Plagiarism Checker X Originality Report

Similarity Found: 26%

Date: Saturday, November 23, 2019
Statistics: 440 words Plagiarized / 1725 Total words
Remarks: Medium Plagiarism Detected - Your Document needs Selective Improvement.

Eurasian Journal of Analytical Chemistry ISSN: 1306-3057 OPEN ACCESS 2018 13 (6): 186-191 INTRODUCTION the lifetime of the Prophet ( ). Bustami, Doctoral Student of Mathematics and Applied Science, Syiah Kuala University . Taufik Fuadi Abidin, Department of Informatics, Faculty of Mathematics and Science, Syiah Kuala University .

Khairul Munadi, Department of Electrical Engineering, Faculty of Engineering, Syiah Kuala University . Zainal Arifin Hasibuan, Department of Information Technology, Faculty of Computer Science, University of Indonesia. Muhammad Fikry, Department of Informatics, Faculty of Engineering, Universitas Malikussaleh.

Classification of Hadith Levels Using Data and Text Mining Techniques Bustami, TaufikFuadi Abidin, Khairul Munadi, Zainal Arifin Hasibuan, Muhammad Fikry 187 Bustami et.al RELATED WORK In this section, we present important studies that have been carried out in hadith data mining ana lysis, decision tree (C4.5) and naïve bayes algorithms. The reason for this study was due to the lack of complexity of analysis in determining the level of hadith.

Naïve Bayes The Naïve Bayes algorithm is a classification method using probability and statistical methods. The Naïve Bayes algorithm predicts opportunities based on existing data so that it is known as the Bayes Theorem. Bayes's theorem is used to calculate posterior probabil ities, P (c | x), from P (c), P (x), and P (x | c).

The classifier carried out by Naïve Bayes is by analyzing the effect of the (x) predictor value on a particular class (c) does not depend on other predictor values. This analogy is also called the conditi onal dependency class. • P (c | x) is a posterior probability (target) which is given a predictor value (attribute). • P(c) is the prior probability of class. • P(x|c)

is the probability of predictor probability which is then given a class. • P(x) is the prior probability of predictor.

Numerical variables must be transformed into existing categories before it forms its own frequency table. The other option is to distribute numeric variables to estimate the frequency. An example is to assume a normal distribution for numeric variables. The probability density function is for a normal distribution which is then defined by two parameters (mean and standard deviation).

Profit information as the amount of information the district by each attribute can provide an ex planation of how predictor values affect the probability of class. Decision Tree (C4.5) C4.5 data mining algorithm is one of the algorithms used to classify or segment or classify and be predictive. Selection of good attributes is an attribute that allows getting the smallest decision tree size. Or attributes that can separate objects according to their class.

Heuristically the attribute selected is the attribute that produces the most "purest" node. The size of the purity is expressed by the level o f impurity, and to calculate it, can be done using the concept of Entropy, Entropy states the impurity of a collection of objects.

188 Eurasian Journal of Analytical Chemistry • S is a case set • k is the number of partitions S • pj is the probability obtained from Sum (Yes) divided by the Total Case Information gain is the most popular criterion for attribute selection. The C4.5 algorithm is the development of the ID3 algorithm. Because of this development, the C4.5 algorithm has the same basic working principles as the ID3 algorithm. It's just that in t he C4.5

algorithm the attribute selection is done by using Gain Ratio with the formula: • a = attribute • gain (a) = information gain on attribute a • Split (a) = split information in attribute a Attributes with the highest Gain Ratio value are selected as tes t attributes for vertices. With gain is information gain. This approach applies normalization to information gain by using what is called split information.

SplitInfo states entropy or potential information with the formula: • S = space (data) sample used for training • A = attribute • Si = number of samples for attributes i where Xi represents the i subset in sample X. S = space (data) sample used for training. A = attribute. |Si| = the number of samples for the value of V. |S| = number of all data samples.

Entropy (Si) = entropy for samples that have a value of i Data and Text Mining The

difference between data mining, association rule mining, and text mining is that in text mining, patterns are extracted from natural language text, but data mining and association rule mining patterns are extracted from databases. The steps in the mining process are (Chiwara): • Text: This represents the given target document for mining in text format.

