Top 10 Read Articles in Computer Science & Information Technology: December 2025

International Journal of Computer Science and Information Technology (IJCSIT)

INSPEC, WJCI Indexed

ISSN: 0975-3826(online); 0975-4660 (Print)

http://airccse.org/journal/ijcsit.html

CLUSTERING ALGORITHM FOR A HEALTHCARE DATASET USING SILHOUETTE SCORE VALUE

Godwin Ogbuabor[1] and Ugwoke, F. N[2]

[1] School of Computer Science, University of Lincoln, United Kingdom
[2]Department of Computer Science, Michael Okpara University of Agriculture Umudike, Abia State, Nigeria

ABSTRACT

The huge amount of healthcare data, coupled with the need for data analysis tools has made data mining interesting research areas. Data mining tools and techniques help to discover and understand hidden patterns in a dataset which may not be possible by mainly visualization of the data. Selecting appropriate clustering method and optimal number of clusters in healthcare data can be confusing and difficult most times. Presently, a large number of clustering algorithms are available for clustering healthcare data, but it is very difficult for people with little knowledge of data mining to choose suitable clustering algorithms. This paper aims to analyze clustering techniques using healthcare dataset, in order to determine suitable algorithms which can bring the optimized group clusters. Performances of two clustering algorithms (Kmeans and DBSCAN) were compared using Silhouette score values. Firstly, we analyzed K-means algorithm using different number of clusters (K) and different distance metrics. Secondly, we analyzed DBSCAN algorithm using different minimum number of points required to form a cluster (minPts) and different distance metrics. The experimental result indicates that both K-means and DBSCAN algorithms have strong intra-cluster cohesion and inter-cluster separation. Based on the analysis, K-means algorithm performed better compare to DBSCAN algorithm in terms of clustering accuracy and execution time.

KEYWORDS

Dataset, Clustering, Healthcare data, Silhouette score value, K-means, DBSCAN

Volume URL : https://airccse.org/journal/ijcsit2018_curr.html

Source URL : https://aircconline.com/ijcsit/V10N2/10218ijcsit03.pdf

REFERENCES

[1] Alsayat, A., & El-Sayed, H. (2016). Efficient genetic K-Means clustering for health care knowledge discovery. In Software Engineering Research, Management and Applications (SERA), 2016 IEEE 14th International Conference on (pp. 45-52). IEEE.

[2] Balasubramanian, T., & Umarani, R. (2012, March). An analysis on the impact of fluoride in human health (dental) using clustering data mining technique. In Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on (pp. 370-375). IEEE.

[3] Banu G. Rasitha & Jamala J.H.Bousal (2015). Perdicting Heart Attack using Fuzzy C Means Clustering Algorithm. International Journal of Latest Trends in Engineering and Technology (IJLTET).

[4] Banu, M. N., & Gomathy, B. (2014). Disease forecasting system using data mining methods. In Intelligent Computing Applications (ICICA), 2014 International Conference on (pp. 130-133). IEEE.

[5] Belciug, S. (2009). Patients length of stay grouping using the hierarchical clustering algorithm. Annals of the University of Craiova-Mathematics and Computer Science Series, 36(2), 79-84.

[6] Belciug, S., Salem, A. B., Gorunescu, F., & Gorunescu, M. (2010, November). Clustering-based approach for detecting breast cancer recurrence. In Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on (pp. 533-538). IEEE.

[7] Bruno, G., Cerquitelli, T., Chiusano, S., & Xiao, X. (2014). A clustering-based approach to analyse examinations for diabetic patients. In Healthcare Informatics (ICHI), 2014 IEEE International Conference on (pp. 45-50). IEEE.

[8] DeFreitas, K., & Bernard, M. (2015). Comparative performance analysis of clustering techniques in educational data mining. IADIS International Journal on Computer Science & Information Systems, 10(2).

[9] Escudero, J., Zajicek, J. P., & Ifeachor, E. (2011). Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (pp. 6470-6473). IEEE.

[10] Han, J., Kamber, M., & Pei, J. (2012). Cluster Analysis-10: Basic Concepts and Methods.

[11] Ibrahim, N. H., Mustapha, A., Rosli, R., & Helmee, N. H. (2013). A hybrid model of hierarchical clustering and decision tree for rule-based classification of diabetic patients. International Journal of Engineering and Technology (IJET), 5(5), 3986-91.

[12] Jabel K. Merlin & Srividhya (2016). Performance analysis of clustering algorithms on heart dataset. International Journal of Modern Computer Science, 5(4), 113-117. International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018 37

[13] Kar Amit Kumar, Shailesh Kumar Patel & Rajkishor Yadav (2016). A Comparative Study & Performance Evaluation of Different Clustering Techniques in Data Mining. ACEIT Conference Proceeding.

[14] Lv, Y., Ma, T., Tang, M., Cao, J., Tian, Y., Al-Dhelaan, A., & Al-Rodhaan, M. (2016). An efficient and scalable density-based clustering algorithm for datasets with complex structures. Neurocomputing, 171, 9-22. [15] Malli, S., Nagesh, H. R., & Joshi, H. G. (2014). A Study on Rural Health care Data sets using Clustering Algorithms. International Journal of Engineering Research and Applications, 3(8), 517- 520.

[16] Maulik, U., & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(12), 1650-1654.

[17] Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on (pp. 63-67). IEEE.

[18] Paul, R., & Hoque, A. S. M. L. (2010, July). Clustering medical data to predict the likelihood of diseases. In Digital Information Management (ICDIM), 2010 Fifth International Conference on (pp. 44-49). IEEE.

[19] Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 219(1), 103-119.

[20] R.Nithya & P.Manikandan & D.Ramyachitra (2015); Analysis of clustering technique for the diabetes dataset using the training set parameter. International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 9.

[21] Sagar, H. K., & Sharma, V. (2014). Error Evaluation on K-Means and Hierarchical Clustering with Effect of Distance Functions for Iris Dataset. International Journal of Computer Applications, 86(16).

[22] Shah, G. H., Bhensdadia, C. K., & Ganatra, A. P. (2012). An empirical evaluation of density-based clustering techniques. International Journal of Soft Computing and Engineering (IJSCE) ISSN, 22312307, 216-223.

[23] Tan, P. N., Steinbach, M., & Kumar, V. (2013). Data mining cluster analysis: basic concepts and algorithms. Introduction to data mining.

[24] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266.

[25] Vijayarani, S., & Sudha, S. (2015). An efficient clustering algorithm for predicting diseases from hemogram blood test samples. Indian Journal of Science and Technology, 8(17).

[26] Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications, 41(4), 1476-1482.

SECURITY THREATS ON CLOUD COMPUTING VULNERABILITIES

Te-Shun Chou

Department of Technology Systems, East Carolina University, Greenville, NC, U.S.A.

ABSTRACT

Clouds provide a powerful computing platform that enables individuals and organizations to perform variety levels of tasks such as: use of online storage space, adoption of business applications, development of customized computer software, and creation of a "realistic" network environment. In previous years, the number of people using cloud services has dramatically increased and lots of data has been stored in cloud computing environments. In the meantime, data breaches to cloud services are also increasing every year due to hackers who are always trying to exploit the security vulnerabilities of the architecture of cloud. In this paper, three cloud service models were compared; cloud security risks and threats were investigated based on the nature of the cloud service models. Real world cloud attacks were included to demonstrate the techniques that hackers used against cloud computing systems. In addition, counter measures to cloud security breaches are presented.

KEYWORDS

Cloud computing, cloud security threats and countermeasures, cloud service models

Volume URL : http://airccse.org/journal/ijcsit2013_curr.html

Source URL : http://airccse.org/journal/jcsit/5313ijcsit06.pdf

REFERENCES

1. DataLossDB Open Security Foundation.http://datalossdb.org/statistics

2. Sophos Security Threat Report 2012.http://www.sophos.com/

3. Amazon.com Server Said to Have Been Used in Sony Attack, May 2011.http://www.bloomberg.com/news/2011-05-13/sony-network-said-to-have-been- invaded-by-hackersusing-amazon-com-server.html

4. Jamil and H. Zaki, "Security Issues in Cloud Computing and Countermeasures," International Journal of Engineering Science and Technology, Vol. 3 No. 4, pp. 2672- 2676, April2011.

5. K. Zunnurhain and S. Vrbsky, "Security Attacks and Solutions in Clouds," 2nd IEEE International Conference on Cloud Computing Technology and Science, Indianapolis, December2010.

6. W. A. Jansen, "Cloud Hooks: Security and Privacy Issues in Cloud Computing," 44th Hawaii International Conference on System Sciences, pp. 1–10, Koloa, Hawaii, January 2011.

7. T. Roth, "Breaking Encryptions Using GPU Accelerated Cloud Instances," Black Hat Technical Security Conference,2011.

8. CERT Coordination Center, Denial of Service.http://www.packetstormsecurity.org/distributed/denial_of_service.html

9. M. Jensen, J. Schwenk, N. Gruschka, and L. L. Iacono, "On Technical Security Issues in Cloud Computing," IEEE International Conference in Cloud Computing, pp. 109-116, Bangalore,2009.

