Int. J. Nonlinear Anal. Appl.
Volume 12, Special Issue, Winter and Spring 2021, 2057-2066
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.6016

Big data implementation in Tesla using classification with rapid miner

Johanes Fernandes Andry[a], Julia Gunadi[a], Glisina Dwinoor Rembulan[b], Hendy Tannady[c,*]

[a] *Information Systems Department, Universitas Bunda Mulia, Jakarta, Indonesia*
[b] *Industrial Engineering Department, Universitas Bunda Mulia, Jakarta, Indonesia*
[c] *Management Department, Kalbis Institute, Jakarta, Indonesia*

(Communicated by Madjid Eshaghi Gordji)

Abstract

In this study, we will analyze how big data is implemented in TESLA Company, in this case, we will use sales data. With the growing of big data and the need for its use in the companies, nowadays big data is everywhere. TESLA is an American automobile and energy storage company founded by engineers Martin Eberhard and Marc Tarpenning in July 2003 under the name Tesla Motors. The company name is a tribute to inventor and electrical engineer Nikola Tesla. Eberhard said that he wanted to build an automobile manufacturer and also a technology company whose core technology is batteries, computer software and proprietary electric motors. As the amount of data that companies must process today continues to increase, companies must keep up with the times by using big data. Big data can be used to move, contain, and access large amounts of unstructured and disparate data in a timely manner. it is good. For the method we use is quantitative data. This calculation will use the Rapid Miner software. The result of this study is the data is 2,146 units, total volume from 118,500 to 47,065,000 based on the number of existing sales, and classification result are from 2621300 to 18766300.

Keywords: Big data, Classification, Data, TESLA.

1. Introduction

Tesla Motors, Inc., at the same time, it has established its own sales network, service center and supercharging station worldwide [17]. The purpose of this paper is to see the price and sales data

[*] Corresponding author

Email addresses: jandry@boundamulia.ac.id (Johanes Fernandes Andry), jgunadi@boundamulia.ac.id (Julia Gunadi), grembulan@boundamulia.ac.id (Glisina Dwinoor Rembulan), hendytannady@gmail.com (Hendy Tannady)

Received: September 2021 *Accepted:* November 2021

that occurs in the TESLA automotive company using the Classification method carried out with Rapid Miner.

Classification is one of data mining methods that the data could categorize by the class labels previously known [2]. With the advancement of IT that is happening in this era, amount of data continues to increase and the challenges that companies will face become more complex. Many companies invest in enabling them to make the best use of big data. Data considered as the source of life of decision-making and the raw material for information to be accounted for. Without high-quality providing data with the right information on the right things at the time, designing, monitoring and evaluating policies is almost impossible [20]. Because of that, a big company like TESLA need high quality of data to build an excellent decision-making system. Nowadays, the an atmosphere of decision-making environment is information quality. Management has very valuable customer transaction data that can be captured, processed, indexed, and stored so that later it can be used as very useful and up-to-date information [3].

With the growing development of big data and the need for its use in the companies, at this moment big data is everywhere and can be easy to manage. In the last few decades, there have been significant improvements in big data, for business analytics purposes, and in living and smart work environments. In recent years, people have paid attention to big data, business analysis and ”smart” living and working environment [10]. Organization are exploring large volume of data that maybe can be useful to create capture values for the organization, beside of that there’s one of the main challenges that the organization will be faced with when the author works with very large data, of course it is done for data management for various purposes, and the organization needs to access structured data. The integration of structured and unstructured data is regarded as data with fixed code meaning and format, most of which are in digital form and are usually used in the database field [7].

Big Data can be quickly added or subtracted into today’s popular business dictionaries; some, think that it is too clear to describe the large amount of data that users, or consumers, will do with the data already available [6]. According to some, big data has made many changes in decision making, especially in the field of accounting because it focuses on the right decision-making process and can explore an effective balance and high objectivity will later consider a large amount of data.

The amount and variety of data far exceed the capabilities of manual analysis, and in some cases have exceeded the capabilities of conventional databases [15]. At the same time computers have grown stronger with better networks and algorithms, able to combine data sets. So that the possibility of the previous analysis is still stored is very likely to happen. Big data that is used properly can be used to analyze further about customers, suppliers, and conditions within the company itself.

