**EPR MODEL FOR EDUCATION SYSTEM**

SUBMITTED IN COMPLETE FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF

**BACHELOR OF ENGINEERING**

IN

**INFORMATION TECHNOLOGY**

BY

**ABHISHEK GUPTA**

**YASH KOTHAWADE**

**KARTIK PARMAR**

UNDER THE GUIDANCE OF

**Prof. JYOTSNA MORE**

(Prof. Department of Information Technology)



**INFORMATION TECHNOLOGY DEPARTMENT**

**XAVIER INSTITUTE OF ENGINEERING**

**UNIVERSITY OF MUMBAI**

**2019 – 2020**

# XAVIER INSTITUTE OF ENGINEERING

**MAHIM CAUSEWAY, MAHIM, MUMBAI-400016.**

**CERTIFICATE**

### This to certify that

ABHISHEK GUPTA (815)

YASH KOTHAWADE (825)

KARTIK PARMAR (838)

Have satisfactorily carried out the PROJECT work titled **“EPR MODEL FOR EDUCATION SYSTEM”** in

complete fulfillment of the degree of Bachelor of Engineering as laid down by the University of Mumbai during the academic year 2019-2020.

**Prof. Jyotsna More**

**Supervisor/Guide**

**Prof. Chhaya Narvekar Dr. Y.D.Venkatesh**

**Head of Department Principal**

# PROJECT REPORT APPROVAL FOR B.E.

**THIS PROJECT REPORT ENTITLED “EPR MODEL FOR EDUCATION SYSTEM”**

**BY**

**ABHISHEK GUPTA (815)**

**YASH KOTHAWADE (825)**

**KARTIK PARMAR (838)**

**IS APPROVED FOR THE DEGREE OF BACHELOR OF ENGINEERING.**

**Examiners**

**1.**

**2.**

**Supervisors**

**1.**

**2.**

**Date:**

**Place: MAHIM, MUMBAI**

**DECLARATION**

I declare that this written submission represents my ideas in my own words and where others’ ideas or words have been included, I have adequately cited and referenced the original sources.

I also declare that I have adhered to all the principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which thus have not been properly cited or from whom proper permission have not been taken when needed.

Abhishek Gupta (815) -------------------------------

Yash Kothawade (825) --------------------------------

Kartik Parmar (838) --------------------------------

Date:

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **SR. NO.** | **TOPIC** | **PAGE**  **NO.** |
| I | LIST OF FIGURES | i |
| II | LIST OF TABLES | ii |
| III | ACKNOWLEDGEMENT | iii |
| 1. | INTRODUCTION   * 1. PROBLEM DEFINITION   2. SCOPE OF THE PROJECT | 1 |
| 2. | LITERATURE REVIEW | 4 |
| 3. | IMPLENTATION STRATEGY | 8 |
|  | REFERENCES | 13 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **SR. NO.** | **FIGURE** | **PAGE NO.** |
| 3.1 | Education data mining process in education systems | 8 |
| 3.2 | Experimental Models in Educational Data Mining | 9 |
| 3.3 | Architecture of System | 10 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **SR. NO.** | **TABLES** | **PAGE NO.** |
| 3.1 | Student Background Attributes | 11 |
| 3.2 | Student Social Activities Attributes | 12 |
| 3.3 | Student Period Results Attributes | 12 |

**Acknowledgement**

We would like to thank **Fr. (Dr.) John Rose S.J. (Director of XIE)** for providing us with such an environment so as to achieve goals of our project and supporting us constantly.

We express our sincere gratitude to our **Honorable Principal Dr. Y.D. Venkatesh** for encouragement and facilities provided to us.

We would like to place on record our deep sense of gratitude to **Prof. Chhaya Narvekar**, Head of Dept. Of Information Technology, Xavier Institute of Engineering, Mahim, Mumbai, for her generous guidance help and useful suggestions.

With deep sense of gratitude we acknowledge the guidance of our project guide **Prof. Jyotsna More**. The time-to-time assistance and encouragement by her has played an important role in the development of our project.

We would also like to thank our entire Information Technology staff who have willingly cooperated with us in resolving our queries and providing us all the required facilities on time.