10. Thunder in the Cloud: $6 Cloud-Based Denial-of-Service Attack, August 2010.http://blogs.computerworld.com/16708/thunder_in_the_cloud_6_cloud_based_denial_of_service_attack

11. DDoS Attack Rains Down on Amazon Cloud, October 2009.http://www.theregister.co.uk/2009/10/05/amazon_bitbucket_outage/

12. 2011 CyberSecurity Watch Survey, CERT Coordination Center at Carnegie Mellon University.

13. D. Catteddu and G. Hogben, "Cloud Computing Benefits, Risks and Recommendations for Information Security," The European Network and Information Security Agency (ENISA), November2009.

14. Insider Threats Related to Cloud Computing, CERT, July 2012.http://www.cert.org/

15. Data Breach Trends & Stats, Symantec, 2012.http://www.indefenseofdata.com/data- breach-trendsstats/

16. 2012 Has Delivered Her First Giant Data Breach, January 2012.http://www.infosecisland.com/blogview/19432-2012-Has-Delivered-Her-First- Giant-DataBreach.html

17. A Few Wrinkles Are Etching Facebook, Other Social Sites, USA Today,
18. 2011.http://www.usatoday.com/printedition/life/20090115/socialnetworking15_st.art.htm l

19. An Update on LinkedIn Member Passwords Compromised, LinkedIn Blog, June, 2012.http://blog.linkedin.com/2012/06/06/linkedin-member-passwords-compromised/

20. Dropbox: Yes, We Were Hacked, August 2012.http://gigaom.com/cloud/dropbox- yes-we-werehacked/

21. Web Based Attacks, Symantec White Paper, February2009.

22. Symantec Internet Security Threat Report, 2011 Trends, Vol. 17, April2012.

23. P. P. Ramgonda and R. R. Mudholkar,"Cloud Market Cogitation and Techniques to Averting SQL Injection for University Cloud," International Journal of Computer Technology and Applications, Vol. 3, No. 3, pp. 1217-1224, January,2012.

24. S. Choudhary and M. L. Dhore, "CIDT: Detection of Malicious Code Injection Attacks on Web Application," International Journal of Computer Applications, Vol. 52, No. 2, pp. 19-26, August2012.

25. Web Application Attack Report For The Second Quarter of 2012.http://www.firehost.com/company/newsroom/web-application-attack-report-second-quarter-2012

26. Visitors to Sony PlayStation Website at Risk of Malware Infection, July 2008.http://www.sophos.com/en-us/press-office/press-releases/2008/07/playstation.aspx

27. N. Provos, M. A. Rajab, and P. Mavrommatis, "Cybercrime 2.0: When the Cloud Turns Dark," ACM Communications, Vol. 52, No. 4, pp. 42–47,2009.

28. S. S. Rajan, Cloud Security Series | SQL Injection and SaaS, Cloud Computing Journal, November2010.

29. Researchers Demo Cloud Security Issue With Amazon AWS Attack, October 2011. http://www.pcworld.idg.com.au/article/405419/researchers_demo_cloud_security_issue_ amazon_aws_attack/

30. M. McIntosh and P. Austel, "XML Signature Element Wrapping Attacks and Countermeasures," 2005 workshop on Secure web services, ACM Press, New York, NY, pp. 20–27,2005.

31. N. Gruschka and L. L. Iacono, "Vulnerable Cloud: SOAP Message Security Validation

Revisited," IEEE International Conference on Web Services, Los Angeles, 2009.

32. Tripathi and A. Mishra, "Cloud Computing Security Considerations Interface," 2011 IEEE International Conference on Signal Processing, Communications and Computing, Xi'an, China, September2011.

33. H. C. Li, P. H. Liang, J. M. Yang, and S. J. Chen, "Analysis on Cloud-Based Security Vulnerability Assessment," IEEE International Conference on E-Business Engineering, pp.490-494, November2010.

34. Amazon: Hey Spammers, Get OffMy Cloud!http://voices.washingtonpost.com/securityfix/2008/07/amazon_hey_spammers_get_off_my.html

35. W. Jansen and T. Grance, "Guidelines on Security and Privacy in Public Cloud Computing," Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, Special Publication 800-144, December2011.

36. Tackling the Insider Threathttp://www.bankinfosecurity.com/blogs.php?postID=140

37. "Cloud Security Risks and Solutions," White Paper, BalaBit IT Security, July2010.

38. S. J. Stolfo, M. B. Salem, and A. D. Keromytis, "Fog computing: Mitigating Insider Data Theft Attacks in the Cloud," IEEE Symposium on Security and Privacy Workshops, pp. 125-128, San Francisco, CA,2012.

39. M. Jensen, C. Meyer, J. Somorovsky, and J. Schwenk, "On the Effectiveness of XML Schema Validation for Countering XML Signature Wrapping Attacks," First International Workshop on Securing Services on the Cloud, Milan, Italy, September 2011.

40. S. Gajek, M. Jensen, L. Liao, and J. Schwenk, "Analysis of Signature Wrapping Attacks and Countermeasures," IEEE International Conference on Web Services, pp. 575–582, Miami, Florida, July2009.

DATA MINING MODEL PERFORMANCE OF SALES PREDICTIVE ALGORITHMS BASED ON RAPIDMINER WORKFLOWS

Alessandro Massaro, Vincenzo Maritati

Angelo Galiano Dyrecta Lab, IT research Laboratory,via Vescovo Simplicio, 45, 70014 Conversano (BA), Italy

ABSTRACT

By applying RapidMiner workflows has been processed a dataset originated from different data files, and containing information about the sales over three years of a large chain of retail stores. Subsequently, has been constructed a Deep Learning model performing a predictive algorithm suitable for sales forecasting. This model is based on artificial neural network –ANN- algorithm able to learn the model starting from sales historical data and by pre-processing the data. The best built model uses a multilayer neural network together with an "optimized operator" able to find automatically the best parameter setting of the implemented algorithm. In order to prove the best performing predictive model, other machine learning algorithms have been tested. The performance comparison has been performed between Support Vector Machine –SVM-, k-Nearest Neighbor k-NN-,Gradient Boosted Trees, Decision Trees, and Deep Learning algorithms. The comparison of the degree of correlation between real and predicted values, the average absolute error and the relative average error proved that ANN exhibited the best performance. The Gradient Boosted Trees approach represents an alternative approach having the second best performance. The case of study has been developed within the framework of an industry project oriented on the integration of high performance data mining models able to predict sales using– ERP- and customer relationship management –CRM- tools.

KEYWORDS

RapidMiner, Neural Network, Deep Learning, Gradient Boosted Trees, Data Mining Performance, Sales Prediction.

Volume URL : http://airccse.org/journal/ijcsit2018_curr.html

Source URL : http://aircconline.com/ijcsit/V10N3/10318ijcsit03.pdf

REFERENCES

[1] Penpece D., & Elma O. E. (2014) "Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry: A Case Study in Turkey", International Journal of Trade Economics and Finance, Vol. 5, No. 5, pp435-440.

[2] Thiesing F. M., & Vornberger, O. (1997) "Sales Forecasting Using Neural Networks", IEEE Proceedings ICNN'97, Houston, Texas, 9-12 June 1997, pp2125-2128.

[3] Zhang, G. P. (2003) "Time series forecasting using a hybrid ARIMA and neural network model", Neurocomputing, Vol. 50, pp159–175.

[4] Sharma, A., & Panigrahi, P. K. (2011) "Neural Network based Approach for Predicting Customer Churn in Cellular Network Services", International Journal of Computer Applications, Vol. 27, No.11, pp0975–8887.

[5] Kamakura, W., Mela, C. F., Ansari A., & al. (2005) " Choice Models and Customer Relationship Management," Marketing Letters, Vol. 16, No.3/4, pp279–291.

[6] Smith, K. A., & Gupta, J. N. D. (2000) "Neural Networks in Business: Techniques and Applications for the Operations Researcher," Computers & Operations Research, Vol. 27, No. 11–12, pp1023- 1044.

[7] Chattopadhyay, M., Dan, P. K., Majumdar, S., & Chakraborty, P. S. (2012) "Application of Artificial Neural Network in Market Segmentation: A Review on Recent Trends," Management Science Letters, Vol. 2, pp425-438.

[8] Berry, J. A. M., & Linoff, G. S. (2004) "Data Mining Techniques For Marketing, Sales, and Customer Relationship Management", Wiley, Second Edition.

[9] Buttle, F. (2009) "Customer Relationship Management Concepts and Technologies", Elsevier, Second Edition.

[10] Thomassey, S. (2014) "Sales Forecasting in Apparel and Fashion Industry: A Review", Springer, chapter 2.

[11] Massaro, A. Barbuzzi, D., Vitti, V., Galiano, A., Aruci, M., Pirlo, G. (2016) "Predictive Sales Analysis According to the Effect of Weather", Proceeding of the 2nd International Conference on Recent Trends and Applications in Computer Science and Information Technology, Tirana, Albania, November 18 - 19, pp53-55.

[12] Parsons, A.G. (2001), "The Association between Daily Weather and Daily Shopping Patterns", Australasian Marketing Journal, Vol. 9, No. 2, pp78–84. [13] Steele, A.T., (1951) "Weather's Effect on the Sales of a Department Store", Journal of Marketing Vol. 15, No. 4, pp436–443.

[14] Murray, K. B., Di Muro, F., Finn, A., & Leszczyc, P. P. (2010) "The Effect of Weather on Consumer Spending", Journal of Retailing and Consumer Services, Vol. 17, No.6, pp512-520.

