A Survey on Smart Education System using Machine Learning Techniques

[1]Kavana M D, [2]T S Deepthi, [3]Salman Shariff, [4]Rachana Prabhu, [5]Tufail Rafiq

[1]Assistant Professor in Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, 570002, Karnataka, India [2]Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, 570002, Karnataka, India [3]Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, 570002, Karnataka, India [4]Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, 570002, Karnataka, India [5]Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, 570002, Karnataka, India

Abstract— Most of the students across the world are mostly in confusion after their higher secondary and during which they have to make the choice for an appropriate career. At a very young age, the students don't have adequate maturity to accurately know about what an individual has to follow in order to choose a congenial career path. As we pass through the stages, we realize that every student undergoes a series of doubts or thought processes on what to pursue, which is the single tallest question. Then comes the next agony whether they have essential skills for the stream they've chosen. Our computerized student counselling system is used to predict the suitable domain for an individual based on their skills and factors related to education. System helps in choosing an appropriate course which will also reduce the failure rate by choosing a wrong career path.One of the biggest challenges for any of the higher learning institutions today is to improve the placement performance of students. The placement prediction is more complex when the complexity of educational entities increase. The proposed system also predicts the placements status of the student and gives a aggregated information about the number of students to help and focus on the ratio that might need an extra push to get placed, This can be done by studying, analyzing and the educational parameters. This system uses ML algorithms for the recommendations.

Date of Submission: 14-05-2023	Date of acceptance: 26-05-2023

I. INTRODUCTION

When it comes to choosing a career it's not only on what course you choose, it's more than what you want to become after your graduation.[6] Career counselling is more about knowing and understanding about yourself and your capabilities and abilities. It is this time every student gets a lot of guidance from various circles (parents, teachers, other educational specialists, etc.) and accordingly the student decides about which course they want to join. Many times, we have come across a situation where a student opts for a course/stream and later repent for having chosen the one. To quote an example, there is a myth that one who does very well and scores highest marks in 12th grade chemistry will tend to choose chemical engineering because they are good in chemistry, however in reality that is not the case. We had multiple rounds of deliberation with students who are currently doing their engineering and students who are currently in 11th and 12th grade. Then we came up with an idea of providing an objective assessment of one's skill set and caliber that recommend a right stream to choose and hence we picked this as our problem statement and started thinking through how we can help the students in addressing this question.

RELATED WORKS
1. Student Placement Analyzer: A Recommendation System Using Machine Learning
YEAR OF PUBLICATION: 2017
METHODOLOGY: Decision Tree Classifier, Sci-kit Learn, Machine Learning, Prediction, Python

LIMITATIONS:
• Less number of outcomes considered such as Dream Company, Core Company, Mass Recruiters, Not Eligible and Not Interested

---

- Only 2 parameters used.

- Applicable for B.Tech Course

2. A Review on Student Placement Chance Prediction
YEAR OF PUBLICATION: 2019
METHODOLOGY: logistic regression, random forest and Decision Tree

LIMITATIONS:
- Literature Survey paper

- Implementation not done

3. Utilizing Exploratory Data Analysis for the Prediction of Campus Placement for Educational Institutions
YEAR OF PUBLICATION: 2020
METHODOLOGY: EDA techniques applied
LIMITATIONS:
- This concept is applied for only MBA specialization

- Here only YES OR NO is predicted

- Huge amount of data required
- Less accurate results

4 . A Prediction Model to Improve Student Placement at a South African Higher Education Institution

YEAR OF PUBLICATION: 2020
METHODOLOGY: Linear Regression, RF

LIMITATIONS:
- This is an advising system related to academic subjects

- Not Suitable for company campus selection
- Algorithms used generate graphical results
- Not suitable for real time application

5. A Job Recommendation Method Optimized by Position Descriptions and Resume Information

YEAR OF PUBLICATION: 2016

METHODOLOGY: item-based collaborative filtering algorithm for job recommendations.
LIMITATIONS:
- This concept applied for job recommendation for the students.

- This can't be applied for campus selection process.
- Algorithms used take more time for data processing.

