

Predictive Analysis of Student Stress Level using Machine Learning

**Dr. J V Gorabal, Mr. Vipul Y S, Mr. Yashas S Bharadwaj,
Mr. Mohammed Anas, Mr. Mohammed Nehad Ameen**

Department of Computer Science and Engineering
ATME College of Engineering, Mysuru, India

Abstract: 1. College students are suffering from many mental health problems including mental stress, somatization, obsessive, interpersonal sensitivity, depression, anxiety, hostility, fear, paranoia and psychosis, which can bring a lot of negative effects to them.

2. Obviously, the mental health problems of college students not only directly affect their own growth, but also affect the stability of the campus. Colleges increasingly prioritize monitoring and preventing students' psychological crises.

3. All Colleges simply analyze whether students have mental health problems or what kind of problems they have. It fails to uncover underlying relationships within psychological data.

4. We require a comprehensive system to address student mental health concerns, with a specific focus on predicting and managing student stress.

5. There are so many factors related to stress such as work load, assignments, family issues, friends issues, attendance, teaching etc...

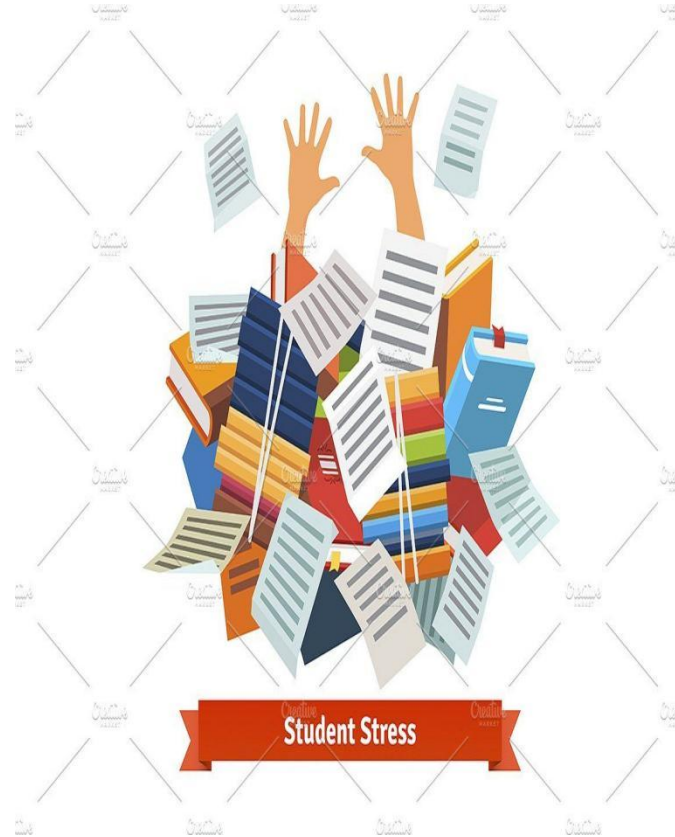
6. Machine learning is a subject to predict future based on the past data. By leveraging machine learning techniques, we can accurately predict student stress levels. Furthermore, our proposed system will provide personalized suggestions based on these stress levels

Keywords: Machine Learning

I. INTRODUCTION

1. Students are facing so many mental health problems such as depression, pressure, stress, interpersonal sensitivity, fear, nervousness etc.. Though many industries and corporate provide mental health related schemes and Efforts to alleviate workplace stress are ongoing, but the issue remains challenging to manage effectively. Predicting stress levels in college students is a significant and demanding task within the education sector today.
2. Stress is regarded as a major thing that is used to create an imbalance in the life of every character and it is additionally regarded as a major issue for psychological adjustments and trauma reduction.
3. Numerous studies work on stress management in school students. Students pursuing secondary and tertiary education commonly experience high levels of ongoing stress. It can be many times decided as day to day movements for a hassle-free mind to pay attention to lecturers
4. To decrease the individual stress rate, human societies have been in a position to boost a complete stage of progress in monitoring the stress stage of students and make them score well in academics.
5. Lack of stress administration can result in some drastic injury which can sometimes affect the education completely and can even cause extreme injury to the fitness of the students at a variety of stages.
6. Individual family background has been conceptualized as a major play that has been taking a path from our childhood. Children residing in rural and urban areas are consistently exposed to distinct environments. Low-income students often face financial and familial challenges, which can contribute to lower academic performance. The fulfillment of the faculty and students is majorly structured on both faculties instructing and learning at home.