[15] Massaro, A., Galiano, A., Barbuzzi, D., Pellicani, L., Birardi, G., Romagno, D. D., & Frulli, L., (2017) "Joint Activities of Market Basket Analysis and Product Facing for Business Intelligence oriented on Global Distribution Market: examples of data mining applications,"

[16] Aguinis, H., Forcum, L. E., & Joo, H. (2013) "Using Market Basket Analysis in Management Research," Journal of Management, Vol. 39, No. 7, pp1799-1824.

[17] Štulec, I, Petljak, K., & Kukor, A. (2016) "The Role of Store Layout and Visual Merchandising in Food Retailing", European Journal of Economics and Business Studies, Vol. 4, No. 1, pp139-152.

[18] Otha, M. & Higuci, Y. (2013) "Study on Design of Supermarket Store Layouts: the Principle of "Sales Magnet"", World Academy of Science, Engieering and Technology, Vol. 7, No. 1, pp209-212.

[19] Shallu, & Gupta, S. (2013) "Impact of Promotional Activities on Consumer Buying Behavior: A Study of Cosmetic Industry", International Journal of Commerce, Business and Management (IJCBM), Vol. 2, No.6, pp379-385.

[20] Al Essa, A. & Bach, C. (2014)" Data Mining and Knowledge Management for Marketing", International Journal of Innovation and Scientific Research, Vol. 2, No. 2, pp321-328.

[21] Kotu, V., & Deshpande B. (2015) "Predictive Analytics and Data Mining- Concepts and Practice with RapidMiner" Elsevier.

[22] Wimmer, H., Powell, L. M. (2015) "A Comparison of Open Source Tools for Data Science", Proceedings of the Conference on Information Systems Applied Research. Wilmington, North Carolina USA.

[23] Al-Khoder, A., Harmouch, H., "Evaluating Four Of The most Popular Open Source and Free Data Mining Tools", International Journal of Academic Scientific Research, Vol. 3, No. 1, pp13-23.

[24] Gulli, A., & Pal, S. (2017) "Deep Learning with Keras- Implement neural networks with Keras on Theano and TensorFlow," Birmingham -Mumbai Packt book, ISBN 978-1-78712-842-2.

[25] Kovalev, V., Kalinovsky, A., & Kovalev, S. (2016) "Deep Learning with Theano, Torch, Caffe, TensorFlow, and deeplearning4j: which one is the best in speed and accuracy?" Proceeding of XIII Int. Conf. on Pattern Recognition and Information Processing, 3-5 October, Minsk, Belarus State University, pp99-103.

[26] "Walmart Recruiting - Store Sales Forecasting" 2018. [Online]. Available: https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data

[27] Huang, H.-C. & Hou, C.-I.. (2017) "Tourism Demand Forecasting Model Using Neural Network", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 9, No. 2, pp19- 29.

[28] Kalyani, J., Bharathi, H. N., & Rao, J. (2016) "Stock Trend Prediction Using News Sentiment Analysis", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 8, No. 3, pp67-76.

PRODUCT SENTIMENT ANALYSIS FOR AMAZON REVIEWS

Arwa S. M. AlQahtani

Department of Computer Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

ABSTRACT

Recently, Ecommerce has Witnessed Rapid Development. As A Result, Online Purchasing has grown, and that has led to Growth in Online Customer Reviews of Products. The Implied Opinions in Customer Reviews Have a Massive Influence on Customer's Decision Purchasing, Since the Customer's Opinion About the Product is Influenced by Other Consumers' Recommendations or Complaints. This Research Provides an Analysis of the Amazon Reviews Dataset and Studies Sentiment Classification with Different Machine Learning Approaches. First, the Reviews were Transformed into Vector Representation using different Techniques, I.E., Bag-Of-Words, Tf-Idf, and Glove. Then, we Trained Various Machine Learning Algorithms, I.E., Logistic Regression, Random Forest, Naïve Bayes, Bidirectional Long-Short Term Memory, and Bert. After That, We Evaluated the Models using Accuracy, F1-Score, Precision, Recall, and Cross-Entropy Loss

Function. Then, We Analyzed The Best Performance Model in Order to Investigate Its Sentiment Classification. The Experiment was Conducted on Multiclass Classifications, Then we Selected the Best Performing Model And Re-Trained It on the Binary Classification.

KEYWORDS

Amazon, Data Analytics, Analysis, Product Sentiment, Ecommerce

Volume URL : https://airccse.org/journal/ijcsit2021_curr.html

Source URL : https://aircconline.com/ijcsit/V13N3/13321ijcsit02.pdf

REFERENCES

[1] S. A. a. A. N. S. Aljuhani, “A Comparison of Sentiment Analysis Methods on Amazon Reviews of Mobile Phones,” International Journal of Advanced Computer Science and Applications, vol. 10, 2019.

[2] L. a. L. B. Zhang, “Aspect and entity extraction for opinion mining,” in Zhang, Lei and Liu, Bing, Berlin, Heidelberg, Springer, 2014, pp. 1--40.

[3] Y.-C. a. K. C.-H. a. C. C.-H. Chang, “Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor,” International Journal of Information Management, vol. 48, pp. 263--279, 2019.

[4] K. S. a. D. J. a. M. J. Kumar, “Opinion mining and sentiment analysis on online customer review,” in IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016.

[5] A. S. a. A. A. a. D. P. Rathor, “Comparative study of machine learning approaches for Amazon reviews,” Procedia computer science, vol. 132, pp. 1552--1561, 2018.

[6] B. a. S. S. Bansal, “Sentiment classification of online consumer reviews using word vector representations,” Procedia computer science, vol. 132, pp. 1147--1153, 2018.

[7] A. a. S. V. a. M. B. ernian, “Sentiment analysis from product reviews using SentiWordNet as lexical resource,” in 2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, 2015.

[8] J. F. V. W. P. G. N. Rodrigo Moraes, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” Expert Systems with Applications, vol. 40, pp. 621--633, 2013. [

9] D. a. X. H. a. S. Z. a. X. Y. Zhang, “Chinese comments sentiment classification based on word2vec and SVMperf,” Expert Systems with Applications, vol. 42, pp. 857--1863, 2015.

[10] Y. a. L. M. a. E. K. K. E. Al Amrani, “Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis,” Procedia Computer Science, pp. 511-520, 2018.

[11] Y. a. K. V. Saito, “Classifying User Reviews at Sentence and Review Levels Utilizing Naïve Bayes,” in 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang Kwangwoon_Do, Korea (South), 2019.

[12] ,. X. C. T. S. M. W. N. J. Sobia Wassan, “Amazon Product Sentiment Analysis using Machine,” Revista Argentina de Clínica Psicológica, pp. 695-703, 2021.

[13] Bahrawi, “Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based.,” JOURNAL OF INFORMATION TECHNOLOGY AND ITS UTILIZATION, vol. 2, pp. 29-33, 2019.

[14] M. a. S. R. Fikri, “A Comparative Study of Sentiment Analysis using SVM and SentiWordNet,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 13, pp. 902-909, 2019.

[15] N. Tamara and Milievi, “Comparing sentiment analysis and document representation methods of Amazon reviews,” 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), pp. 000283--000286, 2018.

[16] K. a. M. W. a. C. W. Ogada, “N-gram Based Text Categorization Method for Improved Data Mining,” Journal of Information Engineering and Applications, vol. 5, pp. 35--43, 2015. [17] R. a. P. B. a. S. S. Al-Rfou, “Polyglot: Distributed word representations for multilingual nlp,” arXiv preprint

arXiv:1307.1662, 2013.

[18] Y. a. A. G. a. J. P. a. K. T. Sharma, “Vector representation of words for sentiment analysis using GloVe,” in 2017 international conference on intelligent communication and computational techniques (icct), Jaipur, 2017.

[19] R. S. C. D. M. Jeffrey Pennington, “GloVe: Global Vectors for Word Representation,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 2014.

[20] A. H. C. H. H. B. G. Marwa Naili, “Comparative study of word embedding methods in topic segmentation,” Procedia Computer Science, vol. 112, pp. 340-349, 2017.

[21] H. a. K. A. Sinha, “A Detailed Survey and Comparative Study of sentiment analysis algorithms,” in 2016 2nd International Conference on Communication Control and Intelligent Systems (CCIS), Mathura, India, 2016.

[22] M. a. O. T. Bouazizi, “A pattern-based approach for multi-class sentiment analysis in Twitter,” in 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 2016.

[23] V. M. N. Harpreet Kaur, “A survey of sentiment analysis techniques,” in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017.

[24] Z. a. F. Y. a. J. B. a. L. T. a. L. W. Li, “{A survey on sentiment analysis and opinion mining for social multimedia,” Multimedia Tools and Applications, vol. 78, pp. 6939--6967, 2019.

[25] A. K. A. a. A. A. B. A. Hassan, “Reviews Sentiment analysis for collaborative recommender system,” Kurdistan journal of applied research, vol. 2, pp. 87--91, 2017.

[26] V. M. a. V. J. a. B. P. Pradhan, “A survey on Sentiment Analysis Algorithms for opinion mining,” International Journal of Computer Applications, vol. 133, pp. 7--11, 2016.