• Text processing: This step concern to text clean up, format, tokenize and others. • Text transformation (attribute generation): Generates attributes from text that has be en processed based on the text provided • Attribute selection: Select attributes for mining data because not all attributes produced will be suitable for mining • Data Mining (Pattern discovery): Mine the selected attribute and then extract it according to the desired pattern. • Interpretation and evaluation: This is about what you are looking for in the next step, i.e.

terminate, results that are perfect for the application that you want and so on. Many people do not know the difference between data mining and k nowledge discovery, others think that data mining is the main stage of the Knowledge Discovery in Database (KDD) process. KDD is based on the whole process that extracts useful knowledge from the amount of data available.

Including evaluations to make deci sions about what is a requirement to become a knowledge. Whereas data mining refers to the application of algorithms to extract patterns from the data. KDD's steps are as follows, • Data Cleaning: Removing noise or outliers which interferes with data retriev aland inconsistent.

• Data Integration: Combining data from considerable data sources • Data Selection: Relevant data by processing data from existing databases. 189 Bustami et.al • Data Transformation: Transforming or entering data into desired forms for data mining by conducting operations or aggregation. • Data Mining: Applying intelligent methods to extract data patterns that it have. • Pattern Evaluation: Evaluating data patternswich exists.

• Knowledge Presentation: Representing knowledge collected. RESULTS AND DISCUSSIONS In this study, an application program was made using PHP's Promrograming language with the SQL database. In the application created, each algorithm has been included in the application.

Matan: :(?? ??) Translate in Indonesian: Dari Ali Radhiayallahu ‘Anhu, Rasulullah Shallallahu ‘Alaihi Wasallam bersabda: “Janganlah kamu mendahulukan orang-orang yang bodoh dari kamu (untuk menjadi iman) dalam shalat, juga janganlah

mendahulukan untuk shalat atas jenazahmu sesungguhnya mereka adalah utusanmu untuk Rabbmu." Table 1: The obtained results from Naïve Bayes experiments From the Naïve Bayes experiment table on the dataset, the highest accuracy is 94.5%, the lowest accuracy level is 38.48%, and the average accuracy is 64.91%. Results of the Decision Tree Experiment (C4.5) In the C4.5

algorithm experiment, it is done by entering the same keyword as the Naive Bayes algorithm experiment. The following are the experimental results of the datasets that have been obtained using the C4.5 algorithm. From the C4.5 experiment table on the dataset, the highest accuracy is 95.01%, the lowest accuracy level is 53.03%, and the average accuracy is 72.73%.

190 Eurasian Journal of Analytical Chemistry Validation Validation is an action that proves that a process/method can provide consistent results in accordance with established specifications and is well documented. After experimenting on the Naive Bayes and C4.5 algorithms, the following algorithm is obtained: Fig .1: Comparison Chart CONCLUSION The experimental results of the Decision Tree algorithm (C4.5), resulted in an accuracy rate of 7.81% greater than that of Naïve Bayes.

From the results of the study, it can be concluded that the C4.5 algorithm has a better level of accuracy compared to Naïve Bayes for the classification of hadith levels. However, it is recommended for further research to experiment with other models so that a better level of accuracy can be obtained. REFERENCES [1] Brahimi, B.,

Touahria, M., & Tari, A. (2016). Data and Text Mining Techniques for Classifying Arabic Tweet Polarity. Journal of Digital Information Management, 14(1). [2] Falotico, R., Liberati, C., & Zappa, P. (2015). Identifying Oncological Patient Information Needs to Improve e -Health Communication: a preliminary text- mining analysis .

Quality and Reliability Engineering International, 31(7), 1115-1126. [3] Fikry, M. (2016). Rancangan basis data kependudukan berdasarkan aspek- aspek kualitas schema database. TECHSI-Jurnal Teknik Informatika, 8(2). [4] Fikry, M. (2016). Aplikasi Java Kriptografi Menggunakan Algoritma Vigenere. Journal Techsi, 8(1), 1-9. [5] Khashfeh, M., Mahmoud, M. a., & Ahmad, M. S. (2018).