Machine Learning
ML concerns with construction and study of system that can learn from data. For example, ML can be used in E-mail message to learn how to distinguish between spam and inbox messages.

Functionality:
    A computer program is said to be learnt from experience E with respect to some task T and some performance P only if the program performance increases with experience E.[1]
    ML is a branch of AI which contains statistical, probabilistic, optimization technique that can learn from past experience and discover the pattern from large complex data sets.

1

For example, we can apply ML techniques in predicting student performance based on their skills. Student performance depends on many factors such as living locality, SSLC result, PUC result, Family income, Parents education, use of internet, use of mobile, use of bike, use of Social Networking and other habits.[3]

We can predict student performance using ML technique before exams so that we can improve student performance by knowing status and skills of student.

ML based technique can be applied to classify the employees in an organization either to be in the class leave or stay based on their behavior.
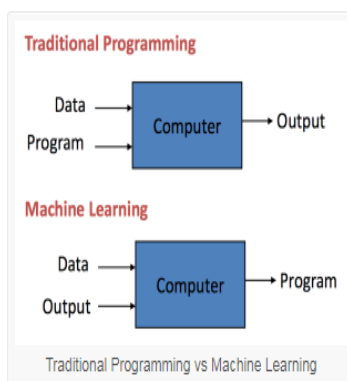
ML PROBLEMS:

1. Is this disease cancer?
2. Identifying types of cancer.
3. What's the market value of a House?
4. In Social Networking we can use ML technique to find best friends based on their interaction.
5. Suggesting the interesting tourist places using ML.
6. Predicting Customer purchasing patterns in Online shopping.

TYPES OF ML:

There are three types of Machine learning (ML), they are:

i. **Supervised Machine Learning** Here we have labels and the input is past examples.
  Ex: 1-4

ii. **Unsupervised Machine Learning**Extraction of patterns without labels.
  Ex. 5 and 6

iii. **Semi-Supervised Machine Learning**Mixture of both Supervised and Unsupervised Machine Learning

Machine learning is like farming or gardening. Seeds is the algorithms, nutrients is the data, the gardner is you and plants is the programs.

Traditional Programming

Data →
Program → Computer → Output

Machine Learning

Data →
Output → Computer → Program

Traditional Programming vs Machine Learning

SUPERVISED ALGORITHMS

KNN Algorithm

● K-Nearest Neighbour is one of the most commonly used Machine Learning algorithms based on Supervised Learning technique.

● K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

● K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

● K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

● K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.

● It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the data-set and at the time of classification, it performs an action on the data-set.

● KNN algorithm at the training phase just stores the data-set and when it gets new data, then it classifies that data into a category that is much similar to the new data.

Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog.[2] So for this identification, we can use the KNN algorithm, as it works on a similarity measure.

Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.

KNN Algorithm Steps

The K-NN working can be explained on the basis of the below algorithm:[7]

Step-1: Select the number K of the neighbors

Step-2: Calculate the Euclidean distance of K number of neighbors

Step-3: Take the K nearest neighbors as per the calculated Euclidean distance

Step-4: Among these k neighbors, count the number of the data points in each category

Step-5: Assign the new data points to that category for which the number of the neighbor is maximum

Step-6: Our model is ready

Naïve Bayes Algorithm Steps

Step 1: Scan the dataset (storage servers)retrieval of required data for mining from the servers such as database, cloud, excel sheet etc

Step 2: Calculate the probability of each attribute value. [n, n_c, m, p]. Here for each attribute we calculate the probability of occurrence using the following formula. (mentioned in the next step). For each class(disease) we should apply the formulae

Step 3: Apply the formulae

$$P(attributevalue(a_i)/subjectvaluev_j)=(n\_c+mp)/(n+m)$$

Where:

- n = the number of training examples for which $v = v_j$
- n_c = number of examples for which $v = v_j$ and $a = a_i$
- p = a priori estimate for $P(a_i|v_j)$
- m = the equivalent sample size

Step 4: Multiply the probabilities provided by P for each class, here we multiple the results of each attribute with p and final results are used for classification.

Step 5: Compare the values and classify the attribute values to one of the predefined set of class.

