A review of the academic achievement of students utilising large-scale data analysis

Prasadu Peddi, Dr. Akash Saxena

Research Scholar, Arni University, Kangra district, Himachal Pradesh

Research Guide, Arni University, Kangra district, Himachal Pradesh

ABSTRACT: Accurate projections are crucial for obtaining dependable outcomes in education, especially when it comes to estimating student performance. Numerous factors can cause significant variations in a student's performance. Therefore, it is critical to have accurate and trustworthy market trend assessments. Predicting a student's future performance can yield substantial advantages. Advanced machine learning and deep learning approaches are needed to comprehend and predict complex non-linear graphs. With the use of Hadoop and deep learning, this project aims to forecast future tuition costs and track student achievement. After conducting a thorough examination and analyzing pertinent literature, we have discovered a number of approaches for forecasting trends in the educational market. We looked at the various types of neural networks and how they function. We have thoroughly investigated numerous methods and a wide range of prediction tools during this study. We have created a deep learning model that can forecast future student performance using this knowledge. There were several approaches used. The process of gathering data from prior years started. After that, we investigated several methods for preprocessing data and went ahead and did so in order to get the data ready for our model. We have determined the opening price, peak price, and volume as the three main determinants of student prices. After that, we constructed the neural network model, tuned it using various hyper-parameters to enhance its functionality, and employed it to forecast future values. All things considered, this task offers a fantastic opportunity to learn. This idea allows us to forecast outcomes in high-stakes scenarios where huge quantities of money are at stake. The project is developed using a multitude of tools and technologies. Several Python libraries were utilized in order to finish this project. Panda and NumPy are often used in data collection and preparation. Keras is used to build the neural network and to carry out the technique. Matplotlib is used to visualize the results.

Keywords: Machine Learning, Big data, Student Performance.

INTRODUCTION

The objective of education is to assist individuals in improving their behaviour by imparting new skills and broadening their existing knowledge base through various learning settings, both formal and informal. It empowers individuals to remain resilient in challenging situations and equips them with the necessary skills to achieve independence, self-reliance, and sustainability. It assists individuals in managing health concerns, adopting healthier eating habits, locating superior child care options, and saving for the future. Consequently, the preservation of human values that benefit individuals and society at large is ensured. Therefore, it plays a crucial role in opening up opportunities for improved professional prospects, elevated social status, and individual satisfaction. It empowers individuals to increase their productivity, assume greater responsibilities in the economy, and improve their living conditions, which is vital for eradicating poverty. Education is crucial for human growth as it provides the foundation for continuous learning and is a fundamental requirement for a democratic society. Increased public awareness of rights and possibilities leads to citizens becoming more accountable and knowledgeable. Education plays a crucial role in empowering individuals to have a voice in government and society, serving as the key driver for a nation's progress.

Delving into the captivating world of modeling, predicting student performance, and analyzing educational trends has piqued interest of professionals from various fields, including academics, investors, and researchers. In realm of education, transactions occur where individuals exchange goods and services in return for monetary compensation. The study and forecast Educational market has garnered considerable attention as a result increasing adoption of artificial intelligence and Hadoop. Education field analysts utilize Hadoop and Map Reduce method to effectively analyze stocks, eliminating need for additional time-consuming infrastructure. The education sector is greatly impacted by the political and macroeconomic landscapes. The data used in education has a sample size that accurately represents real-world transactions. With the increasing amount of data, the educational setting becomes more uncertain during the testing and predictive phase. On one hand, a larger data size is associated with a longer transaction record. Instead of using daily data for this project, we leverage education data to minimize the potential for unclear interference and substantially enhance the sample size within a shorter duration. Extensive

research has been carried out in recent years on the application of machine learning and hadoop techniques in analyzing and predicting financial markets. When it comes to predicting student performance, neural networks have demonstrated superior effectiveness compared to previous models. In addition, Hadoop infrastructure has demonstrated exceptional efficiency in managing vast datasets and executing computations on them. They have a deep understanding of dynamically operating systems as well. Reprocessing the data with a new pattern is necessary for procedure, and the Map Reduce method effectively utilizes mappers and reducers to manage the data.

REVIEW OF LITERATURE

 This study aimed to evaluate the academic performance of first-grade students in Botswana who had received pre-school education, in comparison to their peers who had not. To achieve this, standardized examinations were administered in English, mathematics, language arts, and science. A total of 120 first graders from four different elementary schools in Botswana were selected randomly to participate in the research. The study's findings suggest that preschoolers outperformed their peers who did not attend preschool in all three academic domains.