Abhishek Gupta -----------------------------

Yash Kothawade -----------------------------

Kartik Parmar -----------------------------

**CHAPTER NO. 1**

**INTRODUCTION**

Huge amount of data is being generated by everything around us at all time and is produced by every digital process and social media exchange through system, sensors and mobile devices etc. Big data analytics examines large amount of data to uncover the hidden patterns, correlation and other insights. Using the concept of data mining, a number of algorithms are explored in order to predict the overall performance of the students. To extract meaningful value from big data, one need optimal processing power, analytics capabilities and skills. Based on a previous and current performance of the students the system will train it self and the new performance of a students is generated by applying different data mining algorithms on it.

There are lots of discussions and reviews for Student’s performance based on the previous researches. Schools, Colleges, and other Educational Institutes are running on high pace to provide scholars in this competitive world. These Educational institutes focus on generating graduates with good academic performances as well as extra-curricular activities. The student’s performance prediction is an important part in education system. There are several parameters which effect the student performance level like gender, father’s occupation and marks secured in prior years (Last semesters). The educational process is improved by the prior forecast of student performance. They need to keep track on how the student is performing in particular fields and in what fields they need more training. By using Educational data mining techniques, the educational authorities can have the idea before the starting of the new semester and can have informed decision so it will help them to effectively deal with all problems faced by the students while performing academically or in their personal life as will be already known to them. Large volumes of data are analyzed using educational data mining techniques so as to find different trends and patterns to predict the student performance. Day by day the volumes of data are increasing so to analyze we need to generate algorithms using data mining and then compare them so to get the maximum accuracy rate. The analysis process will be done by applying data mining methodology on institutional education data sets. Data classification is the method, we need to apply in data mining research. Different models are available in representing the Classifier. The selection of algorithm plays a major role in classifiers after giving enormous set of data.

* 1. **PROBLEM DEFINITION:**

Large numbers of graduates/post graduates are produced by universities/institutes every year. Universities/Institutes may follow best of the pedagogies; but still they face the problem of dropout students, low achievers and unemployed students. Understanding and analysing the factors for poor performance is a complex and incessant process hidden in past and present information congregated from academic performance and student’s behavior. Powerful tools are required to analyze and predict the performance of students scientifically. Although, universities/institutions collect an enormous amount of student’s data, but this data remains unutilized and does not help in any decision.

Predicting students’ performance becomes more challenging due to the large volume of data in educational databases. This is due to main two reason. First, the study of existing prediction methods is still insufficient to identify the most suitable methods for predicting the performance of students in the University. Second is due to the lack of investigations on the factors affecting student’s achievements in particular courses.

The current educational system does not involve any prediction about pass or fail percentage based on performance. The system does not deal with dropouts, low achievers and unemployed students. There is no efficient method to caution the students about the deficiency in attendance. It does not identify the weak student and inform the teachers. The faculty cannot find out students’ abilities and their interest easily so that they can enhance them in it. Thus, it may affect with poor university results, placement and career of individual.

Universities and other Educational Institutes are running on high pace to provide scholars in this competitive world. Educational institutes focus on generating graduates with good academic performances as well as extra-curricular activities. They need to keep track on how the student is performing in particular fields and in what fields they need more training.

So, the system should be design to predict the overall performance of any student that too considering every attributes that contribute towards the students performance i.e. not only the attendance factor or considering marks but also some of the other factors such as courses, social activities, etc. a student is involved in.

* 1. **SCOPE:**

Other than academic attributes, there are other components also which are responsible for student’s overall performance like personal and emotional stability.

The student’s data that was collected in this research included a classic sampling process which was a time consuming task, it could be better if the data was collected as part of the admission process of the university, that way, it would be easier to collect the data, as well as, the dataset would have been much bigger, and the university could run these data mining tasks regularly on their students to find out interesting patterns and may be improve their performance.

This system can also be implemented in different non educational institutes like business corporates, sports academies and manufacturing companies where the challenge would be taking into consideration the current market scenario as one of the most important factors affecting quality of their products and employee.

