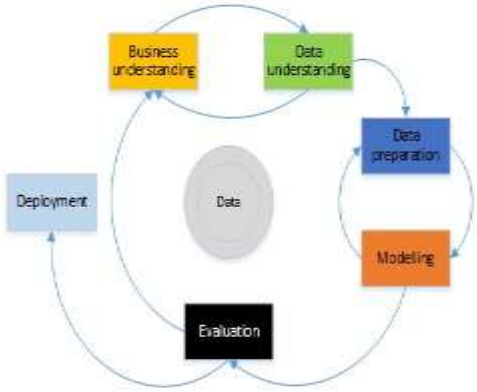


Figure 2. Method of Implementing PKM Activities

Figure 2 shows the stages of CRIPS-DM, a methodology for implementing training. The CRISP-DM stages are a model for solving problems using a Data Science approach. The CRISP-DM process model is a Data Science process cycle which has 6 stages (Mauritsius & Binsar, 2020). The following is an explanation of the CRISP-DM cycle (Dqlab, n.d.):

1) Business Understanding (Business Understanding)
   a. Analyzing Needs
   b. Determining Goals and Plans
   c. Data mining problem formulation.
2) Data Understanding (Data Understanding)
   Data collection, further identifying the data that will be used. The data used in this training comes from the Central Statistics Agency website, namely https://www.bps.go.id/id in the health sector, Kaggle and data is also obtained from health sector journals.
3) Data Preparation (Data Preparation/Processing)
   This is a very important stage for designing a prediction model. To improve the quality of the data to be analyzed, it is necessary to carry out Data Preparation steps, namely (Flin, n.d.) :
   a. Data Cleaning (Data Cleaning)
      Data that has just been collected may have many incongruent parts and some missing parts, so a data cleaning process is needed.
   b. Data Transformation
      Data transformation is used to change data in an appropriate form.
   c. Reducing Data (Data Reduction)
      Aims to increase storage efficiency and reduce data storage and analysis costs.
4) Modeling (Modeling)
   a. Using the resulting model.
   b. The accuracy of a prediction is determined by how big the deviation or error occurs between the predicted data and the actual data .
   c. Validation is a very important stage of modeling, to see the extent of the reliability of the model that will be used in decision making.

5) Evaluation (Evaluation)
Evaluate one or more models in use and determine whether any model meets the objectives at an early stage.
6) Deployment
Using the generated model and generating reports.

5. Activity Evaluation and Reporting
a. Evaluation
After PKM activities are implemented, evaluation needs to be carried out by:
1). Process Evaluation, namely measuring student activity in training
2). Evaluation of Results, namely measuring student understanding of the material presented.
b. Report
The final stage in this activity is compiling a report from the beginning of the activity to the evaluation stage. This report can be used as a reference.

3. RESULTS AND DISCUSSION

3.1 RESULTS

This activity is to increase basic knowledge and skills about the main concepts and roles in Data Science for students of the Public Health Study Program at Hang Tuah University, Pekanbaru. The discussion material for this training is :
1. Introduction to Data Science Analysis (Dqlab, n.d.)
2. Data Science Analysis Function
3. Objectives of Data Science Analysis
4. Data Science Analysis Process
5. Case examples of applying Data Science analysis in the health sector

Figure 3. Training Activity Materials

All material is explained in presentation slides and modules which have been prepared in detail to make it easier for students to understand the material. Practice solving cases using tools, namely Rapidminer Studio software. Examples of cases in practice are problems related to the health sector. The following are the practices carried out in this training using Classification, Clustering, Estimation and Association Algorithms.

1. Classification

Classification Training Using the Decission Tree Algorithm. The data used is diabetes data. The dataset comes from the web, namely Kaggle. The processed data is separated (split) into training data and testing data with a ratio of 80:20. The result is a Model (Knowledge) in the form of a Decision Tree.

