How to Cite:
Vidyasri, P., Rohit, K. B., & Shridhar, A. (2022). Analysis and application of the sports education information data using big data. *International Journal of Health Sciences*, *6*(S1), 9225–9233. https://doi.org/10.53730/ijhs.v6nS1.7089

Analysis and application of the sports education information data using big data

Ms. P. Vidyasri
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai
Email: vidyasrp@srmist.edu.in

K. B. Rohit
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai
Email: rk8838@srmist.edu.in

A. Shridhar
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai
Email: sa6540@srmist.edu.in

***Abstract*---**The education of physical education has changed as a result of big data. To promote the university physical education modernization, this paper points out that big data should be used in the management of sports information in colleges and universities. Apriori algorithm using MapReduce model is proposed to accelerate the calculation and storage capacity of management data. The system also provides various data API interfaces, which integrate all kinds of sports resources in high schools and various institutes, and promote the accuracy of association rule mining. To build a sports big data environment and enhance the decision-making ability and intelligent level of sports management information system, an improvement strategy is proposed.

***Keywords*---**sports education, information data, big data.

Introduction

Because of the development of big data technology, analysis is being done on a large scale in many companies. This is advantageous to some companies, but it can also be disadvantageous to some. Those big data are used in education, new opportunities arise while challenges appear as well. It is necessary to have efficient information systems in schools. This is where sports information

management comes into play. Big Data Analysis and Sports Information Management will help advance the school's education and physical education quality. Because it has become more and more important to understand the value of big data when managing sports, students and administrators alike need to develop new strategies in order to master the complex issue. The problem can be solved by using high technology to better track and predict the success of athletes. What's more, a university sports management information system can provide real-time data for both students and administrators. Our survey revealed that the system's performance and efficiency have risen [1], which is more in line with the needs of university sports management. This has laid the foundation for the application of big data and the in-depth participation and development of university sports. And promoted the social movement of the whole people.

Related works

Today, lots of data is stored in different databases and storage devices. To analyze this data, it is necessary to collect this data into a unified format. XML files are open and flexible, and they can be used independently. They are widely used to describe and store data. Internet services benefit from the use of flexibility, security, dynamic and efficiency to integrate inferior data with the advantages of independence, flexibility, security, [2] dynamics and efficiency. Data sources can be flexibly adapted to external systems and include impurities from a single system unmatched Environment big data system, [3] which uses HBase paper to effectively transfer the latest data attacks. It uses web services technology to capture different data from multiple systems, analyze it in Hadoop platform by capturing relevant data as an XML document from the source system. The document will then be divided into databases where each database contains relevant information.

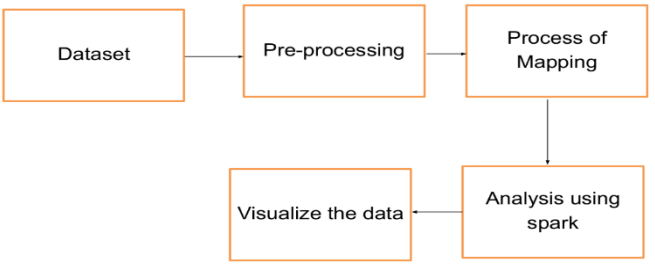

Fig 1. Representation in block diagram

The system is developed through the Hadoop framework. Spring development web services and the Mibatis technology are used in development. External data sources, historical data, and the intelligent data network can be deleted, transformed, merged and summarized in real time. In addition to separate payment data, the system can stimulate resources with intelligent calculation algorithm and manage resource allocation according to time and different applications. [4] The system can apply all kinds of learning resources to students, teachers, and principals for demonstration. The block diagram from Fig 1. And the system architecture from Fig 2. Shows the overall structure of the system.

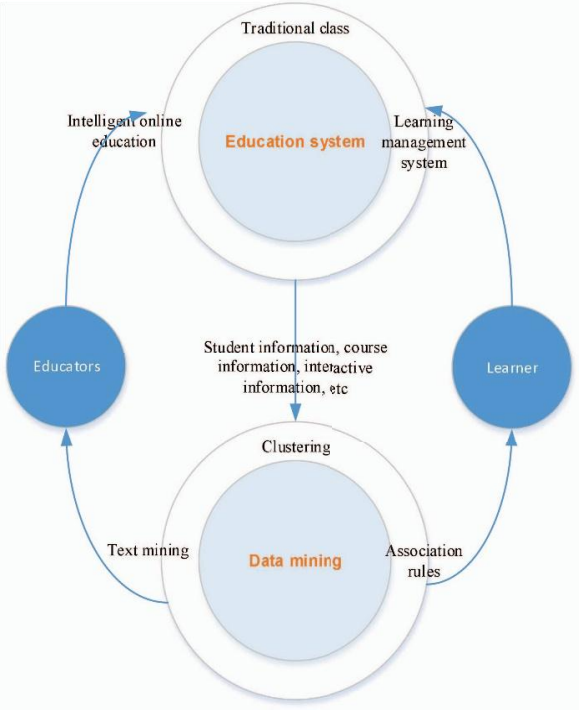

Fig 2. System Architecture

Datasets

A dataset was used that contains academic review data from Amazon.com. [5][16] A fake dataset of Yelp data was also used for additional practice in model implementation. The size of the Amazon dataset is much larger than the Yelp dataset, and it is more difficult to train an algorithm using this data.