**CHAPTER NO. 2**

**LITERATURE REVIEW**

Machine learning is a new approach to learn and analyze a complex and huge data. It is based on algorithms that can learn from data without relying on the conventional programming, i.e., rules-based programming. It emerged individually as a scientific discipline in the late 1990s as steady advances in digitization and cheap computing power enabled data. Scientists have stopped building finished models and have plunged into a novel adventure in training computers, through which an unmanageable volume of data and complexity of the big data can be processed and analyzed using the potentiality of machine learning. Hence machine learning is the emerging trend in the era of information technology

Machine Learning results in higher degree of accuracy, in scenarios where human analytics could not visualize data on their own and make predictions. In particular, learning and evaluating such models have a variety of challenges like machine learning skills in domain area, collection of data and algorithmic complexity, etc. In this paper we have made an attempt to predict results of a batch of students based on the previous performance.

This is the paper that we referred from which we got our idea of working on this topic. It provided us with an actual implementation idea that we have to implement.

It helped us to identify and analyse the impact of student background attributes and student social activities attributes on student performance. Supervised educational data mining namely, Naïve Bayesian, Multilayer Perceptron, decision tree J48 and random forest are applied to build prediction model. The significant and impact of student background and social activities attributes can be visualized and defined from the decision tree structure generated by the models [1].

Extracting the relevant information from a bulk of dataset is the process of Data Mining. Mining knowledge from prearranged and unstructured data is said to be the data mining. If we discover information or knowledge from unstructured data, then it is Knowledge Engineering or Data mining. Data Mining is the essential step in the process of knowledge discovery (KD). The information or knowledge extraction and analysis usually will be having following steps:

● Cleaning the Data set: It is a process of removal of the noise and contradictory information.

● Data Integration: In this process an integration of multiple datasets are happening.

● Data Selection: The information important to our required assignment of job are to be recovered at this stage.

● Data Transformation: It is a procedure to perform sum of the assignment of job. Information will be changed into the structures are suitable for Knowledge Discovery Process.

● Data Mining: Various intelligent patterns of data are extracted in order to make it as useful information to be applied in this stage.

So, the above-mentioned points are the one which will be helpful that we referred from this paper [2].

How to extract trends and patters from the data by applying various algorithms. By applying the Apriori algorithm to the academic records of a group of students and obtaining the best association rules which will help in student profiling. Using Kmeans clustering to group the students categorically and efficiently. In the first section, discussing the Apriori and Kmeans algorithms. Then, discussing how to implement these algorithms on the academic data. Finally, studying the results and deriving useful information from them. So, this paper helped to understand what is Apriori algorithm and a K-means Algorithm is and what is the use of this algorithms. It also described how to implement those algorithms [3].

This paper referred to attributes which are useful and which are useless for predicting a result. Creating a model of individual attributes can help us understand which one is contributing how much for our prediction [4].

What is Naive Bayes Algorithm is and how it works? These are the things about Naive Bayes: In machine learning, Naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. It is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. One of the main assumptions of the Naïve Bayes algorithm is that each feature is independent, which holds good for the problem considered, since the score of the student in each subject is independent, though it could be related with similar subjects. Due to this assumption, this classifier is very effective for this problem [5].

What exactly a ML is and what types of algorithm can be used in a particular scenario. The difference between Supervised and Unsupervised Learning algorithms. How and where to use ML in implementing a real-life problem by using a particular algorithm which is suitable for the problem. How to choose a right ML model and also in knowing the future of ML [6].

How to approach our project and to understand the exact problem in real life so that providing some of the solutions by implementing the project [7].

To get the flow of the system. The first step is collecting the data from the data sources. In our case, the data has been collected from our college itself. The second step is preprocessing the data in order to get a normalized dataset and then labeling the data rows. In the third step, the result of the second step, the training and testing dataset, is fed to the Machine Learning algorithm. The Machine Learning Algorithm builds a model using the training data and tests the model using the test data. Finally, the Machine Learning Algorithm produces a trained model or a trained classifier that can take as an input a new data row and predicts its label [8].

This was used to understand what exactly the classification is and what are the types of classification algorithms and how it can be used for different classification algorithms for a particular requirement [9].