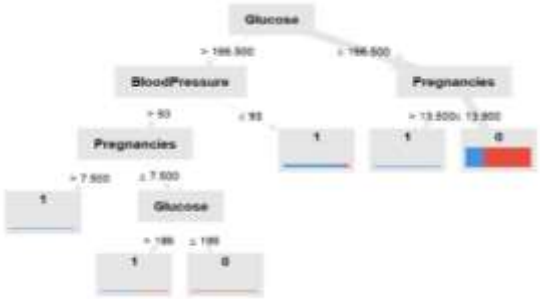
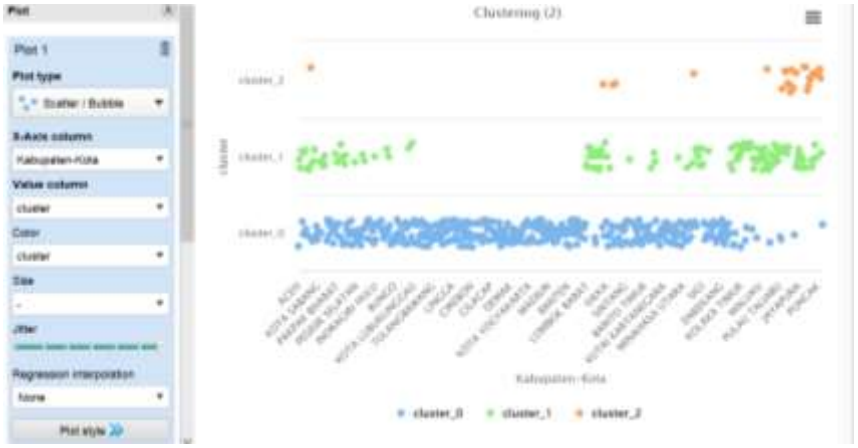


Figure 4. Decision Tree Model View

Figure 4 shows a Decision Tree Model from diabetes data. Through this model, people can find out the variables that determine whether a person suffers from diabetes, so that people can maintain their lifestyle and diet to avoid diabetes.

2. Cluster (Cluster)

Cluster Training uses the K-Means Algorithm. The data used is Child Mortality Rate (CMR) data according to Province-Regency-City in 2020. The data was obtained from the Central Statistics Agency in 2020. The results are data grouping into 3 categories of Child Mortality Rate, namely High, Medium and Low .



Cluster Model

```
Cluster 0: 394 items
Cluster 1: 125 items
Cluster 2: 28 items
Total number of items: 547
```

Figure 5 Cluster model with 3 groups

Figure 5 is a model of the Cluster process with 3 groups, namely Cluster 0 for high CMR numbers, Cluster 1 for medium CMR numbers and Cluster 2 for low CMR numbers. Cluster 0 has 394 regencies-cities, Cluster 1 has 125 regencies-cities and Cluster 2 has 28 regencies-cities.

3. Estimate

Estimation Training using the Linear Regression Algorithm. The data used is ARI disease. The result is Knowledge in the form of Formulation. The variable used is the Independent Variable consisting of 6 variables, namely Industry Ratio (X1), Hospital Percentage (X2), Healthy Latrine Ownership Percentage (X3), Population Density (X4), Average number of people per Household (X5) and Ratio Breeder Business (X6). And the dependent variable is the Percentage of ISPA Sufferers (Y).

Figure 6 Estimation Formulation Model

Figure 6 is the result of the estimation process, the formulation formed from the estimation algorithm is Y = 0,586 – 0,322 X1 + 0,008 X2 – 0,007 X3 – 0,000 X4 – 0,328 X6.

4. Association

Association Training using the FP-Growth Algorithm. The data used is the disease suffered by the patient. The result is knowledge in the form of association rules from patient disease data.

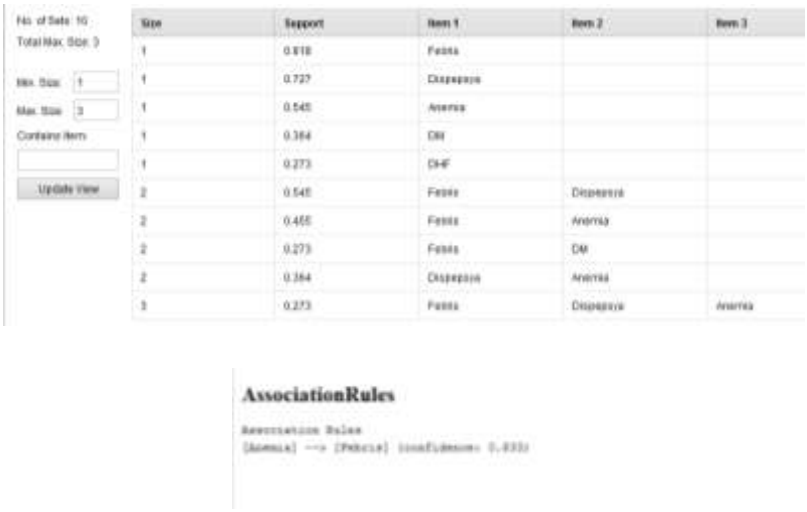


Figure 7. Association Model

Figure 7 shows the results of the Association process, the Association Rules that are formed are If Anemia is Disease then Febris Disease with a Confidence value = 0.833. Based on this rule, if someone has anemia, the patient will also have febrile disease with a confidence level of 83%.