[27] H. a. B. S. a. S. G. Parmar, “Sentiment mining of movie reviews using Random Forest with Tuned Hyperparameters,” in International Conference on Information Science., 2014, Kerala.

[28] J. a. C. M.-W. a. L. K. a. T. K. Devlin, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.

[29] Y. a. L. M. a. L. L. a. F. Z. a. W. F.-X. a. W. J. Yu, “Automatic ICD code assignment of Chinese clinical notes based on multilayer attention BiRNN,” Journal of biomedical informatics, vol. 91, pp. 103-114, 2019.

[30] Y. a. S. X. a. H. C. a. Z. J. Yu, “A review of recurrent neural networks: LSTM cells and network architectures,” Neural computation, vol. 31, pp. 1235--1270, 2019.

[31] C. a. S. C. a. L. Z. a. L. F. Zhou, “A C-LSTM Neural Network for Text Classification,” arXiv preprint arXiv:1511.08630, 2015.

[32] A. a. A. A. a. R. S. K. Tripathy, “Classification of sentiment reviews using n-gram machine learning approach,” Expert Systems with Applications, vol. 57, pp. 117--126, 2016.

[33] S. M. Mudambi, D. Schuff and Z. Zhang, “Why Aren't the Stars Aligned? An Analysis of Online Review Content and Star Ratings,” in 2014 47th Hawaii International Conference on System Sciences, Waikoloa, 2014.

FUTURE AND CHALLENGES OF INTERNET OF THINGS

Falguni Jindal[1] , Rishabh Jamar[2] , Prathamesh Churi[3]

[1,2]Bachelors of Technology in Computer Engineering SVKM's NMIMS Mukesh Patel School of Technology Management and Engineering, Mumbai, India
[3]Assistant Professor (Computer Engineering) SVKM's NMIMS Mukesh Patel School of Technology Management and Engineering, Mumbai, India.

ABSTRACT

The world is moving forward at a fast pace, and the credit goes to ever growing technology. One such concept is IOT (Internet of things) with which automation is no longer a virtual reality. IOT connects various non-living objects through the internet and enables them to share information with their community network to automate processes for humans and makes their lives easier. The paper presents the future challenges of IoT , such as the technical (connectivity , compatibility and longevity , standards , intelligent analysis and actions , security), business (investment , modest revenue model etc.), societal (changing demands , new devices, expense, customer confidence etc.) and legal challenges (laws, regulations, procedures, policies etc.). A section also discusses the various myths that might hamper the progress of IOT, security of data being the most critical factor of all. An optimistic approach to people in adopting the unfolding changes brought by IOT will also help in its growth

KEYWORDS

IoT, Internet of Things, Security, Sensors

For More Details : https://aircconline.com/ijcsit/V10N2/10218ijcsit02.pdf

Volume Link : http://airccse.org/journal/ijcsit2018_curr.html

REFERENCES

[1] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.

[2] Li, S., Da Xu, L., & Zhao, S. (2015). The internet of things: a survey. Information Systems Frontiers, 17(2), 243-259.

[3] Guo, B., Zhang, D., Wang, Z., Yu, Z., & Zhou, X. (2013). Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things. Journal of Network and Computer Applications, 36(6), 1531-1539.

[4] Banafa, A. (2014). IoT Standardization and Implementation Challenges. IEEE. org Newsletter.

[5] Banafa, A. (2015). „What is next for IoT and IIoT". Enterprise Mobility Summit.

[6] Coetzee, L., & Eksteen, J. (2011, May). The Internet of Things-promise for the future? An introduction. In IST-Africa Conference Proceedings, 2011 (pp. 1-9). IEEE.

[7] Cai, H., Da Xu, L., Xu, B., Xie, C., Qin, S., & Jiang, L. (2014). IoT-based configurable information service platform for product lifecycle management. IEEE Transactions on Industrial Informatics, 10(2), 1558-1567.

[8] Khan, R., Khan, S. U., Zaheer, R., & Khan, S. (2012, December). Future internet: the internet of things architecture, possible applications and key challenges. In Frontiers of Information Technology (FIT), 2012 10th International Conference on (pp. 257-260). IEEE.

[9] Liu, Y., & Zhou, G. (2012, January). Key technologies and applications of internet of things. In Intelligent Computation Technology and Automation (ICICTA), 2012 Fifth International Conference on (pp. 197-200). IEEE.

[10] Sadeghi, A. R., Wachsmann, C., & Waidner, M. (2015, June). Security and privacy challenges in industrial internet of things. In Proceedings of the 52nd annual design automation conference (p. 54). ACM.

[11] Banafa, A. (2014). IoT and Blockchain Convergence: Benefits and Challenges. IEEE Internet of Things.

[12] Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT data analytics: Architecture, opportunities, and open research challenges. IEEE Access, 5, 5247-5261.

[13] Desai, P., Sheth, A., & Anantharam, P. (2015, June). Semantic gateway as a service architecture for iot interoperability. In Mobile Services (MS), 2015 IEEE International Conference on(pp. 313-319). IEEE.

[14] Koivu, A., Koivunen, L., Hosseinzadeh, S., Laurén, S., Hyrynsalmi, S., Rauti, S., & Leppänen, V. (2016, December). Software Security Considerations for IoT. In Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2016 IEEE International Conference on (pp. 392-397). IEEE.

[15] Sundmaeker, H., Guillemin, P., Friess, P., & Woelfflé, S. (2010). Vision and challenges for realising the Internet of Things. Cluster of European Research Projects on the Internet of Things, European Commision, 3(3), 34-36.

[16] Vermesan, O., Friess, P., Guillemin, P., Gusmeroli, S., Sundmaeker, H., Bassi, A., ... & Doody, P. (2011). Internet of things strategic research roadmap. Internet of Things-Global Technological and Societal Trends, 1(2011), 9-52.

[17] Sheng, Z., Yang, S., Yu, Y., Vasilakos, A., Mccann, J., & Leung, K. (2013). A survey on the ietf protocol suite for the internet of things: Standards, challenges, and opportunities. IEEE Wireless Communications, 20(6), 91-98.

[18] Theoleyre, F., & Pang, A. C. (Eds.). (2013). Internet of Things and M2M Communications. River Publishers.

[19] Coetzee, L., & Eksteen, J. (2011, May). The Internet of Things-promise for the future? An introduction. In IST-Africa Conference Proceedings, 2011 (pp. 1-9). IEEE.

[20] Ji, Z., & Anwen, Q. (2010, November). The application of internet of things (IOT) in emergency management system in China. In Technologies for Homeland Security (HST), 2010 IEEE International Conference on (pp. 139-142). IEEE.

[21] James Kirkland , “Internet of Things: insights from Red Hat” , Website: https://developers.redhat.com/blog/2015/03/31/internet-of-things-insights-from-red-hat/ , Accesed : 2nd February 2018

Monitoring Student Attendance Using A Smart System at Taif University

Saleh Alghamdi

Department of Information Technology, Taif University, Al-taif, Saudi Arabia

ABSTRACT

The university system in the Kingdom of Saudi Arabia is concerned with student attendance for lectures, and it is the responsibility of lecturers to monitor student attendance for each lecture. By the end of the semester, students get an attendance register indicating which lectures the student has attended and it reports the calculated percentage for each student's attendance in each course. Universities have regulated the mechanisms and the acceptable percentages of student absence. The process for a lecturer to manually check student attendance consumes a lot of time and effort, either during the lecture or when in the process of emptying absenteeism and inserting it into the university's electronic system. Therefore, Saudi universities compete to find modern methods of checking student attendance that will avoid the disadvantages of manually taking attendance. For this reason, they have produced electronic attendance systems, for example, using a student's fingerprint, an eye recognition system, or a mobile phone system to read a QR code designed for the same purpose. All of these systems have the disadvantage that they consume a lot of time, as all students have to line up at the fingerprint reader or the eye detector for identification. Therefore, the problem of the consumption of lecture time is still present, even with these modern systems. Therefore, the aim of this research is to propose a smart mobile application that is able to check the attendance of students without having to consume lecture time or require any effort from the lecturer. The system automatically recognizes the attendance of students through their university ID cards. Each lecturer would use his/ her own mobile phone to use the proposed system to check the attendance of students instead of using manual method to register the attendance of students and the students' ID cards that are detected by coming within range of the lecturer reader would represent present students, and missing student ID cards represent absent students

KEYWORDS

Context Awareness, RFID, Monitoring Student Attendance.

Volume URL : https://airccse.org/journal/ijcsit2019_curr.html

Source URL : https://aircconline.com/ijcsit/V11N1/11119ijcsit08.pdf

REFERENCES

[1] A. N. Yumang, D. Padilla, M. Sejera, A. C. U. Pajarillo, G. V. L. B. Palmiano and M. M. F. Racho, "Attendance checker for students of Mapúa University," 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Manila, 2017, pp. 1-3.

[2] M. M. Islam, M. K. Hasan, M. M. Billah and M. M. Uddin, "Development of smartphone-based student attendance system," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, 2017, pp. 230-233.

[3] M. I. Moksin and N. M. Yasin, "The Implementation of Wireless Student Attendance System in an Examination Procedure," 2009 International Association of Computer Science and Information Technology - Spring Conference, Singapore, 2009, pp. 174-177.