The development of technology is increasingly widespread and affects all areas of life. Technology has become an integral part of everyday human life. This can be demonstrated by the many innovations that have been carried out in this world. For the company, working in a turbulent environment leads to a quality decision-making process, become important element of management. informed consent for subsequent decision-making processes at all levels of the enterprise, and how it is organized, will be increasingly important. the magnitude of Big Data is not only to collect the data itself, but most importantly, and for its visualization, it is crucial to gain business advantage [8].

There are three characteristics of Big data named 3V; Volume, Variety, and Velocity. Volume itself is related to the size of the data; variety is related to the type of data and Velocity is related to the speed of data processing [1].

The presence of big data today has completely changed the work system in a company, where data is large and varied and must be accessed quickly. Create its own pressure when working with big data. With the development of computer capabilities related to big data, data can be fully used for analysis so that no data is wasted even though it is still usable. For example, in a recently released survey and research, half of the respondents gave an opinion about the importance of analytics in the companies they work for and less than 20% of respondents think that they are under pressure to improve business analytics [4]. This can raise questions about how companies actually use big data.

2. Literature Review

2.1. Big Data

Implementation of big data is very important for company, especially Tesla [11]. In this digital era, information and data is an important asset in making decisions. When you want to buy an item, it is likely that you will look for reviews related to the item to decide which product is worthy and worth buying. In everyday life, many people use available information and data for personal or group interests. In the business world, the amount of information needed to make the best decisions is enormous. This large collection of data and information is also called big data, to process it requires Information Technology (IT) [13].

2.2. Rapid Miner

Rapid miner studio application is used for data processing, existing data is taken from Kaggle data then processed by entering a dataset using the classification method [12].

2.3. Classification

The method Classification algorithm is a process that finds rules for partitioning or dividing data into separate groups. The rules for classification generated by the process can be used to classify various data in the future [16]. The classification method is used to predict or predict classes on certain data labels, with the aim of classifying data based on training data sets and values (class labels) in certain attributes and using them in classifying new data [14]. the most popular classification methods are Decision/Classification Trees, Bayesian Classifiers/Naïve Bayes Classifiers, Neural Networks, etc. [5].

3. Methodology

Research is a process where a person observes a phenomenon in depth and collects data to then draw conclusions from that data [9]. From the results of this opinion, the research method is one of the scientific ways used to obtain clear and valid data with the aim of being able to find, develop, and prove the clarity of certain knowledge so that in the end it can be used to understand, solve, and anticipate problems. That exists in a particular field. Goal of this research is to specify the number of sales from TESLA during a certain period using the Rapid Miner. In addition, this study uses qualitative methods. Qualitative research methods are research methods used to analyze data in the form of descriptions of data that cannot be quantified directly [19]. Some of the steps for research carried out in this scientific are:

- Data Selection: get data sources from multiple sources.
- Data preprocessing: is for cleaning unwanted data and looking for missing values and noisy data values from the dataset.
- Data Analyst: one of the data mining techniques is applied to get the results.
- Implementation: the result of data mining techniques applied in the Rapid Miner application.

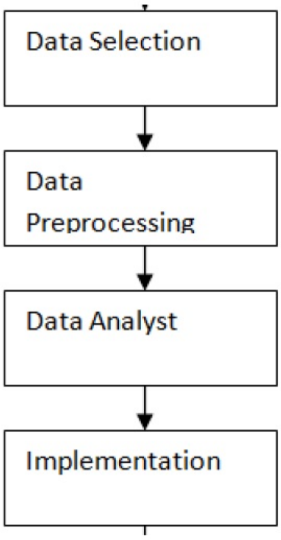


Figure 1: Research Framework

3.1. Research subject

The subjects in this study were TSLA sales data. In this study, the amount of sales data was 2,146 units. Data collection was done by looking for documentation data in Kaggel. The data in this study were taken from the Kaggle site under the name TSLA.

3.2. Instrument Development

Instrument development is very important in research activities. Research instruments have a very, very important contribution in determining the quality of a study, because the validity or clarity of the data obtained will be largely determined by the quality of the instruments used by the researcher, in addition to the data collection procedures used. In developing this instrument, it will be used in the research is a documentation guide. Documentation guidelines are tools used to collect data and documentation archives. Documentation using Kaggle data.