The following is documentation of training activities with students from the Hang Tuah University Pekanbaru Public Health Study Program:

a. Group Foto

b. Questions and Answers during training

c. Delivery of material

d. Evaluation test session

Figure 8 Activity Documentationa

Figure 8 is documentation during the implementation of PkM, Figure 8.a Group Photo after the training was completed, Figure 8.b Student Questions during training using Rapidminer Software, Figure 8.c is during the Delivery of Material and Figure 8.d is the last session, namely implementation of tests for evaluation.

3.2 DISCUSSION

The PkM that has been implemented is then evaluated. The evaluation carried out is

1). Process Evaluation
From this evaluation, the participants who attended this training from start to finish, the participants also actively asked questions and provided their views.

2). Evaluation of Results
This evaluation was carried out by giving tests to students and the results obtained were that 83% of students understood the basic concepts and how to solve cases using the Data Science approach.

4. CONCLUSION

PkM Data Science training to analyze health data can increase students' understanding of the concepts and benefits of Data Science. Through the Evaluation, 83% of students gained a good understanding of Data Science. Suggestions for further activities are to optimize the use of this method by continuing training with advanced concepts from Data Science so that students have a deep understanding of Data Science. So that later students will be able to use it for their research (thesis) and applications in society such as analyzing medical data in health centers or hospitals.Conclusions explain the findings of the study that are relevant to the research question and research objectives without using statistical data. The conclusion section includes the implications of further research.

REFERENCES

Abdillah, N., Susilo, H., & Ihksan, M. (2023). Sosialisasi Pemanfaatan Teknologi Data Mining Untuk Analisis Data Kesehatan Di Klinik Amanah. *Jurnal Abdimas Saintika*, *5*(1), 181–186. https://doi.org/10.30633/JAS.V5I1.1940

Daniel T, L. (2005). DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining. In *Structure and Bonding* (Vol. 134). https://doi.org/10.1007/430_2009_1

Dqlab. (n.d.). *Langkah Awal dalam Pemrosesan Data: Data Preprocessing dalam...* Retrieved April 9, 2022, from https://www.dqlab.id/langkah-awal-dalam-pemrosesan-data-dalam-data-mining

Flin. (n.d.). *Metodologi CRISP-DM Beserta Contoh Kasusnya - Flin Setyadi*. Retrieved November 27, 2022, from https://flinsetyadi.com/metodologi-crisp-dm-beserta-contoh-kasusnya/

Ha, J., Kambe, M., & Pe, J. (2012). Data Mining: Concepts and Techniques. *Data Mining: Concepts and Techniques*, 1–703. https://doi.org/10.1016/C2009-0-61819-5

Habibah, N. N., Nazir, A., Iskandar, I., Syafria, F., Oktavia, L., & Syurfi, I. (2023). Pemodelan Klasifikasi Untuk Menentukan Penyakit Diabetes dengan Faktor Penyebab Menggunakan Decision Tree C4.5 Pada Wanita. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(4), 654–661. https://doi.org/10.30865/JSON.V4I4.6202

Han, J. (2011). *Han and Kamber: Data Mining---Concepts and Techniques, 2nd ed., Morgan Kaufmann, 2006*. https://hanj.cs.illinois.edu/bk3/bk3_slidesindex.htm

Mauritsius, T., & Binsar, F. (2020). *Cross-Industry Standard Process for Data Mining (CRISP-DM) – MMSI BINUS University*. https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/

Rofiqo, N., Windarto, A. P., & Hartama, D. (2018). PENERAPAN CLUSTERING PADA PENDUDUK YANG MEMPUNYAI KELUHAN KESEHATAN DENGAN DATAMINING K-MEANS. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *2*(1). https://doi.org/10.30865/KOMIK.V2I1.929

Zunaidi, M., Nasyuha, A. H., & Sinaga, S. M. (2020). Penerapan Data Mining Untuk Memprediksi Pertumbuhan Jumlah Penderita Human Immunodeficiency Virus (HIV) Menggunakan Metode Multiple Linier Regression (Studi Kasus Dinas Kesehatan Provinsi Sumatera Utara). *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, *3*(1), 137–147. https://doi.org/10.53513/JSK.V3I1.205

.