[4] L. Ma, J. Pan, L. Cao and J. Shen, "The Design of Wireless Students Management System Based on Fingerprint Sensor," 2010 International Conference on E-Business and E-Government, Guangzhou, 2010, pp. 2153-2155.

[5] A. Purohit, K. Gaurav, C. Bhati and A. Oak, "Smart attendance," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2017, pp. 415- 419.

[6] J. A. Rubella, M. Suganya, K. Senathipathi, B. S. Kumar, K. R. Gowdham and M. Ranjithkumar, "Fingerprint based license checking for auto-mobiles," 2012 Fourth International Conference on Advanced Computing (ICoAC), Chennai, 2012, pp. 1-8.

[7] Han Xi, Niu Wenliang and Li Zheying, "Application of fingerprint recognition on the laboratory management," 2008 9th International Conference on Signal Processing, Beijing, 2008, pp. 2960-2963.

[8] N. Wenliang, H. Xi, L. Zheying and W. Shuying, "An Open Laboratory Management System Based on Fingerprints Recognition," 2009 Second International Symposium on Computational Intelligence and Design, Changsha, 2009, pp. 510-512.

[9] R. Samet and M. Tanriverdi, "Face Recognition-Based Mobile Automatic Classroom Attendance Management System," 2017 International Conference on Cyberworlds (CW), Chester, 2017, pp. 253-256.

[10] Y. Mao, H. Li and Z. Yin, "Who missed the class? — Unifying multi-face detection, tracking and recognition in videos," 2014 IEEE International Conference on Multimedia and Expo (ICME), Chengdu, 2014, pp. 1-6.

[11] S. M. Čisar, R. Pinter, V. Vojnić, V. Tumbas and P. Čisar, "Smartphone application for tracking students' class attendance," 2016 IEEE 14th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, 2016, pp. 227-232.

[12] S. M. A. Rahman and O. N. A. Hatali, "Trends of the Academic Staff to Use Electronic Attendance in the College of Applied Science, Ibri," 2015 Fifth International Conference on eLearning (econf), Manama, 2015, pp. 225-230.

[13] J. Park, K. An, D. Kim and J. Choi, "Multiple human tracking using multiple kinects for an attendance check system of a smart class," 2013 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Jeju, 2013, pp. 130-130. International Journal of Computer Science

[14] G. J. Saxena et al., "Three stage robust attendance recording and monitoring using social media platform, image processing and cloud computing," 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, 2017, pp. 194-199.

[15] U. E. Peter, C. K. A. Joe-Uzuegbu, L. Uzoechi and F. K. Opara, "Biometric-based attendance system with remote real-time monitoring for tertiary institutions in developing countries," 2013 IEEE International Conference on Emerging & Sustainable Technologies for Power & ICT in a Developing Society (NIGERCON), Owerri, 2013, pp. 1-8.

[16] Xiao, Sheng & Liang, Wenlong & Tang, Yishan. (2018). Classroom Attention Restoration Using Computer Game Rewarding Mechanism. pp. 1-6.

[17] J. He, A. Atabekov and H. M. Haddad, "Internet-of-Things Based Smart Resource Management System: A Case Study Intelligent Chair System," 2016 25th International Conference on Computer Communication and Networks (ICCCN), Waikoloa, HI, 2016, pp. 1-6

AUTHOR

Saleh Ahmed Alghamdi, Assistant Professor of College of Computers and Information Technology, department of Information Technology, Taif University, Taif, Saudi Arabia. Saleh completed Bachelor of Education degree in the department of Computer Science, Teachers' college, Riyadh, Saudi Arabia, GPA 4.72 out of 5 With the second honor degree, 2004. Then he got Master of Information Technology, from Latrobe University, Melbourne, Australia. 2008-2010. After that Saleh got Doctor of Philosophy (Computer Science), Royal Melbourne Institute of Technology (RMIT) University, Melbourne, Australia. 2010- 2014, thesis title "A Context-aware Navigational Autonomy Aid for the Blind". Now the main area of Interest in research is: Context Awareness, Positioning and Navigation and Visually Impaired Assistance.

DATA WAREHOUSE AND BIG DATA INTEGRATION

Sonia Ordoñez Salinas and Alba Consuelo Nieto Lemus

Faculty of Engineering, Distrial F.J.C University, Bogotá, Colombia

ABSTRACT

Big Data triggered furthered an influx of research and prospective on concepts and processes pertaining previously to the Data Warehouse field. Some conclude that Data Warehouse as such will disappear; others present Big Data as the natural Data Warehouse evolution (perhaps without identifying a clear division between the two); and finally, some others pose a future of convergence, partially exploring the possible integration of both. In this paper, we revise the underlying technological features of Big Data and Data Warehouse, highlighting their differences and areas of convergence. Even when some differences exist, both technologies could (and should) be integrated because they both aim at the same purpose: data exploration and decision making support. We explore some convergence strategies, based on the common elements in both technologies. We present a revision of the state-of-the-art in integration proposals from the point of view of the purpose, methodology, architecture and underlying technology, highlighting the common elements that support both technologies that may serve as a starting point for full integration and we propose a proposal of integration between the two technologies.

KEYWORDS

Big Data, Data Warehouse, Integration, Hadoop, NoSql, MapReduce, 7V's, 3C's, M&G

Volume URL : http://airccse.org/journal/ijcsit2017_curr.html

Source URL : https://aircconline.com/ijcsit/V9N2/9217ijcsit01.pdf

REFERENCES

[1] P. Bedi, V. Jindal, and A. Gautam, “Beginning with Big Data Simplified,” 2014.

[2] R. Kimball, M. Ross, W. Thorthwaite, B. Becker, and M. J, The Data Warehouse Lifecycle Toolkit, 2nd Edition. 2008.

[3] C. Todman, Designing A Data Warehouse: Supporting Customer Relationship Management. 2001.

[4] W. H. Inmon, Building the Data Warehouse, 4th Edition. 2005. [5] “Oracle Database 12c for Data Warehousing and Big Data .” [Online]. Available: http://www.oracle.com/technetwork/database/bi-datawarehousing/data-warehousing-wp-12c1896097.pdf. [Accessed: 09-Sep-2015].

[6] M. Cox and D. Ellsworth, “Application-Controlled Demand Paging for Out-of-Core Visualization,” 1997. [Online]. Available: http://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf. [Accessed: 09-Apr-2015].

[7] S. Chaudhuri and U. Dayal, “An overview of data warehousing and OLAP technology,” ACM SIGMOD Rec., vol. 26, no. 1, pp. 65–74, 1997. [8] T. Maiorescu, “General Information on Business Intelligence,” pp. 294–297, 2010. [9] “Data Warehouses and OLAP: Concepts, Architectures and Solutions: 9781599043647: Library and Information Science Books | IGI Global.” .

[10] Y. Demchenko, C. De Laat, and P. Membrey, “Defining Architecture Components of the Big Data Ecosystem,” Collab. Technol. Syst. (CTS), 2014 Int. Conf., pp. 104–112, 2014.

[11] G. NBD-PWG, “ISO/IEC JTC 1 Study Group on Big Data,” 2013. [Online]. Available: http://bigdatawg.nist.gov/cochairs.php. [Accessed: 24-Oct-2015].

[12] D. L. W.H. Inmon, Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault. Amsterdam,Boston: Elsevier, 2014.

[13] G. N. W.H. Inmon, Derek Strauss, DW 2.0: The Architecture for the Next Generation of Data Warehousing (Morgan Kaufman Series in Data Management Systems) (): : Books. Burlington, USA: Morgan Kaufmann Publishers Inc., 2008.

[14] R. Kimball, “The Evolving Role of the Enterprise Data Warehouse in the Era of Big Data Analytics,” Kimball Gr., 2011.

[15] M. Muntean and T. Surcel, “Agile BI - The Future of BI,” Inform. Econ., vol. 17, no. 3, pp. 114–124, 2013.

[16] D. Agrawal, “The Reality of Real-Time Business Intelligence,” in Business Intelligence for the RealTime Enterprise, vol. 27 , M. Castellanos, U. Dayal, and T. Sellis, Eds. Springer Berlin Heidelberg , 2009, pp. 75–88.

[17] R. Castillo, J. Morata, and L. del Arbol, “Operational Data Store (ODS) - 933.pdf,” Actas del III taller nacional de minería de datos y aprendizaje, pp. 359–365, 2005.

[18] S. YiChuan and X. Yao, “Research of Real-time Data Warehouse Storage Strategy Based on Multilevel Caches,” Phys. Procedia, vol. 25, no. 0, pp. 2315–2321, 2012.

[19] A. Ma. P. Díaz-zorita, “Evaluación de la herramienta de código libre Apache Hadoop,” Universidad Carlos III de Madrid Escuela Politécnica Superior, 2011.

[20] R. Kimball, “Newly Emerging Best Practices for Big Data,” Kimball Group, p. 14, 2012.

[21] M. Maier, “Towards a Big Data Reference Architecture,” no. October, pp. 1–144, 2013.

[22] O. Corporation, “ORACLE ENTERPRISE ARCHITECTURE WHITE PAPER. An Enterprise Architect ’ s Guide to Big Data,” no. February, 2015.

[23] F. Kramer, H. Muller, and K. Turowski, “Acceleration of Single Inserts for Columnar Databases -- An Experiment on Data Import Performance Using SAP HANA,” in Signal-Image Technology and Internet-Based Systems (SITIS), 2014 Tenth International Conference on, 2014, pp. 672–676.