Pre-processing

Pre-processing is the foremost and essential step in any dataset analysis [6], which includes removing unwanted attributes, punctuations, stop words, and redundant words. This certifies proper training of the model.

Feature Extraction

This process involves almost all the methods that eliminate the dataset's unwanted information.[7] This step is very vital for finding gaps and relationships between attributes and use them to filter valid conclusions. The libraries from NLTK package are used to construct corpus of words. [13] The stop words like then, to be removed and other unnecessary words like is, why etc. added by spammers on a regular basis are grouped under stop words in the English language.

Sampling Data

Because the dataset includes many reviews, it must be processed before feeding to the classifier. Sampling is used to assign different labels to authentic fake reviews and then create a new column of these labels. The dataset is then returned as a data frame.

Algorithms used

Naïve Bayes Algorithm

Naive Bayes is a classification algorithm that utilizes conditional independence tests to determine whether two events are likely to be related. Given an event and a set of possible causes, this algorithm calculates the probability that the event is caused by each cause. However, if it finds that the probability of the event being caused by a cause is above some threshold, then it considers that cause as not having caused the event.[8] By iteratively removing causes that do not meet the threshold for causation, it can arrive at a more accurate estimation of the probability of an event being caused by a specific reason.

To find the probability of an event Y given that X has occurred, [9] we must first calculate the prior probability P(Y) of Y.[10][15] Next we must consider the evidence provided by X, which is an attribute value of an unknown instance, and solve P(Y|X), or the posterior probability of Y.

Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning [11]. It performs a regression task. Regression models a target prediction value based on independent variables. [12] It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on [14] – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used.

Equations

Naïve Bayes Algorithm:

$$P\left(\frac{y}{X}\right) = P\left(\frac{X}{y}\right) P(y)/P(X)$$

Y - class variable

X - feature dependent vector

Linear Regression:

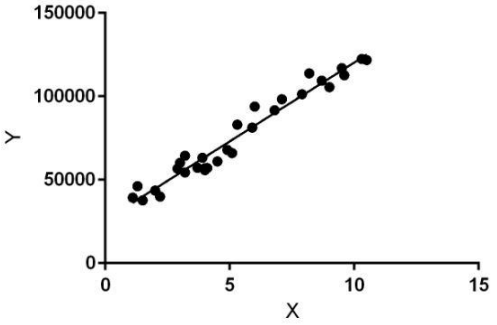

Fig 3. Graphical representation of Linear regression

$$Y = \theta_1 + \theta_2.X$$

x: input training data (univariate – one input variable(parameter))
y: labels to data (supervised learning)
θ_1: intercept
θ_2: coefficient of x