3.3. Data Analysis Technique

The data analysis technique used in this research is descriptive. There are two types of data in this action research, namely qualitative data and quantitative data. But what we use is quantitative. Quantitative data is in the form of numbers and is also the result of measurement and calculation. This calculation will use the Rapid Miner software, and also with the K-mean. The data will be compared using two classification algorithms from data mining. The proposed algorithm is a combination of K-Nearest Neighbor algorithm with Naive Bayes Classifier algorithm, then evaluate and validate the result with confusion matrix. The next stage is to compare the results of accuracy and time complexity of each algorithm, to obtain the model of the classification algorithm which obtains the highest accuracy and time complexity. Another approach can be done by creating a set of functions that can measure several properties of a grouping as a function of several grouping parameters [18].

4. Analysis and Discussion

This stage, we will use the Rapid Miner and Microsoft Office Excel 2019 software. A rapid Miner is a tool used to analyse data. The purpose of using Rapid Miner is to analyse the dataset that will

Table 1: TSLA

Date	Open	High	Low	Close	Adj Close	Volume
29/06/2010	19.000.000	25.000.000	17.540.001	23.889.999	23.889.999	18766300
30/06/2010	25.790.001	30.420.000	23.299.999	23.830.000	23.830.000	17187100
01/07/2010	25.000.000	25.920.000	20.270.000	21.959.999	21.959.999	8218800
02/07/2010	23.000.000	23.100.000	18.709.999	19.200.001	19.200.001	5139800
06/07/2010	20.000.000	20.000.000	15.830.000	16.110.001	16.110.001	6866900
07/07/2010	16.400.000	16.629.999	14.980.000	15.800.000	15.800.000	6921700
08/07/2010	16.139.999	17.520.000	15.570.000	17.459.999	17.459.999	7711400
09/07/2010	17.580.000	17.900.000	16.549.999	17.400.000	17.400.000	4050600
12/07/2010	17.950.001	18.070.000	17.000.000	17.049.999	17.049.999	2202500

be needed as needed. The purpose of this research is to find out alternatives and show the chosen decision. Moreover, the benefit is that we use the decision tree to find how many people are looking for their type. This research will follow the classification stages in this research stage.

Research Subject, the previous chapter, we know that research subject is an individual that participates in research. Information (or 'data') is collected from or about the individual to help answer the question under study. As mentioned in previous chapter, The subjects in this study were TSLA sales data. In this study, the amount of sales data was 2,146 units. Data collection was done by looking for documentation data in Kaggel. The data in this study were taken from the Kaggle site under the name TSLA.

Instrument development, as discussed in the previous chapter, research instruments are indispensable in carrying out this research. Validity or clarity is mostly determined by the quality of the data instruments used by the researcher. In this study, the authors will use data documentation taken from Kaggle.

Data analysis technique, authors already know that data analysis is defined as a process of to cleaned, to transformed, and to model data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and making a decision based upon the data analysis. . There are two types of data in this action research, namely qualitative data and quantitative data. But what we use is quantitative. Quantitative data is in the form of numbers and is also the result of measurement and calculation. This calculation will use the Rapid Miner software, and also with the K-mean. K-Means Classification is an algorithm in data mining that can be used to classify data.

4.1. Data Source

The source for the primary data used in this study is the TSLA dataset in .xlsx format. The data dataset used in this study is Date, Open, High, Low, Close, Adjust Close, and Volume. The amount of data is 2146 data, and there are seven attributes.

The unique attribute is the volume type. To find data collection to become data mining, we enter data into Rapid Miner. How to enter it by using Retrieve. The purpose of using Retrieve is to block the mode used to access data. The data that will appear as in table 1.

Then, we need the raw data that we have, which aims to display from that data that we have. For the next part, we see the statistic table, where in the table we see how many statistics are on each attribute. In Statistics, there are some row like name, type, missing, and statistics. It will be better if there is no missing data, because missing data can make the analysis complicated. Also, there are types with several types: integer, polynomial, binominal, and many others. And for the

Table 2: Statistics

Name	Type	Missing	Min	Max	Average
Volume	Integer	0	118500	47065000	5572721.689
Date	Date	0	Earliest, Jun 29,2010	Latest, Feb 3,2020	Duration: 3506 Days
Open	Polynominal	0	Least, 99.000.000	Most, 28.000.000	Values: 28.000.000[2130 more]
Close	Polynominal	0	Least, 99.550.003	Most, 27.420.000	Values: 27.420.000[2223 more]
Adj Close	Polynominal	0	Least, 99.550.003	Most, 27.420.000	Values: 27.420.000[2223 more]
High	Polynominal	0	Least, 99.279.999	Most, 28.000.000	Values: 28.000.000[2126 more]
Low	Polynominal	0	Least, 99.499.997	Most, 27.299.999	Values: 27.299.999[2134 more]