[24] M. R. Patil and F. Thia, Pentaho for Big Data Analytics, vol. 2013. PACKT PUBLISHING, 2013.

[25] S. G. Manikandan and S. Ravi, “Big Data Analysis Using Apache Hadoop,” in IT Convergence and Security (ICITCS), 2014 International Conference on , 2014, pp. 1–4.

[26] J. Nandimath, E. Banerjee, A. Patil, P. Kakade, and S. Vaidya, “Big data analysis using Apache Hadoop,” 2013 IEEE 14th Int. Conf. Inf. Reuse Integr., pp. 700–703, 2013.

[27] A. Katal, M. Wazid, and R. H. Goudar, “Big data: Issues, challenges, tools and Good practices,” in Contemporary Computing (IC3), 2013 Sixth International Conference on , 2013, pp. 404–409.

[28] A. Pal and S. Agrawal, “An Experimental Approach Towards Big Data for Analyzing Memory Utilization on a Hadoop cluster using HDFS and MapReduce .,” pp. 442–447, 2014.

[29] R. Zhang, D. Hildebrand, and R. Tewari, “In unity there is strength: Showcasing a unified big data platform with MapReduce Over both object and file storage,” in Big Data (Big Data), 2014 IEEE International Conference on , 2014, pp. 960–966.

[30] “Welcome to ApacheTM Hadoop®!” [Online]. Available: https://hadoop.apache.org/. [Accessed: 26- Mar-2015].

[31] “HDFS Architecture Guide.” [Online]. Available: http://hadoop.apache.org/docs/r1.2.1/hdfs_design.html. [Accessed: 26-Mar-2015].

[32] S. Brin and L. Page, “The Anatomy of a Large-Scale Hypertextual Web Search Engine,” in Computer Networks and ISDN Systems, 1998, pp. 107–117. [

[33] D. Garlasu, V. Sandulescu, I. Halcu, G. Neculoiu, O. Grigoriu, M. Marinescu, and V. Marinescu, “A big data implementation based on Grid computing,” in Roedunet International Conference (RoEduNet), 2013 11th, 2013, pp. 1–4.

[34] A. Jorgensen, C. Price, B. Mitchell, and J. Rowlan, Microsoft Big Data Solutions. John Wiley & Sons, Inc., 2014.

[35] R. T. Kaushik, M. Bhandarkar, and K. Nahrstedt, “Evaluation and Analysis of GreenHDFS: A SelfAdaptive, Energy-Conserving Variant of the Hadoop Distributed File System,” in Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference

on, 2010, pp. 274–287.

[36] J. G. Shanahan and L. Dai, “Large Scale Distributed Data Science Using Apache Spark,” in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 2323–2324.

[37] R. S. Xin, J. Rosen, M. Zaharia, M. J. Franklin, S. Shenker, and I. Stoica, “Shark: SQL and Rich Analytics at Scale,” in Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, 2013, pp. 13–24.

[38] J. Li, J. Wu, X. Yang, and S. Zhong, “Optimizing MapReduce Based on Locality of K-V Pairs and Overlap between Shuffle and Local Reduce,” in Parallel Processing (ICPP), 2015 44th International Conference on, 2015, pp. 939–948.

[39] E. Brewer, “CAP Twelve Years Later: How the ‘Rules’ Have Changed,” InfoQ, 2012. [Online]. Available: http://www.infoq.com/articles/cap-twelve-years-later-how-the-rules-have-changed. [Accessed: 26-Mar-2015].

[40] G. Vaish, Getting started with NoSQL. 2013.

[41] V. N. Gudivada, D. Rao, and V. V. Raghavan, “NoSQL Systems for Big Data Management,” 2014 IEEE World Congr. Serv., pp. 190–197, 2014.

[42] Cassandra, “The Apache Cassandra Project,” httpcassandraapacheorg, 2010. [Online]. Available: http://cassandra.apache.org/.

[43] D. Borthakur, “Petabyte Scale Databases and Storage Systems at Facebook,” in Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, 2013, pp. 1267–1268.

[44] J. Huang, X. Ouyang, J. Jose, M. Wasi-ur-Rahman, H. Wang, M. Luo, H. Subramoni, C. Murthy, and D. K. Panda, “High-Performance Design of HBase with RDMA over InfiniBand,” in Parallel

[45] G. Weintraub, “Dynamo and BigTable - Review and comparison,” Electr. Electron. Eng. Isr. (IEEEI), 2014 IEEE 28th Conv., pp. 1–5, 2

[46] D. Pereira, P. Oliveira, and F. Rodrigues, “Data warehouses in MongoDB vs SQL Server: A comparative analysis of the querie performance,” in Information Systems and Technologies (CISTI), 2015 10th Iberian Conference on, 2015, pp. 1–7.

[47] K. Dehdouh, F. Bentayeb, O. Boussaid, and N. Kabachi, “Columnar NoSQL CUBE: Agregation operator for columnar NoSQL data warehouse,” in Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on, 2014, pp. 3828–3833.

[48] Y. Liu and T. M. Vitolo, “Graph Data Warehouse: Steps to Integrating Graph Databases Into the Traditional Conceptual Structure of a Data Warehouse,” in Big Data (BigData Congress), 2013 IEEE International Congress on, 2013, pp. 433–434.

[49] M. Chevalier, M. El Malki, A. Kopliku, O. Teste, and R. Tournier, “Benchmark for OLAP on NoSQL technologies comparing NoSQL multidimensional data warehousing solutions,” in Research Challenges in Information Science (RCIS), 2015 IEEE 9th International Conference on, 2015, pp. 480–485.

[50] F. Färber, S. K. Cha, J. Primsch, C. Bornhövd, S. Sigg, and W. Lehner, “SAP HANA Database:

Data Management for Modern Business Applications," SIGMOD Rec., vol. 40, no. 4, pp. 45–51, 2012.

[51] K. M. A. Hasan, M. T. Omar, S. M. M. Ahsan, and N. Nahar, "Chunking implementation of extendible array to handle address space overflow for large multidimensional data sets," in Electrical Information and Communication Technology (EICT), 2013 International Conference on, 2014, pp. 1– 6.

[52] S. Müller and H. Plattner, "An In-depth Analysis of Data Aggregation Cost Factors in a Columnar Inmemory Database," in Proceedings of the Fifteenth International Workshop on Data Warehousing and OLAP, 2012, pp. 65–72.

[53] H. Plattner, "A Common Database Approach for OLTP and OLAP Using an In-memory Column Database," in Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, 2009, pp. 1–2.

[54] J. Schaffner, A. Bog, J. Krüger, and A. Zeier, "A Hybrid Row-Column OLTP Database Architecture for Operational Reporting," in Business Intelligence for the Real-Time Enterprise SE - 5, vol. 27, M. Castellanos, U. Dayal, and T. Sellis, Eds. Springer Berlin Heidelberg, 2009, pp. 61–74.

[55] V. K. Vavilapalli, A. C. Murthy, C. Douglas, S. Agarwal, M. Konar, R. Evans, T. Graves, J. Lowe, H. Shah, S. Seth, B. Saha, C. Curino, O. O'Malley, S. Radia, B. Reed, and E. Baldeschwieler, "Apache Hadoop YARN: Yet Another Resource Negotiator," in Proceedings of the 4th Annual Symposium on Cloud Computing, 2013, pp. 5:1–5:16.

[56] "Apache Pig Philosophy." [Online]. Available: http://pig.apache.org/philosophy.html. [Accessed: 26- Mar-2015].

[57] "Architecture - Apache Drill." [Online]. Available: http://drill.apache.org/architecture/. [Accessed: 26-Mar-2015].

[58] "Storm, distributed and fault-tolerant realtime computation." [Online]. Available: https://storm.apache.org/. [Accessed: 26-Mar-2015].

[59] "Apache Hive TM." [Online]. Available: https://hive.apache.org/. [Accessed: 26-Mar-2015]. [60] "Sqoop -." [Online]. Available: http://sqoop.apache.org/. [Accessed: 26-Mar-2015].

[61] "Impala." [Online]. Available: http://www.cloudera.com/content/cloudera/en/products-andservices/cdh/impala.html. [Accessed: 26-Mar-2015].

[62] "Apache Thrift - Home." [Online]. Available: https://thrift.apache.org/. [Accessed: 26-Mar-2015].

[63] "Apache ZooKeeper - Home." [Online]. Available: https://zookeeper.apache.org/. [Accessed: 26-Mar-2015].

[64] D. Borthakur, J. Gray, J. Sen Sarma, K. Muthukkaruppan, N. Spiegelberg, H. Kuang, K. Ranganathan, D. Molkov, A. Menon, S. Rash, R. Schmidt, and A. Aiyer, "Apache Hadoop Goes Realtime at Facebook," in Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, 2011, pp. 1071–1080.

[65] B. Ghit, A. Iosup, and D. Epema, "Towards an Optimized Big Data Processing System," in Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on, 2013, pp. 83–86.

[66] P. Agarwal, G. Shroff, and P. Malhotra, “Approximate Incremental Big-Data Harmonization,” in Big Data (BigData Congress), 2013 IEEE International Congress on, 2013, pp. 118–125.