Experiment result

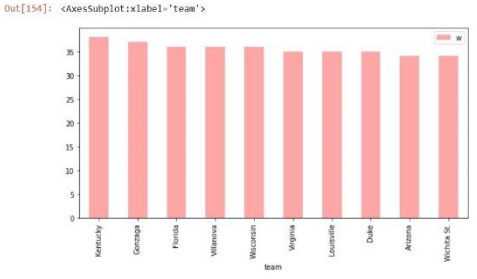

Fig 4. Teams with maximum winnings

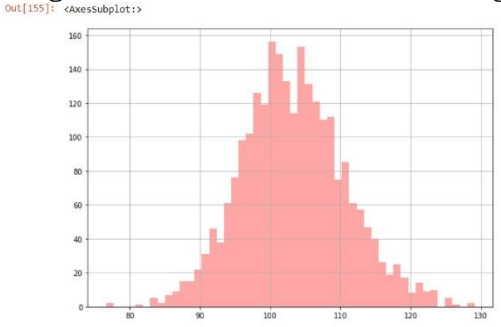

Fig 5. Team's adjacent offensive efficiency

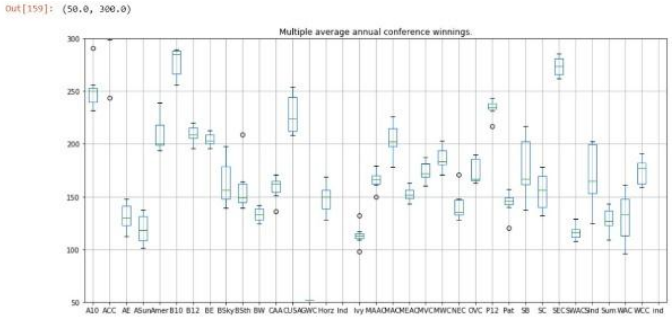

Fig 6. Multiple average annual conference winnings

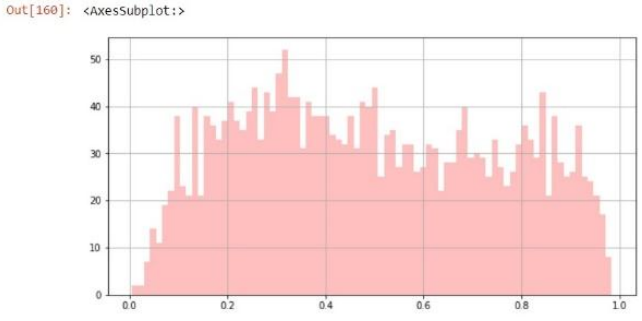

Fig 7. Graphical representation of Barthag

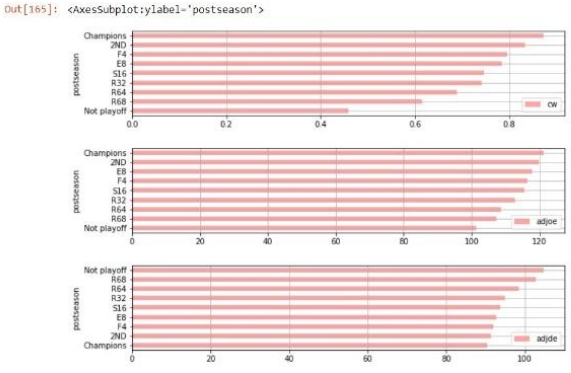

Fig 8. Postseason's winning result of three prime factors

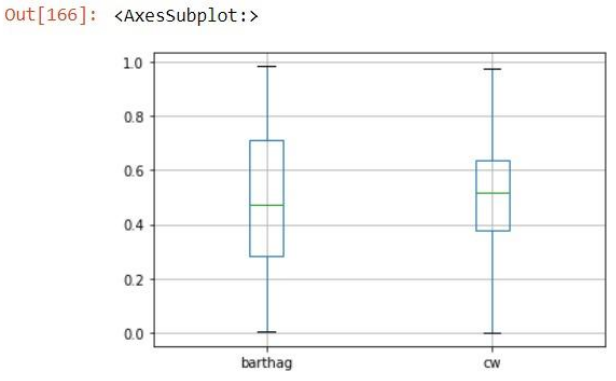

Fig 9. Graphical comparison between barthag and cw(total games played/total wins)

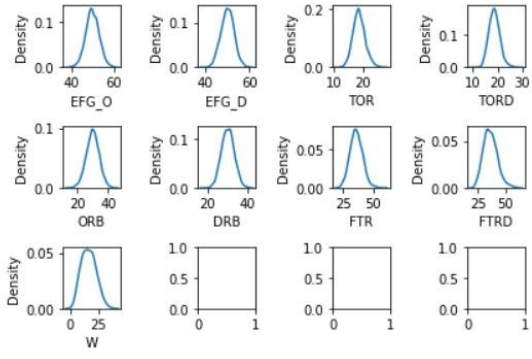

Fig 10. Average results from main factors

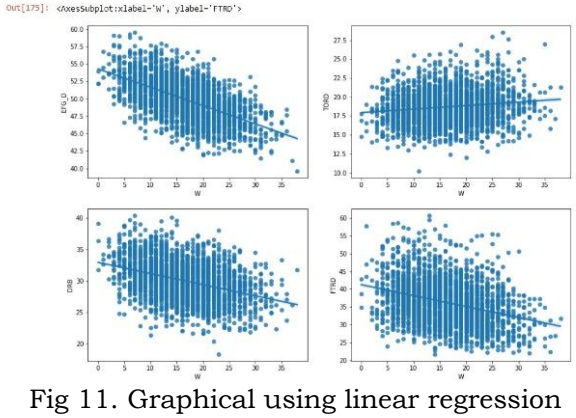

Fig 11. Graphical using linear regression

Out[184]:

	Actual	Predicted	AbsDiff
2072	22	22.003795	0.003795
2125	23	22.996072	0.003928
1018	22	22.004891	0.004891
342	19	18.992386	0.007614
1644	11	10.987165	0.012835
557	13	13.013088	0.013088
2139	24	24.019892	0.019892
2378	25	25.029728	0.029728
1422	23	23.034919	0.034919
465	14	14.036557	0.036557

Fig 12. Comparison with actual result with our predicted ones

Conclusion

Personalized learning is an educational approach related to domains like Artificial Intelligence (AI) and Machine Learning (ML) technology. Learners collect data and analyze the basics, visualize the results, and spend more time on physical activities, practical experiences, and interactions with teachers. School administrators can use AI-based teacher support to report real-time student status and present various alternatives to students when they are having difficulty in class. The AI technology also helps administrators manage evaluation and lessons more efficiently. Finally, AI helps educators by reducing time spent on administrative tasks so they have more time to invest in improving teaching quality.