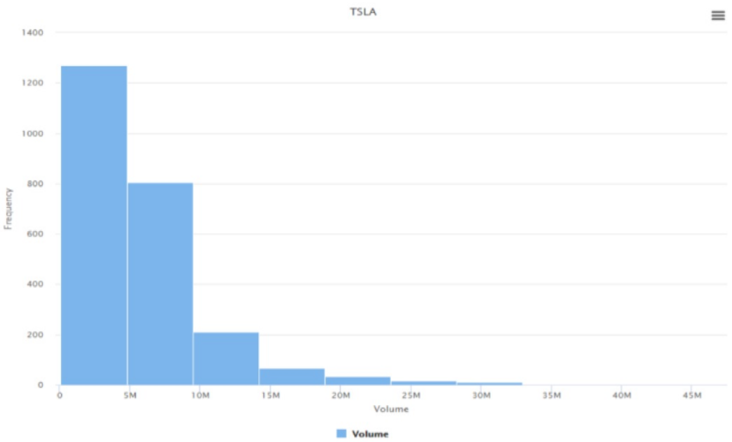


Figure 2: Volume Graph

last step, statistics are useful for knowing the most, the smallest, the least, and the average amount of each data category. For the following statistic data, see table 2.

The jam is contained with the name visualization which aims to bring up the graphics in each attribute. We can choose the graphics according to what we want ourselves. The following is an example of a graphic image 2:

In Figure 2. Explain the total volume from 118,500 to 47,065,000 based on the number of existing sales. From that, the largest data is 1,269 volumes and the smallest is 0 volumes. From the graphic data, it can show that the graph from 0 is decreasing the sales figure of an item. From the results of the graph, 0-5m has a volume of 1,269 volumes, then 5m-1.m has a volume of 804 volumes, 10m-15m has a volume of 210 volumes, 15m-20m has a volume of 67 volumes, then 20m-25m has a volume of 17 , 25m-30m has a volume of 12 volumes.

4.2. Classification

The purpose of using classification is to predict the class from the given TSLA dataset. The overall process that is made will be like the following picture 3:

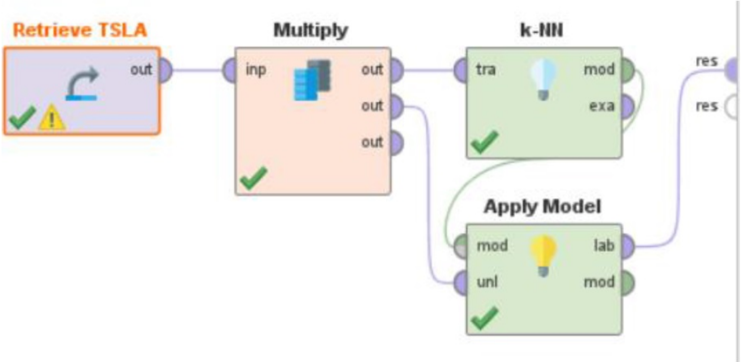
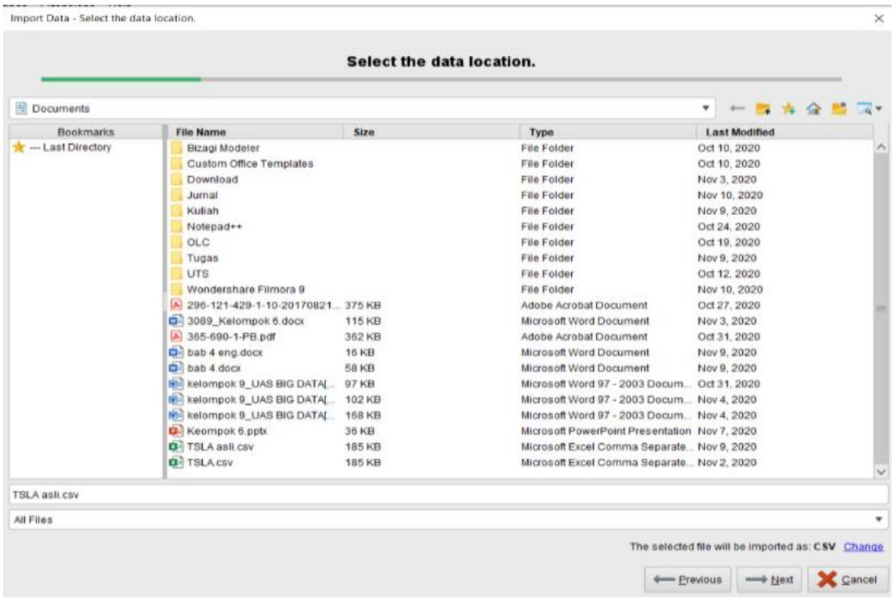


Figure 3: Process of Apply Model for Classification



Select the data location.