[67] Y. Elshater, P. Martin, D. Rope, M. McRoberts, and C. Statchuk, “A Study of Data Locality in YARN,” 2015 IEEE Int. Congr. Big Data, pp. 174–181, 2015.

[68] A. H. B. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, “Big data: The next frontier for innovation, competition, and productivity,” McKinsey Glob. Inst., no. June, p. 156, 2011.

[69] J. S. Marron, “Big Data in context and robustness against heterogeneity,” Econom. Stat., vol. 2, pp. 73–80, 2017.

[70] L. Kugler, “What Happens When Big Data Blunders?,” Commun. ACM, vol. 59, no. 6, pp. 15–16, 2016.

[71] S. Sagiroglu, R. Terzi, Y. Canbay, and I. Colak, “Big data issues in smart grid systems,” in 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA), 2016, pp. 1007–1012.

[72] A. Gandomi and M. Haider, “Beyond the hype: Big data concepts, methods, and analytics,” Int. J. Inf. Manage., vol. 35, no. 2, pp. 137–144, 2015.

[73] Jameela Al-Jaroodi, Brandon Hollein, Nader Mohamed, "Applying software engineering processes for big data analytics applications development", Computing and Communication Workshop and Conference (CCWC) 2017 IEEE 7th Annual, pp. 1-7, 2017.

FUTURE AND CHALLENGES OF INTERNET OF THINGS

Falguni Jindal[1] , Rishabh Jamar[2] , Prathamesh Churi[3]

[1,2] Bachelors of Technology in Computer Engineering SVKM's NMIMS Mukesh Patel School of Technology Management and Engineering, Mumbai, India
[3] Assistant Professor (Computer Engineering) SVKM's NMIMS Mukesh Patel School of Technology Management and Engineering, Mumbai, India

ABSTRACT

The world is moving forward at a fast pace, and the credit goes to ever growing technology. One such concept is IOT (Internet of things) with which automation is no longer a virtual reality. IOT connects various non-living objects through the internet and enables them to share information with their community network to automate processes for humans and makes their lives easier. The paper presents the future challenges of IoT , such as the technical (connectivity , compatibility and longevity , standards , intelligent analysis and actions , security), business (investment , modest revenue model etc.), societal (changing demands , new devices, expense, customer confidence etc.) and legal challenges (laws, regulations, procedures, policies etc.). A section also discusses the various myths that might hamper the progress of IOT, security of data being the most critical factor of all. An optimistic approach to people in adopting the unfolding changes brought by IOT will also help in its growth.

KEYWORDS

IoT, Internet of Things, Security, Sensors.

Volume URL : https://airccse.org/journal/ijcsit2018_curr.html

Source URL : https://aircconline.com/ijcsit/V10N2/10218ijcsit02.pdf

REFERENCES

[1] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.

[2] Li, S., Da Xu, L., & Zhao, S. (2015). The internet of things: a survey. Information Systems Frontiers, 17(2), 243-259.

[3] Guo, B., Zhang, D., Wang, Z., Yu, Z., & Zhou, X. (2013). Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things. Journal of Network and Computer Applications, 36(6), 1531-1539.

[4] Banafa, A. (2014). IoT Standardization and Implementation Challenges. IEEE. org Newsletter.

[5] Banafa, A. (2015). „What is next for IoT and IIoT". Enterprise Mobility Summit

[6] Coetzee, L., & Eksteen, J. (2011, May). The Internet of Things-promise for the future? An introduction. In IST-Africa Conference Proceedings, 2011 (pp. 1-9). IEEE.

[7] Cai, H., Da Xu, L., Xu, B., Xie, C., Qin, S., & Jiang, L. (2014). IoT-based configurable information service platform for product lifecycle management. IEEE Transactions on Industrial Informatics, 10(2), 1558-1567.

[8] Khan, R., Khan, S. U., Zaheer, R., & Khan, S. (2012, December). Future internet: the internet of things architecture, possible applications and key challenges. In Frontiers of Information Technology (FIT), 2012 10th International Conference on (pp. 257-260). IEEE.

[9] Liu, Y., & Zhou, G. (2012, January). Key technologies and applications of internet of things. In Intelligent Computation Technology and Automation (ICICTA), 2012 Fifth International Conference on (pp. 197-200). IEEE.

[10] Sadeghi, A. R., Wachsmann, C., & Waidner, M. (2015, June). Security and privacy challenges in industrial internet of things. In Proceedings of the 52nd annual design automation conference (p. 54). ACM.

[11] Banafa, A. (2014). IoT and Blockchain Convergence: Benefits and Challenges. IEEE Internet of Things.

[12] Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT data analytics: Architecture, opportunities, and open research challenges. IEEE Access, 5, 5247-5261.

[13] Desai, P., Sheth, A., & Anantharam, P. (2015, June). Semantic gateway as a service architecture for iot interoperability. In Mobile Services (MS), 2015 IEEE International Conference on(pp. 313-319). IEEE.

[14] Koivu, A., Koivunen, L., Hosseinzadeh, S., Laurén, S., Hyrynsalmi, S., Rauti, S., & Leppänen, V. (2016, December). Software Security Considerations for IoT. In Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2016 IEEE International Conference on (pp. 392-397). IEEE.

[15] Sundmaeker, H., Guillemin, P., Friess, P., & Woelfflé, S. (2010). Vision and challenges for

realising the Internet of Things. Cluster of European Research Projects on the Internet of Things, European Commision, 3(3), 34-36.

[16] Vermesan, O., Friess, P., Guillemin, P., Gusmeroli, S., Sundmaeker, H., Bassi, A., ... & Doody, P. (2011). Internet of things strategic research roadmap. Internet of Things-Global Technological and Societal Trends, 1(2011), 9-52.

[17] Sheng, Z., Yang, S., Yu, Y., Vasilakos, A., Mccann, J., & Leung, K. (2013). A survey on the ietf protocol suite for the internet of things: Standards, challenges, and opportunities. IEEE Wireless Communications, 20(6), 91-98.

[18] Theoleyre, F., & Pang, A. C. (Eds.). (2013). Internet of Things and M2M Communications. River Publishers.

[19] Coetzee, L., & Eksteen, J. (2011, May). The Internet of Things-promise for the future? An introduction. In IST-Africa Conference Proceedings, 2011 (pp. 1-9). IEEE. International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018 25

[20] Ji, Z., & Anwen, Q. (2010, November). The application of internet of things (IOT) in emergency management system in China. In Technologies for Homeland Security (HST), 2010 IEEE International Conference on (pp. 139-142). IEEE.

[21] James Kirkland , “Internet of Things: insights from Red Hat” , Website: https://developers.redhat.com/blog/2015/03/31/internet-of-things-insights-from-red-hat/ , Accesed : 2nd February 2018

AUTHORS

Falguni Jindal is a final year student pursuing B.Tech in Computer Science from SVKM's NMIMS Mukesh Patel School of Technology Management and Engineering (MPSTME), Mumbai, India. She is a passionate student and has a strong determination for gathering knowledge and learning new things every day. Falguni has published two research papers in the field of IOT and Web Security respectively. Currently, she is also working on a few other projects in other domains of Computer Science.

Rishabh Jamar is a final year student pursuing B.Tech in Computer Science from SVKM's NMIMS Mukesh Patel School of Technology Management and Engineering (MPSTME), Mumbai, India. He is hard working, enthusiastic and his quest for more knowledge led him to gain interest in exploring new domains like Network Security, Artificial Intelligence, Data Analytics and Internet of Things. He has published four research papers in the same fields at national and International level. He has also done a major project on internet security and several other minor projects in different domains of Computer Science.

Prof. Prathamesh Churi is Assistant Professor in Computer Engineering Department of SVKM's NMIMS Mukesh Patel School of Technology Management and Engineering (MPSTME), Mumbai, India. He has Completed his Bachelor's degree in Engineering (Computer science) from University of Mumbai and completed his Master's Degree in Engineering (Information Technology) from University of Mumbai. He started his journey as a professor and has been working successfully in this field since past 3 years where outcome of learning is different for every day. He is having outstanding technical knowledge in the field of Network Security and Cryptography, Education Technology, Internet of Things. He has published many research papers in the same field at national and International level. He is a reviewer, TPC member, Session Chair, guest speaker of many IEEE/ Springer Conferences and Institutes at International Level. . He has bagged with many awards in the education field. His relaxation and change lies in pursuing his hobbies which mainly includes expressing views be it in public ¬writing columns or blogging.

CLUSTERING ALGORITHM FOR A HEALTHCARE DATASET USING SILHOUETTE SCORE VALUE

Godwin Ogbuabor[1] and Ugwoke, F. N[2]

[1] School of Computer Science, University of Lincoln, United Kingdom
[2]Department of Computer Science, Michael Okpara University of Agriculture Umudike, Abia State, Nigeria

ABSTRACT

The huge amount of healthcare data, coupled with the need for data analysis tools has made data mining interesting research areas. Data mining tools and techniques help to discover and understand hidden patterns in a dataset which may not be possible by mainly visualization of the data. Selecting appropriate clustering method and optimal number of clusters in healthcare data can be confusing and difficult most times. Presently, a large number of clustering algorithms are available for clustering healthcare data, but it is very difficult for people with little knowledge of data mining to choose suitable clustering algorithms. This paper aims to analyze clustering techniques using healthcare dataset, in order to determine suitable algorithms which can bring the optimized group clusters. Performances of two clustering algorithms (Kmeans and DBSCAN) were compared using Silhouette score values. Firstly, we analyzed K-means algorithm using different number of clusters (K) and different distance metrics. Secondly, we analyzed DBSCAN algorithm using different minimum number of points required to form a cluster (minPts) and different distance metrics. The experimental result indicates that both K-means and DBSCAN algorithms have strong intra-cluster cohesion and inter-cluster separation. Based on the analysis, K-means algorithm performed better compare to DBSCAN algorithm in terms of clustering accuracy and execution time.