FLOW DIAGRAM
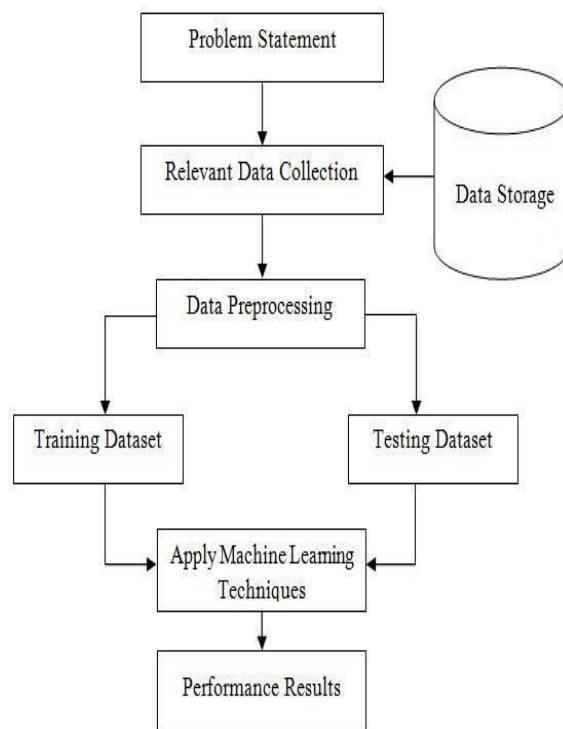
II.   CONCLUSION

Many research works are done on the educational data. All works are building models using PYTHON Language and using some ready tools like R tool, Rapid Miner etc.[5] This Smart placement system helps in recommending the suitable course for the students and predicts the placement status of the student. Machine

learning is an emerging technology used to process data and extract the useful information. To recommend suitable course for the final year students and for placement status prediction, ML algorithms used to process education data and predicts the outputs. This system helps education sector to take right decision in less time so as to improvise the overall performance of the institute, guide each and every student to choose the right path and also help the negative skilled students to lift up their placement chances.

REFERENCES

[1]. A.M. El-Halee's, and M.M. Abu Tair, "Mining educational data to improve students' performance: A case study," International Journal of Information and Communication Technology Research, 2011, pp. 140-146.M. tech Er. Rimmy Chuchra. " Use of data mining techniques for the evaluation of student performance: a case study ". International Journal of Computer Science and Management Research, 1(3):425-433, October 2012.

[2]. Muhfuza Haque Md. Hedayetul Islam Shovon. " Prediction of student academic performance by an application of k-means clustering algorithm ". International Journal of Advanced Research in Computer Science and Software Engineering, 2(7):353-355, July2012.

[3]. S.Karthik M.Sukanya, S.Biruntha and T.Kalaikumaran, "Mining Data mining: Performance improvement in education sector using classification and clustering algorithm," ICCCE In Proceedings of the International Conference on Computing and Control Engineering, 2012.

[4]. Brijesh Kumar Bhardwaj and Saurabh Pal. Data mining:"A prediction for performance improvement using classification & (IJCSIS) International Journal of Computer Science ^d_ Information Security, 9(4), April,2011.

[5]. H. Anandakumar and K. Umamaheswari, "An Efficient Optimized Handover in Cognitive Radio Networks using Cooperative Spectrum Sensing," Intelligent Automation & Soft Computing, pp. 1-8, Sep. 2017. doi:10.1080/10798587.2017.1364931

[6]. Haldorai, A. Ramu, and S. Murugan, " Social Aware Cognitive Radio Networks," Social Network Analytics for Contemporary Business Organizations, pp. 188-202. doi:10.4018/978-1-5225- 5097-6.ch010

[7]. H. Anandakumar and K. Nisha, "Enhanced multicast cluster-based routing protocol for delay tolerant mobile networks," International Journal of Information and Communication Technology, vol. 7, no.6, p. 676, 2015.

[8]. Md. Hedayetul Islam Shovon and Mahfuza Haque." An approach of improving student's academic performance by using k-means clustering algorithm and decision tree". (IJACSA)International Journal of Advanced Computer Science and Applications, 3(8):146-149, August,2012.