An analysis of text mining factors enhancing the identification of relevant studies . Journal of Theoretical & Applied Information Technology, 96(12). [6] Kim, Y., & Jeong, S. R. (2018). Competitive intelligence in Korean Ramen Market using Text Mining and Sentiment Analysis. ?????????? , 19(1), 155-166.. [7] Kinne, J., & Axenbeck, J. (2018). Web mining of firm websites: A framework for web scraping and a pilot study for Germany.

ZEW-Centre for European Economic Research Discussion Paper. [8] Philips, A. A. B. (2005). Asal-Usul dan Perkembangan Fiqh: Analisis Historis Atas Mazhab, Doktrin dan Kontribusi, terj. M. Fauzi Arifin, Bandung: Nuansa. [9] Philips, Abu Ame enah Bilal. Usool Hadith: The Methodology of Hadith Evaluation. IslamKotob , 1990. [10] Rosandy, T. (2017).

Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4. 5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus: Kspps/Bmt Al -fadhila. Jurnal Teknologi Informasi Magister, 2(01), 52-62. [11] Saloot, M. A., Idris, N., Mahmud, R., Ja'afar, S., Thorleuchter, D., & Gani, A. (2016). Hadith data mining and classification: a comparative analysis. Artificial Intelligence Review, 46(1), 113-128. [12] Sarma, G. (2017).

Scientific literature text mining and the case for Open Access. The Journal of Open Engineering. [13] Somantri, O. (2017). Text Mining Untuk Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes (NB). Jurnal Telematika, 12(1), 7-12. 60 65 70 75 NB C4.5 Comparison Chart 191 Bustami et.al [14] Singh, R., Rani, A., Kumar, P., Singh, C., Shukla, G., & Kumar, A. Hemicellulolytic Activity in the Crop Residues.

International journal of pharmacy research & technology, 7(1), 18-20. [15] Mathumathi, K.M. and Senthilprakash, K. (2017) energy sentient qos implemented node -disjoint multipath routing protocolfor manet. International journal of communication and computer technologies, 5 (2), 67-75. [16] Surendar, A. (2018). Role of Microbiology in the Pharmaceutical & Medical Device.

International Journal of Pharmaceutical Research, 10(3).

INTERNET SOURCES:

12% - http://www.eurasianjournals.com/pdf-104473-36951?filename=Classification%20of%20Hadith.pdf
1% - http://www.eurasianjournals.com/Classification-of-Hadith-Levels-Using-Data-and-Text-Mining-Techniques,104473,0,2.html
1% - http://jurnal.unsyiah.ac.id/index.php/JDA/about/editorialTeam
1% - https://unsyiah.ac.id/faculty-and-staff?Fsd_page=44&Fsd_sort=unit_kerja.desc
<1% - https://www.researchgate.net/profile/Yugo_Isal
<1% - https://acadgild.com/blog/naive-bayesian-model
<1% - https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

<1% - http://journals.sagepub.com/doi/pdf/10.1177/0013164410388832
<1% - http://wjert.org/download/article/15062016/1467339706.pdf
<1% -
https://www.researchgate.net/publication/221442285_Modeling_and_Inference_with_Relational_Dynamic_Bayesian_Networks
1% -
http://ijsetr.org/wp-content/uploads/2016/04/IJSETR-VOL-5-ISSUE-4-1009-1014.pdf
<1% - https://www.ischool.utexas.edu/~yanz/Content_analysis.pdf
<1% - https://rcatlord.github.io/GSinR/modules/model/exercises/model_data.html
<1% -
https://intellipaat.com/blog/tutorial/statistics-and-probability-tutorial/the-normal-distribution/
<1% - https://quizlet.com/318506474/i-flash-cards/
<1% - https://data-flair.training/blogs/data-mining-algorithms/
<1% -
http://assets.cambridge.org/97805216/80547/excerpt/9780521680547_excerpt.pdf
<1% - https://quizlet.com/174221425/mining-and-warehousing-flash-cards/
<1% - http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf
<1% - https://javadevnotes.com/java-array-of-objects
<1% - https://arxiv.org/pdf/1802.04109.pdf
<1% -
https://www.quora.com/In-a-decision-tree-how-do-we-select-which-attribute-to-split-the-data-by-at-each-level
<1% - https://www.sciencedirect.com/science/article/pii/S1568494612002864
<1% -
https://www.researchgate.net/publication/318657266_Data_Mining_Approach_for_Breast_Cancer_Patient_Recovery
<1% -
https://www.researchgate.net/publication/224585894_An_improved_ID3_decision_tree_algorithm
<1% - https://www.hindawi.com/journals/mpe/2019/3526539/
<1% - https://www.investopedia.com/terms/s/sample.asp
<1% - http://web.cs.wpi.edu/~cs4445/b12/Exams/topics_exams.html
1% - https://www.researchgate.net/post/Text_mining_and_Data_mining
<1% - https://github.com/jlroo/data-mining/wiki/DOCUMENTATION
<1% -
https://rstudio-pubs-static.s3.amazonaws.com/386722_522e84e1ff264537ace05ad2a1cdc009.html
<1% - https://siddhi.io/en/v5.0/docs/query-guide/
<1% - https://www.sciencedirect.com/science/article/pii/S221083271400026X

