

A Review of Data Mining Implementation in Various Fields

Mustabshiroh, Roswan Latuconsina and Tito Waluyo Purboyo
Department of Computer Engineering, Faculty of Electrical Engineering,
Telkom University, Bandung, Indonesia

Abstract: Data mining is the process of searching patterns in big data sets and then will be converted into information that is easy to understand. A lot of research has developed the use of data mining to create something that is beneficial to society. Data mining has two general uses for defining patterns to be easily understood by explaining the characteristics of data and predicting. Data mining can be utilized in areas such as business, science, entertainment, etc. with different methods depending on needs.

Key words: Data mining, characteristics, predicting, business, science, searching patterns

INTRODUCTION

Data mining is a combination of a number of computer science disciplines which search new patterns from a very large set of data with various of methods.

The last few years, data are known to be more complex and are increasing rapidly. According to the investigation, the volume of data in 2011 reached 1.8 ZB and in 2012 increased by more than 50% (Gantz and Reinsel, 2013). To describe the large data, it is then known then the term of big data. Data mining can be used to analyze big data to make decisions better. Data mining draws conclusions from many data to be easily understood by humans, so that, it can be processed further. Therefore, data mining is related to the database.

Based on the usefulness of data mining, data mining can be exploited in various angles of view. Data mining can be used for scientific and commercial purposes. Many scientists use data mining to facilitate their research, especially in large-scale calculations or analyzing data that is real-time or large.

Data mining is many applied in various fields like as business, science, entertainment and others. Each field uses data mining with a specific method depending on the need. One field may use more than one method in data mining depending on the data that will be used and the purpose of the use.

MATERIALS AND METHODS

Implementation of data mining

Business: Dr. Arno Penzias in an interview with the 1999 computerworld magazine said that data mining would be

a mainstay in the use of company databases. Data mining will be very important for companies in controlling a very large customer data. Data mining can help companies to maximize profits with minimal expenses (Cao *et al.*, 2008).

Market basket analysis: Studying consumer behavior by looking for patterns of link in data to find products that are often purchased based on transaction data. At supermarkets, data mining can be used in the arrangement of goods layout by placing items that are interconnected with each other (Trnka, 2010). For example, most consumers often buy bread and jam at the same time, so, it is appropriate that bread and jams are placed close to each other. In addition to the layout of goods, this technique can be used as a media campaign (Trnka, 2010). Only by promoting one product can indirectly promote related product then the sales will be increased.

Database is important because the consumer data in the transaction will be stored then processed as needed by doing the analysis to find a combination of products that customers often buy using data mining (Trnka, 2010).

Data mining can be used to predict the time of an item is needed where the amount of sales is very large, so that, the seller can find out the time to increase the inventory (Fazel *et al.*, 2010).

Recommendation system: Data mining can be used to create systems designed to recommend something to others by habit. Recommendation systems are often used on e-Commerce. The way it researches is to match products that will be recommended to customers based on the pull power of the previous search history and purchases (Lawrence *et al.*, 2001).

Recommendations system are built by combining several techniques in data mining such as data filtering and data mining (Wang *et al.*, 2004). Through the habits of buyers in searching for items, data screening and data mining will be done to identify the type of items that are buyer's favorite.

Fraud detection: Build the system using anomaly detection technique by reading the hidden patterns in the database, so that, a fraud will be known. Cheating can be very vulnerable to occur especially in paid transactions such as electricity bills or telephone bills. Cheating can be done for yourself by manipulating the data themselves or catapult it to others, so that, it impacts other customers.

Data mining is most often used to detect fraud. Analyzing the database to look for the behaviors that are out of the ordinary is the foundation to find anything strange. The pattern search is done gradually, starting from analyzing as a whole if unusual pattern is found it will be processed by narrowing the readings it can be decided whether there is cheating (Cabral *et al.*, 2008).

RESULTS AND DISCUSSION

Health

Diagnose the disease: Data mining can read patterns to diagnose diseases, ranging from natural diseases such as heart disease to viral diseases. Data mining explores patterns of patient data such as age, blood pressure and complaints.

The use of data mining can support the decision of laboratory results to be more accurate by combining several algorithms that can be used such as fuzzy algorithms that are used in supporting the decision to diagnose heart disease (Babu *et al.*, 2017).