To design the UI part of the system and how the final prediction varies by changing the attributes values which ensures that the system is working properly [10].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SR. NO | YEAR | NAME OF PAPER | AUTHOR | CONTETNT |
| 1 | 2018 | Data Mining Analysis on Student’s Academic Performance through Exploration of Student’s Background and Social Activities. | Ching-Chieh. Kiu | This is the paper that we referred from which we got our idea of working on this topic. It provided us with an actual implementation idea that we have to implement. |
| 2 | 2019 | Students Performance Prediction Using Data Mining Techniques. | T. R. Kumar, T. Vamsidhar, B. Harika, T. M. Kumar and R. Nissy | This paper focused on the knowledge extraction and analysis. |
| 3 | 2012 | Application of Data Mining in Educational Databases for Predicting Academic Trends and Patterns. | P. Suhem, Z. Zain, and M. Fatima | This paper helped us to understand what is Apriori algorithm and a K-means Algorithm is and what is the use of this algorithms. |
| 4 | 2015 | Student Performance Prediction using Machine Learning. | Havan Agrawal, Harshil Mavani. | This paper helped us to understand like which attributes are useful and which are useless for predicting a result |

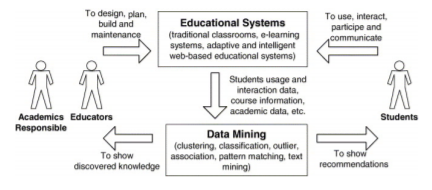
|  |  |  |
| --- | --- | --- |
| SR NO. | SITE | CONTENT |
| 1 | <https://en.wikipedia.org/wiki/Naive_Bayes_classifier> | This link refers what is Naive Bayes Algorithm is and how it works. |
| 2 | http://whatis.techtarget.com/definition/machine-learning | How to choose a right ML model and it also helped us in knowing the future of ML. |
| 3 | https://dl.acm.org/doi/10.1145/3318396.3318419 | Understand the exact problem in real life so that we can provide some of the solutions by implementing the project. |
| 4 | https://www.researchgate.net/publication/332893829\_  Predicting\_Students'\_Performance\_Using\_  Machine\_Learning\_Techniques | This link helped us to get the flow of the system. |
| 5 | https://www.geeksforgeeks.org/basic-concept-classification-data-mining/ | Understand what exactly the classification is and what are the types of classification algorithms. |

|  |  |  |
| --- | --- | --- |
| SR .NO | YOUTUBE LINK | VIDEO NAME |
| 1 | https://youtu.be/Vvv1ODX4pYY | How to design a student performance prediction using machine learning |

**CHAPTER NO. 3**

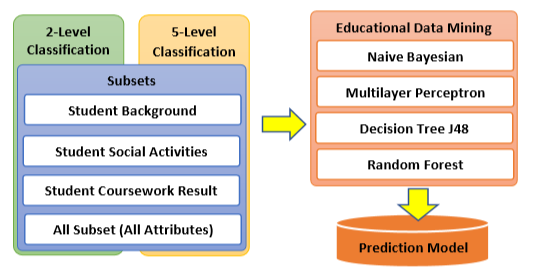
**IMPLEMENTATION STRATEGY**

The Fig 3.1 shows that the Educational Data Mining (EDM) has been actively applied to improve student performance in in education systems. Early prediction and analysis of at-risk student identification in classroom education may be helpful for both students and teachers. Teachers can have sufficient time to perform education interventions to improve student’s performance. The cycle of applying data mining in educational systems is depicted in Fig 3.1.



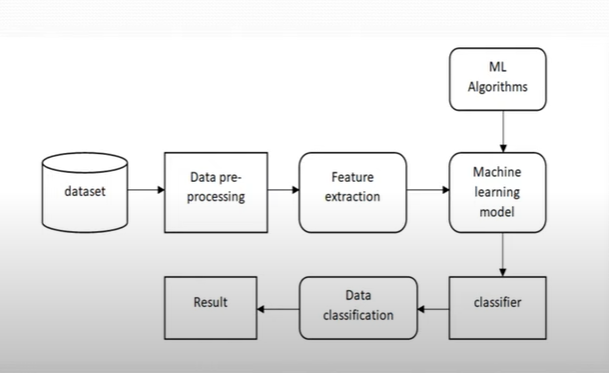
**Fig 3.1: Education data mining process in education systems**

Fig 3.2 shows the four supervised educational data mining techniques, namely Naïve Bayesian, Multilayer Perceptron, Decision Tree J48 and Random Forest. The evaluation has been performed on the three subset attributes on 2-level classification and 5-level classification models. The experimental analysis also performed on all attributes which is referred as all subsets dataset.