<1% - https://www.skillsyouneed.com/learn/critical-reading.html
<1% - https://stackoverflow.com/q/2820357
<1% - https://en.wikipedia.org/wiki/Data_mining
<1% - https://www.witpress.com/Secure/elibrary/papers/DATA02/DATA02015FU.pdf
<1% - https://www.effectivegovernance.com.au/wp-content/uploads/2012/10/Seven-steps-to-effective-board-and-director-evaluations-Beck-and-Kiel.pdf
<1% - http://www.scielo.br/pdf/jistm/v11n2/1807-1775-jistem-11-2-0379.pdf
<1% - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4503556/
<1% - https://www.hindawi.com/journals/sp/2017/1901876/
<1% - https://www.researchgate.net/publication/319404250_A_Text_Mining_Based_Literature_Analysis_for_Learning_Theories_and_Computer_Science_Education
<1% - https://www.researchgate.net/profile/Mohammad_Sadegh_Aslanpour/publication/318341798_Proactive_Auto-Scaling_Algorithm_PASA_for_Cloud_Application/links/5b4cf1e90f7e9b240fe4f302/Proactive-Auto-Scaling-Algorithm-PASA-for-Cloud-Application.pdf
<1% - https://imaduddien-matahati.blogspot.com/2010/01/
<1% - https://csce.ucmss.com/cr/books/2018/LFS/CSREA2018/BIC3404.pdf
<1% - https://www.researchgate.net/publication/323712169_Very_Fast_C45_Decision_Tree_Algorithm
<1% - http://www.eurasianjournals.com/pdf-104473-35468?filename=Classification%20of%20Hadith.pdf
<1% - http://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/1002
<1% - http://www.iaeng.org/publication/WCECS2010/WCECS2010_pp454-459.pdf
<1% - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4453835/
<1% - https://link.springer.com/chapter/10.1007%2F978-3-319-67056-0_20
<1% - https://www.sciencedirect.com/science/article/pii/S1361841508000224
1% - https://onlinelibrary.wiley.com/doi/full/10.1111/jpim.12233
1% - https://ojs.unimal.ac.id/index.php/techsi/article/view/136
1% - http://www.jatit.org/volumes/Vol96No12/24Vol96No12.pdf
<1% - http://ftp.zew.de/pub/zew-docs/dp/dp14049.pdf
<1% - https://dinus.ac.id/detailstaf/pertemuan%209
<1% - https://link.springer.com/article/10.1007%2Fs10462-019-09692-w
<1% - https://www.sciencedirect.com/science/article/pii/S0740002017310638
<1% - https://scicurve.com/author/Singh%20A
<1% - https://onlinelibrary.wiley.com/journal/20427174
<1% -

https://www.researchgate.net/publication/305630430_QoS-Aware_Routing_Mechanism_in_OpenFlow-Enabled_Wireless_Multimedia_Sensor_Networks
<1% - https://pubs.acs.org/doi/full/10.1021/jf60183a013
1% - http://ijpronline.com/