File Name	Size	Type	Last Modified
Bizagi Modeler		File Folder	Oct 10, 2020
Custom Office Templates		File Folder	Oct 10, 2020
Download		File Folder	Nov 3, 2020
Jurnal		File Folder	Nov 10, 2020
Kuliah		File Folder	Nov 9, 2020
Notepad++		File Folder	Oct 24, 2020
OLC		File Folder	Oct 19, 2020
Tugas		File Folder	Nov 9, 2020
UTS		File Folder	Oct 12, 2020
Wondershare Filmora 9		File Folder	Nov 10, 2020
296-121-429-1-10-20170821...	375 KB	Adobe Acrobat Document	Oct 27, 2020
3089_Kelompok 6.docx	115 KB	Microsoft Word Document	Nov 3, 2020
365-690-1-PB.pdf	362 KB	Adobe Acrobat Document	Oct 31, 2020
bab 4 eng.docx	16 KB	Microsoft Word Document	Nov 9, 2020
bab 4.docx	58 KB	Microsoft Word Document	Nov 9, 2020
kelompok 9_UAS BIG DATA(...	97 KB	Microsoft Word 97 - 2003 Docum...	Oct 31, 2020
kelompok 9_UAS BIG DATA(...	102 KB	Microsoft Word 97 - 2003 Docum...	Nov 4, 2020
kelompok 9_UAS BIG DATA(...	168 KB	Microsoft Word 97 - 2003 Docum...	Nov 4, 2020
Keompok 6.pptx	36 KB	Microsoft PowerPoint Presentation	Nov 7, 2020
TSLA asli.csv	185 KB	Microsoft Excel Comma Separate...	Nov 9, 2020
TSLA.csv	185 KB	Microsoft Excel Comma Separate...	Nov 2, 2020

TSLA asli.csv

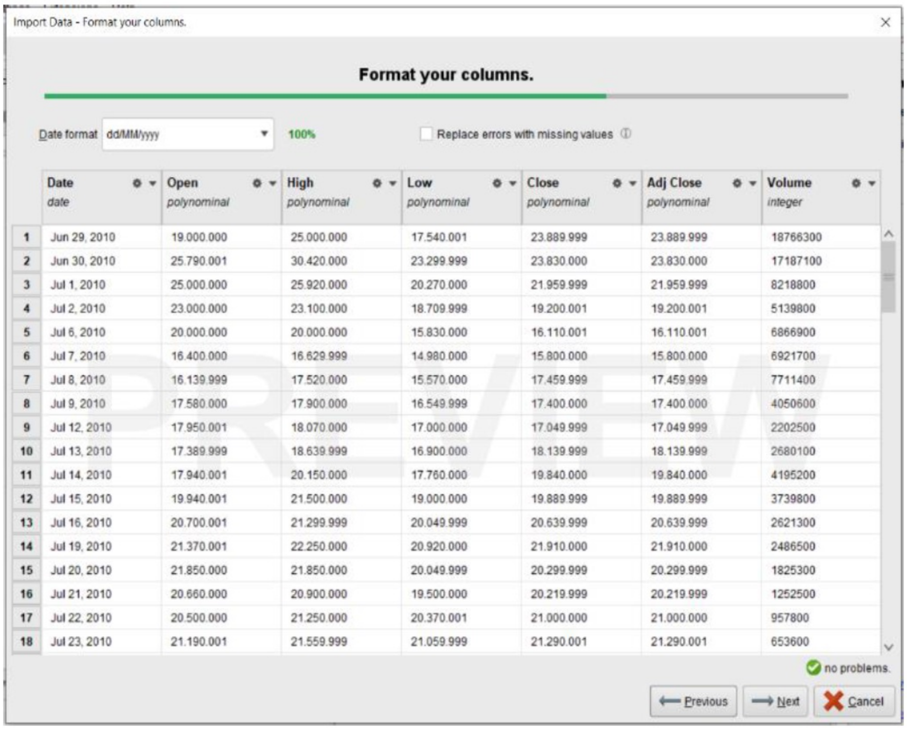
The selected file will be imported as: CSV Change