A study was conducted on 400 students to evaluate the importance of personality tests, cognitive abilities, demographic variables, and mental processes in achieving success in upper secondary science programs. The group of students consisted of an equal number of males and females, all hailing from the high school at Aligarh Muslim University in Aligarh, India. The selection technique utilized cluster sampling, dividing the study's participants into smaller groups. Clusters were generated randomly to aid in future investigation. The study revealed that boys from less privileged backgrounds demonstrated a greater likelihood of achieving academic success, whereas girls from more affluent backgrounds displayed exceptional performance in the field of science. A study was conducted to investigate the impact of socioeconomic factors on the learning environments at home and school in [3]. The research sample included 240 students from eight co-ed schools in Dharwar, Karnataka, India. The students in this group were in grades 8 through 10 and attended schools that offered education in both English and Kannada languages. The interrelationships between characteristics such as high school students' socioeconomic position, academic achievement, and home and school contexts were examined using Karl Pearson's moment test of correlation by the researchers. Studies show that schools serving students from different socioeconomic backgrounds tend to use different languages for instruction. Schools with lower socioeconomic backgrounds often use Kannada as the primary language, while

schools with higher socioeconomic backgrounds primarily use English. Their findings indicated that students who were enrolled in schools where English was the main language of instruction demonstrated notably superior academic performance in comparison to students attending schools that used Kannada. The research in this field suggests that the way teachers deliver their lessons can have a significant impact on students' academic achievement.

[4] Conducted a survey of 300 undergraduate students (225 males and 75 females) to evaluate their academic achievement. The Punjab University of Pakistan has numerous affiliations with colleges.

The study conducted in [5] examined the fluctuations in learners' motivation and their learning strategies. The study involved a group of 246 business students who were participating in open-learning programs in Australia. A mathematical equation model has been used to construct models for various structural issues. The model reveals a significant finding regarding gender disparities in the learning processes of males and females. Utilized various data mining techniques to predict the academic progress of computer science students in Hellenic Open University's (HOU) online courses [6]. The machine learning techniques commonly used include perception learning, instance learning, decision learning, decision trees, Bayesian networks, and rule learning. These techniques rely on algorithms that represent the HOU (Hypothesis Output Unit). Machine learning algorithms have proven to be highly accurate in predicting students' performance on the final test. The researchers included demographic and performance-related data while constructing their model. There are seven factors that make up the demographic attribute sets: The variables of interest encompass individuals' sex, age, marital status, number of children who are employed, occupation, level of computer proficiency, and the correlation between students' employment and their access to computers. The performance attribute collection included four pairs of characteristics: students' grades on written assignments and their attendance or non-attendance during in-person discussions. The class (pass or fail) of each student was predicted using various techniques, including C4.5, Naive Bayes, 3-NN, RIBBER, and WINNOW. We evaluated the precision of the predictive model for each set of demographic attributes. We carefully integrated one performance indicator at a time for each category and evaluated the precision of our projected estimates. The report claimed that this method achieved the highest level of expected accuracy. The algorithms Naive Bayes and RIPPER were utilized to analyze data from the University of Malaya's accounting for beginners course with the aim of identifying the factors that influence students' grades. Female students are often seen as having more employment responsibilities, which may explain

their strong performance in accounting compared to male students. This research suggests that a strong grasp of high school accounting math, grades, and economics-related math can be a reliable predictor of academic success.

The study examined the difference in academic performance between males and females in math and language arts at the secondary school level, as well as other cognitive-motivational characteristics. The sample included 521 students, with 236 males and 285 females, aged 14-18, who were selected from the second round of compulsory secondary education in Jeon province. Hello, Jeon. An analysis was conducted to evaluate the difference in academic achievement between male and female secondary school students, based on their examination results. Researchers conducted a study that highlighted disparities in the academic performance of children, specifically based on their gender. The study conducted by [9] explored the relationship between students' fitness grades and their Stanford Achievement Test results, which are used to measure their intellectual aptitude. The survey encompassed a grand total of 884,715 students in California, spanning across the sixth, seventh, and ninth grades during the 2002 academic year. The performance of students was found to be correlated with their fitness levels. Research has shown that the physical fitness of children can have a significant impact on their overall health and performance. In addition, the research findings indicate that students from higher socioeconomic backgrounds (SES) demonstrated a stronger correlation between physical fitness and academic performance, in contrast to students from lower socioeconomic backgrounds (SES). In addition, this link was particularly strong when comparing females to males.

The e-learning and e-assessing system utilized Bayesian Networks to predict the performance of eighth grade mathematics students [10]. The Massachusetts Comprehensive Assessment System (MCAS) was a mandatory standardized test that evaluated students' proficiency in English, mathematics, science, and social studies from third to tenth grade. Teachers and students in Massachusetts greatly appreciate it as passing the test is a mandatory component for high school graduation.

The approach in question operated under the assumption that all talents were given equal consideration, without any prior knowledge of student responses. It was presented in a clear and understandable manner. The network quickly calculated the probability of the students' knowledge based on the evidence obtained.

Reference [11] documents the BioGRID database. The web interface provided a convenient way to search for and access data through internal links. The graphical representations were crafted to facilitate effortless retrieval of information. A graph was generated using an

information visualization system. Unfortunately, cross-species predictions were not conducted.

In our study, we explored the influence of sibling relationships on the academic performance of American children [2]. Data was collected from the Child Development Supplement (CDS) and the Panel Study of Income Dynamics (PSID) from January 1996 to December 1997. The study includes a total of 3,563 children, with ages ranging from 0 to 12 years old, who were selected from 2,380 households. The race of the children is not specified. Children from families with both parents demonstrated higher performance on age-adjusted achievement exams in comparison to children from single-parent households. After analyzing the results of age-controlled examinations, it was noted that individuals who were born first displayed higher performance compared to their siblings born later in the same family.