One algorithm that can be used to diagnose heart disease is the k-means algorithm. The k-means algorithm performs data processing with a clustering system that calculates the average distance to the centroid point and makes the calculation result a new centroid point, the calculation is done to find a centroid point which is the result of data processing. The number of centroid points is determined as much as k which will be placed randomly. The centroid point update can be done by the Eq. 1:

$$A_i = \frac{1}{N} \sum_{a=1}^N X_a \quad (1)$$

A_i = The new centroid point from the i-cluster

N = Number of data in the i-cluster

X_a = The first data on the I-cluster (Yi and Moon, 2013)

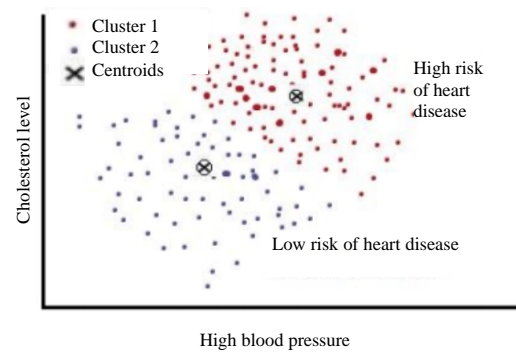


Fig 1: k-mean clustering for heart disease patient (Babu *et al.*, 2017)

The data used to diagnose heart disease can be in the body factors that are most likely to cause heart disease such as cholesterol levels and high blood pressure.

Figure 1 shows the results of k-mean clustering in the form of graphs showing 2 centroid points. The first point is low risk of heart disease and the second point is high risk of heart disease. Update the centroid point until there is no change then calculate the closest distance from a data to the nearest centroid point. The centroid point that has the highest data stickiness is the result of data processing.

Conclusion can also be done by looking for the value of closeness between training data with patient data that will be diagnosed using a database of previous patients that are perceived as having the same characteristics (Farias *et al.*, 2012).

Data mining in determining illnesses other than through symptoms can also be through examining other organs such as detecting gastric disorders through iris. Iris is chosen because it can show characteristics or signs that indicate consistency in a person's body. Iris can show someone has a stomach disorder from the color level surrounding it (Dewi *et al.*, 2016).

The detection process uses an image from the iris which will be processed further by turning it into a new image with a smaller pixel size, approximately 200×200 pixels (Dewi *et al.*, 2016).

In Fig. 2 the process of reading the iris is shown to determine the gastric condition. The process focuses on the color surrounding the iris that represents the condition of the stomach. In the final reading process, a blurred image is repaired with more detail at the color point around the iris. The results of the process will be used as input in data mining. In addition to detecting gastric disorders, this method can also be used to detect other interference by reading irises at different angles as input (Dewi *et al.*, 2016).

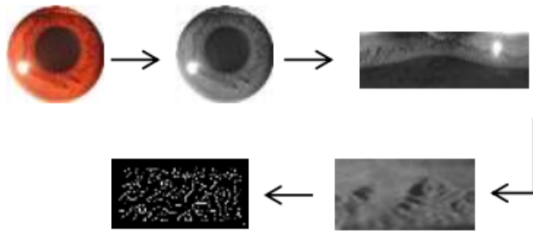


Fig. 2: Overview of the Iris reading process (Dewi *et al.*, 2016)

One example of another organ that can be seen as a health condition other than the stomach is the heart. The condition of the heart organ can be selected on the left iris from an angle of 60-90° (Permatasari *et al.*, 2016).

Health services: Many types of health services are often used by humans, ranging from hospital health services to health services on smartphones and some of them use data mining. For example, in hospital services, data mining helps analyze the costs of hospitalized patients to suit the financial considerations, types of illness and length of stay of patients.

Many personal health services have used data mining. Health controller application is one example that uses data mining. Data mining will look for patterns of exercise habits and the user's heart beat to determine what health care is the most suitable (Tseng *et al.*, 2017).

Research: Research continues to be done along with the increasing number of diseases not the least of them using data mining. Data mining can be used for drug development and studying side effects.

In one of the international conferences explaining that data mining can mean the pattern of writing on health forums for research. Data in the form of user Tweets about the type and treatment of an illness will be processed. Look for the pattern of the Tweets by breaking them into positive and negative opinions to find out the impact of using the drugs conveyed through the forum. Those opinions become the ratio of satisfaction to know which drugs are suitable for use. For the case of certain symptoms, data mining will take the opinion of a doctor or an online medical dictionary to find the appropriate medication (Sutar, 2017).

Entertainment: Data mining is widely used in entertainment. Examples: spotify and joox applications use data mining on the recommendation system based on the type of music that users often hear.

Prediction: In the entertainment industry especially in cinema, data mining can be used to estimated success of the film. Data mining helps analyzes the relationship between genre and film ratings, film actors and ratings and film genres and actors to predict the success of a film. The analysis was carried out with additional attributes, namely the year of release, film director and producer. Famous film directors and producers will be added value in the analysis (Ahmad *et al.*, 2017).