Figure 4: Import Data- Select The Data Location

In Figure 3, there are 4 classification processes used for this classification. The operators used are read data, multiply, k-NN, and Apply Model. Read Data in Rapid Miner supports a lot of formats such as Ms. Excel, CSV, SPSS, Ms. Access and others. In this research, we will use the TSLA dataset. The TSLA dataset we use comes from the Kaggle data. To enter Ms. data. Excel or the like into rapid miner, that is, we have to open the RapidMiner application and after that we go straight to Import data. After that it will appear like this 4:

In Figure 4, we are told to choose the data we have. In this research, because we are using TSLA data, then we click on TSLA.csv data. After that, click next to the next process, until it will appear as shown below 5:

In figure 5, we will focus on the column format of our read data. In the image, we will change the column to the attributes we have. There will be various things in this format, such as change types, change roles, rename columns, and exclude columns. There are several types of change types, such as date, polynomial, binary, integer, etc. Then there is the role change, which aims to become a special attribute. There are three special attributes, namely, weight, label, and id. Then there is the rename column, which aims to change the column name, and finally, the exclude column aims to delete a column. Columns that have been deleted will appear again if we click the include column.



Date format dd/MM/yyyy

□ Replace errors with missing values

	Date date	Open polynominal	High polynominal	Low polynominal	Close polynominal	Adj Close polynominal	Volume integer
1	Jun 29, 2010	19.000.000	25.000.000	17.540.001	23.889.999	23.889.999	18766300
2	Jun 30, 2010	25.790.001	30.420.000	23.299.999	23.830.000	23.830.000	17187100
3	Jul 1, 2010	25.000.000	25.920.000	20.270.000	21.959.999	21.959.999	8218800
4	Jul 2, 2010	23.000.000	23.100.000	18.709.999	19.200.001	19.200.001	5139800
5	Jul 6, 2010	20.000.000	20.000.000	15.830.000	16.110.001	16.110.001	6866900
6	Jul 7, 2010	16.400.000	16.629.999	14.980.000	15.800.000	15.800.000	6921700
7	Jul 8, 2010	16.139.999	17.520.000	15.570.000	17.459.999	17.459.999	7711400
8	Jul 9, 2010	17.580.000	17.900.000	16.549.999	17.400.000	17.400.000	4050600
9	Jul 12, 2010	17.950.001	18.070.000	17.000.000	17.049.999	17.049.999	2202500
10	Jul 13, 2010	17.389.999	18.639.999	16.900.000	18.139.999	18.139.999	2680100
11	Jul 14, 2010	17.940.001	20.150.000	17.760.000	19.840.000	19.840.000	4195200
12	Jul 15, 2010	19.940.001	21.500.000	19.000.000	19.889.999	19.889.999	3739800
13	Jul 16, 2010	20.700.001	21.299.999	20.049.999	20.639.999	20.639.999	2621300
14	Jul 19, 2010	21.370.001	22.250.000	20.920.000	21.910.000	21.910.000	2486500
15	Jul 20, 2010	21.850.000	21.850.000	20.049.999	20.299.999	20.299.999	1825300
16	Jul 21, 2010	20.660.000	20.900.000	19.500.000	20.219.999	20.219.999	1252500
17	Jul 22, 2010	20.500.000	21.250.000	20.370.001	21.000.000	21.000.000	957800
18	Jul 23, 2010	21.190.001	21.559.999	21.059.999	21.290.001	21.290.001	653600