Reference

[1] Weed, Mike. After 20 years, what are the Big Questions for sports tourism research? Journal of Sport Tourism, 2014, 19(1):1-4

[2] Ispa-Landa S. Gender, Race, and Justifications for Group Exclusion Urban Black Students Bussed to Affluent Suburban Schools.Sociology of Education, 2013, 86(3):218-233

[3] Wong L H, Looi C K. What seams do we remove in mobile-assisted seamless learning? A critical review of the literature. Computers & Education, 2011, 57(4):2364-2381

[4] Almansoori K A, Clark M. Research Article Increasing Trends in Orthopedic Fellowships Are Not due to Inadequate Residency Training. Education Research International, 2014, 2015(2015):215- 226

[5] Patel D, Shah D, Shah M. The Intertwine of Brain and Body: A Quantitative Analysis on How Big Data Influences the System of Sports. Annals of Data Science, 2020, 7(1):1-16

[6] S. M. Marcora, W. Staiano, and V. Manning, “Mental fatigue impairs physical performance in humans,” J. Appl. Physiol., vol. 106, no. 3, pp. 857–864, 2009.

[7] Karnan, B., Kuppusamy, A., Latchoumi, T. P., Banerjee, A., Sinha, A., Biswas, A., & Subramanian, A. K. (2022). Multi-response Optimization of Turning Parameters for Cryogenically Treated and Tempered WC–Co Inserts. Journal of The Institution of Engineers (India): Series D, 1-12.

[8] Banu, J. F., Muneeshwari, P., Raja, K., Suresh, S., Latchoumi, T. P., & Deepan, S. (2022, January). Ontology-Based Image Retrieval by Utilizing Model Annotations and Content. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 300-305). IEEE.

[9] Latchoumi, T. P., & Parthiban, L. (2021). Quasi oppositional dragonfly algorithm for load balancing in a cloud computing environment. Wireless Personal Communications, 1-18.

[10] Garikapati, P. R., Balamurugan, K., Latchoumi, T. P., & Shankar, G. (2022). A Quantitative Study of Small Dataset Machining by Agglomerative Hierarchical Cluster and K-Medoid. In Emergent Converging Technologies and Biomedical Systems (pp. 717-727). Springer, Singapore.

[11] I. Kathner, S. C. Wriessnegger, G. R. M ¨ uller-Putz, A. K ¨ ubler, and ¨ S. Halder, “Effects of mental workload and fatigue on the p300, alpha and theta band power during operation of an erp (p300) brain–computer interface,” Biol. Psychol., vol. 102, pp. 118–129, 2014.

[12] Latchoumi, T. P., Kalusuraman, G., Banu, J. F., Yookesh, T. L., Ezhilarasi, T. P., & Balamurugan, K. (2021, November). Enhancement in manufacturing systems using Grey-Fuzzy and LK-SVM approach. In 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT) (pp. 72-78). IEEE.

[13] Latchoumi, T. P., Swathi, R., Vidyasri, P., & Balamurugan, K. (2022, March). Develop New Algorithm To Improve Safety On WMSN In Health Disease Monitoring. In 2022 International Mobile and Embedded Technology Conference (MECON) (pp. 357-362). IEEE.

[14] C. Zhao, M. Zhao, Y. Yang, J. Gao, N. Rao, and P. Lin, “The reorganization of human brain networks modulated by driving mental fatigue,” IEEE J. Biomed. Health, vol. 21, no. 3, pp. 743–755, 2016.

[15] Y. Yang, Z. Gao, Y. Li, Q. Cai, N. Marwan, and J. Kurths, “A complex network-based broad learning system for detecting driver fatigue from eeg signals,” IEEE T. Syst. Man, Cy-S., 2019.

[16] Z. Gao, S. Li, Q. Cai, W. Dang, Y. Yang, C. Mu, and P. Hui, “Relative wavelet entropy complex network for improving eeg-based fatigue driving classification,” IEEE T. Instrum. Meas., vol. 68, no. 7, pp. 2491– 2497, 2018.