In 2003, a study was conducted by researchers from Southwestern Agriculture University to explore the potential impact of various factors on students' performance in an introductory soils course (PSS 2432). In addition to the student's success in the introductory soils course, various factors were taken into account. These factors encompassed their SAT score, percentile rank in their high school class, major in high school chemistry, gender, and the academic term in which the course was offered. With the capability to perform multi-regression and correlation analysis There is a strong correlation between grades in PSS2432, SAT percentile rank, high school class, gender, and major.

RESEARCH OBJECTIVE

The major objective of every student is to improve performance by making effective decisions and attain outstanding outcomes. Accurate data forecasts are essential for attaining optimal outcomes in the field of education. Various prediction and classification approaches exist, each characterised by its unique methodology. The research team gathers data on student performance, however, there is no assurance that any variables can reliably forecast whether a student would thrive, struggle, or fall within the average range.

PROPSED METHODOLOGY

We are working with and trying to solve the problem of the possible loss of educational due to student performance. How well kids do in school depends on many things. The value of a degree changes depending on many factors related to education that affect how well a

student does, such as the start and end times, the number of exams, the volume, and the adjusted close. The question is how a kid can do the best in school, which makes news headlines. If a better algorithm is made to predict how well a student will do in school in the short term, and some datasets (big data) are reviewed using an efficient algorithm and covariance is found, then software can help each student do better. Fundamental analysis, which tries to guess how profitable an education system will be in the future, looks at two main things: how the school is doing now and how well its students are doing. Trends in how well students are doing can be found using technical analysis plots and statistical data. To solve the problem stated above, we are working on an algorithm that can accurately predict school costs with a small error margin, which will protect student performance.

CONCLUSION

Below are the boundaries of the current inquiry for your information. This investigation will just encompass five educational districts within the Indian state of Telangana. Nine different types of schools were chosen for each educational district based on specific criteria. These criteria included the type of financial support (public or private), the number of students in each class, the language of instruction (Tamil or English), and whether the classes were mixed gender or not. Every university or college has as its fundamental mission the guaranteeing of a first-rate education for its students. The success of a school depends on the student's performance. Achieving admission to a programme or a high academic % upon completion of coursework is considered a success for a student. Research on student success and failure rates was given by a number of authors. Finding and predicting students' success rates has been the focus of a great deal of research. A number of writers have created models for making predictions, such as ID3, Bayesian networks, and methods for linear and multiple linear regression.

REFERENCES

1. Andrew Kusiak ; Xiupeng Wei ; Anoop Prakash Verma ; Evan Roz, (2013), "Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach", ISSN: 0196-2892, Volume: 51, Issue: 4, PP: 2337-2342.
2. Caiping Hu ; Xiaolin Qin ; Jun Zhang, 2007, "Spatial Prediction Models for Mining Spatial Data", 2007 IEEE International Conference on Integration Technology, PP: 369372.

3. Divya Tomar, Sonali Agarwal. Twin Support Vector Machine: A review from 2007 to 2014. Egyptian Informatics Journal 2015; 16: 55–69.

4. G. Zhou, Y. Zhu, G. Wang, Cache conscious star-join in mapreduce environments, in: 2nd International Workshop on Cloud Intelligence (colocated with VLDB 2013), Cloud-I '13, Riva del Garda, Trento, Italy, August 26, 2013, 2013, pp.1:1–1:7.

5. Diana Heredia ; Yegny Amaya ; Edwin Barrientos, 2015, "Student Dropout Predictive Model Using Data Mining Techniques", ISSN: 1548-0992, Volume: 13, Issue: 9, PP: 3127-3134.

6. Fong-Long Huang ; Neng-Huang Pan ; Ming-Shing Yu ; Jun-Yi Wu, 2011, "Break prediction of prosody for Hakka'S TTS systems based on data mining approaches", ISSN: 2160-133X, 2011 International Conference on Machine Learning and Cybernetics, PP: 51-55.

7. Tao Huang, Liang Lan, Xuexian Fang. Promises and Challenges of Big Data Computing in Health Sciences. Big Data Research 2015; 2: 2–11.

8. K. Menaka, S. Karpagavalli. Breast Cancer Classification using Support Vector Machine and Genetic Programming. International Journal of Innovative Research in Computer and Communication Engineering; Vol. 1, Issue 7, September 2013.

9. Valmik B Nikam & B.B. Meshram, 2013, "Modeling Rainfall Prediction Using Data Mining Method: A Bayesian Approach", ISSN: 2166-8523, 2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation, PP: 132-136.

10. Yang Lan Qin & Xu Xin, 2012, "Research on the price prediction in supply chain based on data mining technology", 2012 International Symposium on Instrumentation & Measurement, Sensor Network and Automation (IMSNA), PP: 460-463.

11. Zhao Yongyi & Pan Qiang, 2010, "Intelligent data mining for economie prediction and analysis", 2010 2nd IEEE International Conference on Information Management and Engineering, PP: 510-513.