Figure 5: Import Data-Format Your Columns

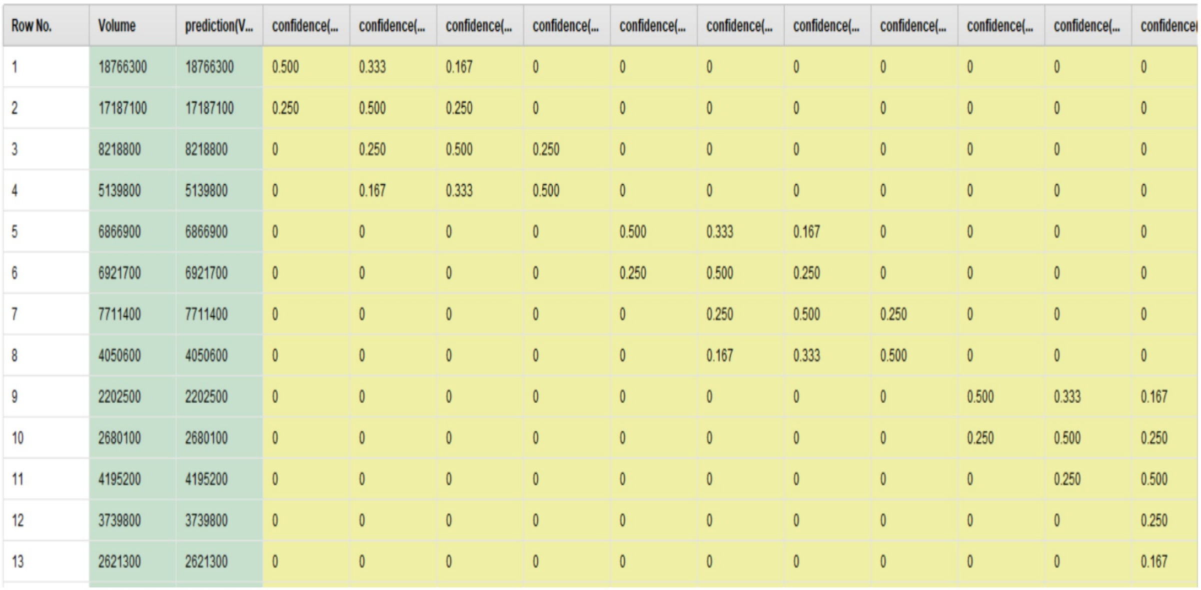
After that, click next to show the last stage, namely save data. Save data can be done anywhere.

Then there is the Multiply operator, which is used to make Copies of objects in Rapidminer. The copy itself will make it unrelated, so it has no effect on another copy when changing one copy. Then there is the k-NN operator to produce a model based on the k-Nearest Neighbor algorithm. The goal is to use the k-Nearest Neighbor algorithm for classification and regression. The k-NN Algoritman based on the distance about three close as a predictive value of the new instance. In this study, the value at k was determined to be $k = 3$. And finally, the Apply Model operator is used to apply a previously trained model using data raining on unlabeled data. The purpose of using the Apply Model is to obtain predictions on unlabeled data that do not yet have a label. What is needed when the Apply Model is data testing. The testing data must have the same sequence, type, and attribute prean as the data. The following are the results of Classification 6:

In the Figure 6, there are 2 colors, namely green and yellow. The green one is the volume in the TSLA read data that we use in this research. The reason the volume can turn green is because the volume when inputted the change role data is changed to a label. In change roles, there are 3 types, namely label, id, and weight. Because the volume has a special attribute so the color will be green. While the yellow color is the result of the Apply Model. Meanwhile, the white one is our read data that we use today.

5. Conclusion

The presence of big data today has completely changed the work system in a company, where data is large and varied and must be accessed quickly. the implementation of big data for the case at the Tesla company was obtained from this Kaggle data by using the classification method using a text mining processing application, namely Rapid Miner Studio, the results of the study are as follows is 2,146 units, total volume from 118,500 to 47,065,000 based on the number of existing sales,



Row No.	Volume	prediction(V...	confidence(...	confidence(...	confidence(...	confidence(...	confidence(...	confidence(...	confidence(...	confidence(...	confidence(...	confidence(...	confidence
1	18766300	18766300	0.500	0.333	0.167	0	0	0	0	0	0	0	0
2	17187100	17187100	0.250	0.500	0.250	0	0	0	0	0	0	0	0
3	8218800	8218800	0	0.250	0.500	0.250	0	0	0	0	0	0	0
4	5139800	5139800	0	0.167	0.333	0.500	0	0	0	0	0	0	0
5	6866900	6866900	0	0	0	0	0.500	0.333	0.167	0	0	0	0
6	6921700	6921700	0	0	0	0	0.250	0.500	0.250	0	0	0	0
7	7711400	7711400	0	0	0	0	0	0.250	0.500	0.250	0	0	0
8	4050600	4050600	0	0	0	0	0	0.167	0.333	0.500	0	0	0
9	2202500	2202500	0	0	0	0	0	0	0	0	0.500	0.333	0.167
10	2680100	2680100	0	0	0	0	0	0	0	0	0.250	0.500	0.250
11	4195200	4195200	0	0	0	0	0	0	0	0	0	0.250	0.500
12	3739800	3739800	0	0	0	0	0	0	0	0	0	0	0.250
13	2621300	2621300	0	0	0	0	0	0	0	0	0	0	0.167

Figure 6: Classfication results

and classification result are from 2621300 to 18766300.