KEYWORDS

Dataset, Clustering, Healthcare data, Silhouette score value, K-means, DBSCAN

Volume URL : https://airccse.org/journal/ijcsit2018_curr.html

Source URL : https://aircconline.com/ijcsit/V10N2/10218ijcsit03.pdf

REFERENCES

[1] Alsayat, A., & El-Sayed, H. (2016). Efficient genetic K-Means clustering for health care knowledge discovery. In Software Engineering Research, Management and Applications (SERA), 2016 IEEE 14th International Conference on (pp. 45-52). IEEE.

[2] Balasubramanian, T., & Umarani, R. (2012, March). An analysis on the impact of fluoride in human health (dental) using clustering data mining technique. In Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on (pp. 370-375). IEEE.

[3] Banu G. Rasitha & Jamala J.H.Bousal (2015). Perdicting Heart Attack using Fuzzy C Means Clustering Algorithm. International Journal of Latest Trends in Engineering and Technology (IJLTET).

[4] Banu, M. N., & Gomathy, B. (2014). Disease forecasting system using data mining methods. In Intelligent Computing Applications (ICICA), 2014 International Conference on (pp. 130-133). IEEE.

[5] Belciug, S. (2009). Patients length of stay grouping using the hierarchical clustering algorithm. Annals of the University of Craiova-Mathematics and Computer Science Series, 36(2), 79-84.

[6] Belciug, S., Salem, A. B., Gorunescu, F., & Gorunescu, M. (2010, November). Clustering-based approach for detecting breast cancer recurrence. In Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on (pp. 533-538). IEEE.

[7] Bruno, G., Cerquitelli, T., Chiusano, S., & Xiao, X. (2014). A clustering-based approach to analyse examinations for diabetic patients. In Healthcare Informatics (ICHI), 2014 IEEE International Conference on (pp. 45-50). IEEE.

[8] DeFreitas, K., & Bernard, M. (2015). Comparative performance analysis of clustering techniques in educational data mining. IADIS International Journal on Computer Science & Information Systems, 10(2).

[9] Escudero, J., Zajicek, J. P., & Ifeachor, E. (2011). Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (pp. 6470-6473). IEEE.

[10] Han, J., Kamber, M., & Pei, J. (2012). Cluster Analysis-10: Basic Concepts and Methods.

[11] Ibrahim, N. H., Mustapha, A., Rosli, R., & Helmee, N. H. (2013). A hybrid model of hierarchical clustering and decision tree for rule-based classification of diabetic patients. International Journal of Engineering and Technology (IJET), 5(5), 3986-91.

[12] Jabel K. Merlin & Srividhya (2016). Performance analysis of clustering algorithms on heart dataset. International Journal of Modern Computer Science, 5(4), 113-117. International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018 37

[13] Kar Amit Kumar, Shailesh Kumar Patel & Rajkishor Yadav (2016). A Comparative Study & Performance Evaluation of Different Clustering Techniques in Data Mining. ACEIT Conference Proceeding.

[14] Lv, Y., Ma, T., Tang, M., Cao, J., Tian, Y., Al-Dhelaan, A., & Al-Rodhaan, M. (2016). An

efficient and scalable density-based clustering algorithm for datasets with complex structures. Neurocomputing, 171, 9-22. [15] Malli, S., Nagesh, H. R., & Joshi, H. G. (2014). A Study on Rural Health care Data sets using Clustering Algorithms. International Journal of Engineering Research and Applications, 3(8), 517- 520.

[16] Maulik, U., & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(12), 1650-1654.

[17] Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on (pp. 63-67). IEEE.

[18] Paul, R., & Hoque, A. S. M. L. (2010, July). Clustering medical data to predict the likelihood of diseases. In Digital Information Management (ICDIM), 2010 Fifth International Conference on (pp. 44-49). IEEE.

[19] Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 219(1), 103-119.

[20] R.Nithya & P.Manikandan & D.Ramyachitra (2015); Analysis of clustering technique for the diabetes dataset using the training set parameter. International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 9.

[21] Sagar, H. K., & Sharma, V. (2014). Error Evaluation on K-Means and Hierarchical Clustering with Effect of Distance Functions for Iris Dataset. International Journal of Computer Applications, 86(16).

[22] Shah, G. H., Bhensdadia, C. K., & Ganatra, A. P. (2012). An empirical evaluation of density-based clustering techniques. International Journal of Soft Computing and Engineering (IJSCE) ISSN, 22312307, 216-223.

[23] Tan, P. N., Steinbach, M., & Kumar, V. (2013). Data mining cluster analysis: basic concepts and algorithms. Introduction to data mining.

[24] Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266.

[25] Vijayarani, S., & Sudha, S. (2015). An efficient clustering algorithm for predicting diseases from hemogram blood test samples. Indian Journal of Science and Technology, 8(17).

[26] Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications, 41(4), 1476-1482.

AN ALGORITHM FOR PREDICTIVE DATA MINING APPROACH IN MEDICAL DIAGNOSIS

Shakuntala Jatav[1] and Vivek Sharma[2]

[1]M.Tech Scholar, Department of CSE, TIT College, Bhopal
[2] Professor, Department of CSE, TIT College, Bhopal

ABSTRACT

The Healthcare industry contains big and complex data that may be required in order to discover fascinating pattern of diseases & makes effective decisions with the help of different machine learning techniques. Advanced data mining techniques are used to discover knowledge in database and for medical research. This paper has analyzed prediction systems for Diabetes, Kidney and Liver disease using more number of input attributes. The data mining classification techniques, namely Support Vector Machine(SVM) and Random Forest (RF) are analyzed on Diabetes, Kidney and Liver disease database. The performance of these techniques is compared, based on precision, recall, accuracy, f_measure as well as time. As a result of study the proposed algorithm is designed using SVM and RF algorithm and the experimental result shows the accuracy of 99.35%, 99.37 and 99.14 on diabetes, kidney and liver disease respectively.

KEYWORDS

Data Mining, Clinical Decision Support System, Disease Prediction, Classification, SVM, RF.

Volume URL : https://airccse.org/journal/ijcsit2019_curr.html

Source URL : https://aircconline.com/ijcsit/V10N1/10118ijcsit02.pdf

REFERENCES

[1] Nidhi Bhatla, Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques",IJERT,Vol 1, Issue 8, 2012.

[2] Syed Umar Amin, Kavita Agarwal, Rizwan Beg, "Genetic Neural Network based Data Mining in Prediction of Heart Disease using Risk Factors", IEEE, 2013.

[3] A H Chen, S Y Huang, P S Hong, C H Cheng, E J Lin, "HDPS: Heart Disease Prediction System", IEEE, 2011.

[4] M. Akhil Jabbar, B. L Deekshatulu, Priti Chandra, "Heart Disease Prediction using Lazy Associative Classification", IEEE, 2013.

[5] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", IJCA, Volume 47– No.10, June 2012. [6] P. Bhandari, S. Yadav, S. Mote, D.Rankhambe, "Predictive System for Medical Diagnosis with Expertise Analysis", IJESC, Vol. 6, pp. 4652-4656, 2016.

[7] Nishara Banu, Gomathy, "Disease Forecasting System using Data Mining Methods", IEEE Transaction on Intelligent Computing Applications, 2014.

[8] A. Iyer, S. Jeyalatha and R. Sumbaly, "Diagnosis of Diabetes using Classification Mining Techniques", IJDKP, Vol. 5, pp. 1-14, 2015.

[9] Sadiyah Noor Novita Alfisahrin and Teddy Mantoro, "Data Mining Techniques for Optimatization of Liver Disease Classification", International Conference on Advanced Computer Science Applications and Technologies, IEEE, pp. 379-384, 2013.

[10] A. Naik and L. Samant, "Correlation Review of Classification Algorithm using Data Mining Tool: WEKA, Rapidminer , Tanagra ,Orange and Knime", ELSEVIER, Vol. 85, pp. 662-668, 2016.

[11] Uma Ojha and Savita Goel, "A study on prediction of breast cancer recurrence using data mining techniques", International Conference on Cloud Computing, Data Science & Engineering, IEEE, 2017.

[12] Naganna Chetty, Kunwar Singh Vaisla, Nagamma Patil, "An Improved Method for Disease Prediction using Fuzzy Approach", International Conference on Advances in Computing and Communication Engineering, IEEE, pp. 568-572, 2015.

[13] Kumari Deepika and Dr. S. Seema, "Predictive Analytics to Prevent and Control Chronic Diseases", International Conference on Applied and Theoretical Computing and Communication Technology, IEEE, pp. 381-386, 2016.

[14] Emrana Kabir Hashi, Md. Shahid Uz Zaman and Md. Rokibul Hasan, "An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques", IEEE, 2017.