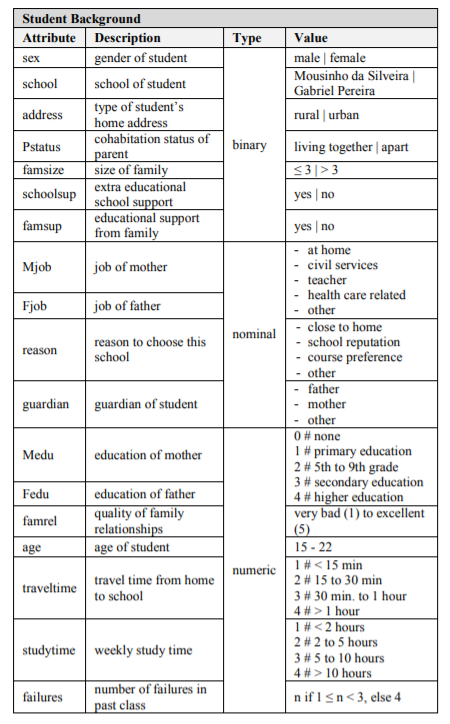
**Fig 3.2: Experimental Models in Educational Data Mining**

Fig 3.3 shows the architecture of system it includes dataset or data warehouse is utilized to input the raw data there might be a lot of inconsistency in it so data preprocessing is important after that Machine learning algorithms is used to provide Machine Learning model then classification algorithm is used to predict the result.

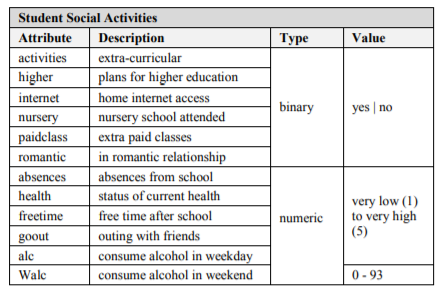


**Fig 3.3: Architecture of System**

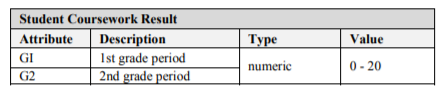
**Table No 3.1: Student Background Attributes**



**Table No 3.2: Student Social Activities Attributes**



**Table No 3.3: Student Period Results Attributes**



**REFERENCES**

1. C. Kiu, "Data Mining Analysis on Student’s Academic Performance through Exploration of Student’s Background and Social Activities," 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), Subang Jaya, Malaysia, 2018, pp. 1-5, doi: 10.1109/ICACCAF.2018.8776809.
2. T. R. Kumar, T. Vamsidhar, B. Harika, T. M. Kumar and R. Nissy, "Students Performance Prediction Using Data Mining Techniques," 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, Tamilnadu, India, 2019, pp. 407-411, doi: 10.1109/ISS1.2019.8907945.
3. P. Suhem, Z. Zain, and M. Fatima, “Application of Data Mining in Educational Databases for Predicting Academic Trends and Patterns”, 2012 IEEE International Conference on Technology Enhanced Education (ICTEE), 2012, pp 1 – 4.
4. Havan Agrawal, Harshil Mavani, 2015, Student Performance Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 04, Issue 03 (March 2015), http://dx.doi.org/10.17577/IJERTV4IS030127.
5. <https://en.wikipedia.org/wiki/Naive_Bayes_classifier>. (Accessed on 27-3-20, timestamp 14.30)
6. http://whatis.techtarget.com/definition/machine-learning.(Accessed on 2-4-20, timestamp 12.30)
7. https://dl.acm.org/doi/10.1145/3318396.3318419. (Accessed on 2-4-20, timestamp 13.00)
8. https://www.researchgate.net/publication/332893829\_Predicting\_Students'\_Performance\_Using\_Machine\_Learning\_Techniques. (Accessed on 10-4-20, timestamp 15.30)
9. <https://www.geeksforgeeks.org/basic-concept-classification-data-mining/>. (Accessed on 10-4-

20, timestamp 16.30)

1. (Designing UI) https://youtu.be/Vvv1ODX4pYY. (Accessed on 15-5-20, timestamp 17.30)