References

[1] N.E.V. Anna and E.F. Mannan, *Big data adoption in academic libraries: a literature review*, Libr. Hi Tech. News 37(4) (2020) 1–5.

[2] J. Arunadevi, S. Ramya and M.R. Raja, *A study of classification algorithms using Rapidminer*, Int. J. Pure Appl. Math. 119(12) (2018) 15977–15988.

[3] N. Bharadwaj, *Strategic decision making in an information-rich environment: A synthesis and an organizing framework for innovation research*, Rev. Mark. Res. 15 (2018) 3–30.

[4] K. Ebner, T. Bühnen and N. Urbach, *Think big with big data: Identifying suitable big data strategies in corporate environments*, Proc. Annu. Hawaii Int. Conf. Syst. Sci. (2014) 3748–3757.

[5] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*, Springer Science & Business Media, 2011.

[6] S. Green, J.E. McKinney, K. Heppard and L. Garcia, *Article information: Big data, digital demand, and decision-making*, Int. J. Account. Inf. Manag. 26(4) (2017) 541–555.

[7] A. Intezari and S. Gressel, *Information and reformation in KM systems: big data and strategic decision-making*, J. Knowl. Manag. 21(1) (2017) 71–91.

[8] H. Koŝcielniak and A. Puto, *Big data in decision making processes of enterprises*, Procedia Comput. Sci. 65(Iccmit) (2015) 1052–1058.

[9] T. Loya and G. Carden, *Higher Education Strategy and Planning: A Profesional Guide*, Business Intelligence and Analytics, 2018.

[10] E.D. Madyatmadja, L. Liliana, J.F. Andry and H. Tannady, *Risk analysis of human resource information systems using Cobit 5*, J. Theor. Appl. Inf. Technol. 98(21) (2020) 3357–3367.

[11] E.D. Madyatmadja, M. Marvel, J.F. Andry, H. Tannady and A. Chakir, *Implementation of big data in hospital using cluster analytics*, Int. Conf. Inf. Manag. Technol. (2021) 496–500.

[12] E.D. Madyatmadja, A. Rianto, J.F. Andry, H. Tannady and A. Chakir, *Analysis of big data in healthcare using decision tree algorithm*, Proc. 2021 1st Int. Conf. Comput. Sci. Artific. Intell. (2021) 313–317.

[13] E.D. Madyatmadja, D.J.M. Sembiring, S.M.B.P. Angin, D. Ferdy and J.F. Andry, *Big data in educational institutions using RapidMiner to predict learning effectiveness*, J. Comput. Sci. 17(4) (2021) 403–413.

[14] S. Neelamegam and E. Ramaraj, *Classification algorithm in data mining: An overview*, Int. J. P2P Netw. Trends Technol. 4(8) (2013) 369–374.

[15] F. Provost and T. Fawcett, *Data science and its relationship to big data and data-driven decision making*, Big Data 1(1) (2013) 51–59.

[16] V. Rajeswari and K. Arunesh, *Analysing soil data using data mining classification techniques*, Indian J. Sci. Technol. 9(19) (2016) 1–4.

[17] L. Ruoxi, W. Hao, X. Jiayi, X. Xiaowei and X. Yanting, *Capital structure analysis of Tesla Motors, Inc*, Department of Economics and Finance City University of Hong Kong, 2014.

[18] Y.F. Safri, R. Arifudin and M.A. Muslim, *K-nearest neighbor and naive bayes classifier algorithm in determining the classification of healthy card Indonesia giving to the poor*, Sci. J. Inf. 5(1) (2018).

[19] P.K. Sari and A. Purwadinata, *Analysis characteristics of car sales in E-commerce data using clustering model*, J. Data Sci. Appl. 2(1) (2019) 68–77.

[20] K. Vassakis, E. Petrakis and I. Kopanakis, *Mobile Big Data, A Roadmap from Models to Technologies*, Springer, Cham, 2018.