Science

Prediction: Data mining is most often used to predict a thing. Education often use data mining to predict grades or graduations. In universities, lecturers can use data mining to estimate the grades of their students by looking for patterns in the grades of previous courses. This pattern will predict courses that are suitable for students to choose along with their grades (Ramesh *et al.*, 2013).

In addition to estimate value, data mining is also used in the selection of new student admissions. Many universities use data mining for student admission selection. The reading pattern using data alumni and students who have been educated at least 4 years as a training data. The data of self and the grade of prospective student will be searched its value of proximity to training data. Prospective students who are deemed to have the greatest proximity value are likely to be accepted as university students.

Evaluation: Educational institutions can use data mining to evaluate their performance in improving the quality of their teaching. Using student data such as student behavior in the classroom or outside the classroom, liveliness class, parent participation, class attendance records, etc., become attributes used for reading patterns to determine whether students enter either, moderate or low grade. The difference in the class will be adjusted to the ability of students. The lessons in each class will be made in different ways (Rahman and Islam, 2017).

Agriculture: Many of the pertinent sectors use data mining to predict, for example, predicting pest or harvest attacks. Mining data analyzes patterns of factors that affect plants such as weather, soil conditions, water conditions, etc. (Table 1).

Maintenance: Data mining helps farmers or land preparers in maintenance to produce superior and quality yields. Maintenance can take the form of maintenance on the plant itself or maintenance of plant growth supporting

Table 1: Review and discussion

Implementation	Researcher's name	Discussion
Market basket	Trmka (2010)	Learn the pattern of interconnection between products for the arrangement of goods layout in supermarkets using analysis customer transaction databases to determine which product combinations are often purchased
Recommender	Wang <i>et al.</i> (2004)	The recommendation system is built with several techniques such as data filtering and mining by analyzing system customer interest in the product in the form of a recommendation score
Fraud detection	Cabral <i>et al.</i> (2008)	Using an anomaly detection technique to look for hidden patterns that start from reading the pattern as a whole if a pattern is found outside the habit then the pattern reading will be narrowed to prove whether it is a form of cheating
Diagnose the	Farias <i>et al.</i> (2012)	Taking conclusions to diagnose the disease can be done by reading the hidden patterns of patient data stored in the disease database such as age, sex, blood pressure and then looking for value of proximity to training data from previous patients to find the suitability of the type of disease.
Health service	Tseng <i>et al.</i> (2017)	Data mining helps users control their health through the application by translating patterns of relationship between exercise habits and heart rate to regulate suitable exercise
Research	Sutar (2017)	Tweet of users will be in the process of doing text mining to separate positive and negative opinions to know the side effects of the drug. Positive opinion will be the ratio of satisfaction to determine which drugs are suitable for use
Prediction	Ahmad <i>et al.</i> (2017)	To predict the success of the film data mining is done on a movie database containing genres, ratings, actors, release years, directors and film producers. Before the search process the pattern is carried out cleaning and integration first
Evaluation	Rahman and Islam (2017)	Evaluate learning by separating classes according to ability. Look for patterns by processing student's personal data such as behavior and absentee notes in class to determine suitable students placed in the appropriate class
Maintenance	Jia (2009)	Classification of agricultural land can be carried out by data mining by reading patterns on the types of soil and content that are in the soil like PH, etc., helping to assess the land

factors such as soil. Data mining helps to overcome the problem of lost land and land quality. Data mining performs pattern readings on soil types accompanied by content and PH on soil helps to know the soil quality (Jia, 2009).

CONCLUSION

In this study we discuss some of the uses and data mining applications in such areas as business, entertainment, health and agriculture. Where in each field, data mining can be on duty for various things. Data mining is most often used for recommendation and prediction systems. There are still many areas that use data mining that cannot be submitted in this study such as law, mining, social, entrepreneurship, etc.

REFERENCES

- Ahmad, J., P. Duraisamy, A. Yousef and B. Buckles, 2017. Movie success prediction using data mining. Proceedings of the 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), July 3-5, 2017, IEEE, Delhi, India, ISBN: 978-1-5090-3039-2, pp: 1-4.
- Babu, S., E.M. Vivek, K.P. Famina, K. Fida and P. Aswathi *et al.*, 2017. Heart disease diagnosis using data mining technique. Proceedings of the 2017 International Conference on Electronics, Communication and Aerospace Technology (ICECA) Vol. 1, April 20-22, 2017, IEEE, Coimbatore, India, ISBN: 978-1-5090-5685-9, pp: 750-753.
- Cabral, J.E., J.O. Pinto, E.M. Martins and A.M. Pinto, 2008. Fraud detection in high voltage electricity consumers using data mining. Proceedings of the 2008 International Conference on IEEE/PES Transmission and Distribution Conference and Exposition, April 21-24, 2008, IEEE, Chicago, Illinois, USA., ISBN: 978-1-4244-1903-6, pp: 1-5.
- Cao, L., P.S. Yu, C. Zhang and H. Zhang, 2008. Data Mining for Business Applications. Springer, Berlin, Germany., ISBN:978-0-387-79420-4, Pages: 299.
- Dewi, A.K., A. Novianty and T.W. Purboyo, 2016. Stomach disorder detection through the iris image using backpropagation neural network. Proceedings of the 2016 International Conference on Informatics and Computing (ICIC), October 28-29, 2016, IEEE, Mataram, Indonesia, ISBN: 978-1-5090-1649-5, pp: 192-197.
- Farias, F.D.S., L.V.D. Souza, R.C.M. Sousa, C.A.M. Caldas and L.F. Gomes *et al.*, 2012. Data mining applied to diagnose diseases caused by lymphotropic virus: A performance analysis. IEEE. Latin Am. Trans., 10: 1319-1323.
- Fazel, A., M. Saraee and P. Shamsinejad, 2010. Mining time series data: Case of predicting consumption patterns in steel industry. Proceedings of the 2010 2nd International Conference on Software Engineering and Data Mining (SEDM), June 23-25, 2010, IEEE, Chengdu, China, ISBN: 978-1-4244-7324-3, pp: 501-505.
- Gantz, J. and D. Reinsel, 2013. The digital universe in 2020: Big data, bigger digital shadows and biggest growth in the far East. IDC. Analyze Future, 2007: 1-16.

- Jia, Z., 2009. An expert system based on spatial data mining used decision tree for agriculture land grading. Proceedings of the 2009 2nd International Conference on Intelligent Computation Technology and Automation, ICICTA'09 Vol. 4, October 10-11, 2009, IEEE, Changsha, Hunan, China, ISBN: 978-0-7695-3804-4, pp: 142-145.
- Lawrence, R.D., G.S. Almasi, V. Kotlyar, M.S. Viveros and S. Duri, 2001. Personalization of supermarket product recommendations. *Data Mining Knowledge Discov.*, 5: 11-32.
- Permatasari, L.I., A. Novianty and T.W. Purboyo, 2016. Heart disorder detection based on computerized iridology using support vector machine. Proceedings of the 2016 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), September, 13-15, 2016, IEEE, Bandung, Indonesia, ISBN: 978-1-5090-0745-5, pp: 157-161.
- Rahman, M.H. and M.R. Islam, 2017. Predict student's academic performance and evaluate the impact of different attributes on the performance using data mining techniques. Proceedings of the 2017 2nd International Conference on Electrical and Electronic Engineering (ICEEE), December 27-29, 2017, IEEE, Rajshahi, Bangladesh, ISBN: 978-1-5386-3340-3, pp: 1-4.
- Ramesh, V., P. Parkavi and K. Ramar, 2013. Predicting student performance: A statistical and data mining approach. *Intl. J. Comput. Appl.*, 63: 35-39.
- Sutar, S.G., 2017. Intelligent data mining technique of social media for improving health care. Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), June 15-16, 2017, IEEE, Madurai, India, ISBN: 978-1-5386-3901-6, pp: 1356-1360.
- Trnka, A., 2010. Market basket analysis with data mining methods. Proceedings of the 2010 International Conference on Networking and Information Technology (ICNIT), June 11-12, 2010, IEEE, Manila, Philippines, ISBN: 978-1-4244-7579-7, pp: 446-450.
- Tseng, V.S., C.H. Chou, K.Q. Yang and J.C. Tseng, 2017. A big data analytical framework for sports behavior mining and personalized health services. Proceedings of the 2017 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), December 1-3, 2017, IEEE, Taipei, Taiwan., ISBN: 978-1-5386-4204-7, pp: 178-183.
- Wang, Y.F., Y.L. Chuang, M.H. Hsu and H.C. Keh, 2004. A personalized recommender system for the cosmetic business. *Exp. Syst. Appl.*, 26: 427-434.
- Yi, F. and I. Moon, 2013. Extended K-means algorithm. Proceedings of the 2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) Vol. 2, IEEE, Hangzhou, China, ISBN: 978-0-7695-5011-4, pp: 263-266-10.1109/IHMSC.2013.210.