7. Current system is manual process where it is difficult to identify the stress in the college students. Currently, there is a lack of automation for predicting students' stress levels.



II. RELATED WORKS

Survey Papers

1. An Intelligent Student Advising System Using Collaborative Filtering

Description: We propose a web based intelligent student advising system using collaborative filtering, a technique commonly used in recommendation systems assuming that users with similar characteristics and behaviors will have similar preferences. With our advising system, students are sorted into groups and given advice based on their similarities to the groups. If a student is determined to be similar to a group students, a course preferred by that group might be recommended to the student.

Disadvantages

- System used to predict suitable course for students and data-set not compatible to predict student results.
- Not all student behaviors connected to course advising.
- Students are organized into groups, and the system then predicts the most suitable courses for each group of students.

2. Mining Students' Data for Performance Prediction

Description : The ability to predict a student's performance is very important in educational environments. Students' academic performance is based upon diverse factors like personal, social, Psychological and other environmental variables. A very promising tool to attain this objective is the use of Data Mining. Data mining techniques are used to discover hidden information patterns and relationships of large amount of data, which is very much helpful in decision making. A single data contains a lot of information. The type of information is produced by the data and it decides the processing method of data. A lot of data that can produce valuable information, in education sector contains this valuable information. Which helps the education sector to capture and compile low cost information for this

information and communication technology is used. Now-a-days educational database is increased rapidly because of the large amount of data stored in it. The loyal students motivate the higher education systems, to know them well; the best way is by using valid management and processing of the students' database. Data mining approach provides valid information from existing student to manage relationships with upcoming students.

Disadvantages

- The system utilizes student behaviors to predict their academic performance. Datasets not compatible for class results prediction.
- The student performance prediction model incorporated less relevant parameters, such as father's income, mother's income, and qualifications, which may not significantly contribute to accurate predictions.

3. An Effective Algorithm for Mining Positive and Negative Association Rules

Author : Honglei Zhu ,Zhigang Xu

Year: 2008

Description : Recently, mining negative association rules has received some attention and been proved to be useful in real world. This paper presents an efficient algorithm (PNAR) for mining both positive and negative association rules in databases. The algorithm extends traditional association rules to include negative association rules. When mining negative association rules, we adopt another minimum support threshold to mine frequent negative item-sets. With a correlation coefficient measure and pruning strategies, the algorithm can find all valid association rules quickly and overcome some limitations of the previous mining methods. The experimental results demonstrate its effectiveness and efficiency.

Disadvantages

- Algorithms used here takes more time processing
- Less efficient results

4. The Application of Association Rules Mining in the Analysis of Students' Test Scores

Author : Jigang Zheng and Jingmei Zhang

Year: 2016

Description : Data mining from large amounts of data to extract efficient, implicit, potential useful knowledge and information, it has in many other areas has been successfully applied. However, the application in the field of education is relatively less. The central work of institutions of higher learning is teaching, the focus is to improve the quality of education, but students are precisely the basis for measuring the quality of teaching. The characteristic of this paper is to apply association rule mining method of university students' grades, previously unknown effects on student achievement factors of mining, to provide some valuable reference for teachers and administrators, to provide the necessary decision support for teaching and student management, to better carry out teaching, in order to improve the quality of teaching in Colleges and universities.

Disadvantages

- Less parameter used in the project
- Less efficient results

5. Data Mining Applications In Healthcare Sector

Author : M. Durairaj, V. Ranjani

Year: 2013

Description : In this paper, we have focused to compare a variety of techniques, approaches and different tools and its impact on the healthcare sector. The goal of data mining application is to turn that data are facts, numbers, or text which can be processed by a computer into knowledge or information. The main purpose of data mining application in healthcare systems is to develop an automated tool for identifying and disseminating relevant healthcare information. This paper aims to make a detailed study report of different types of data mining applications in the healthcare sector

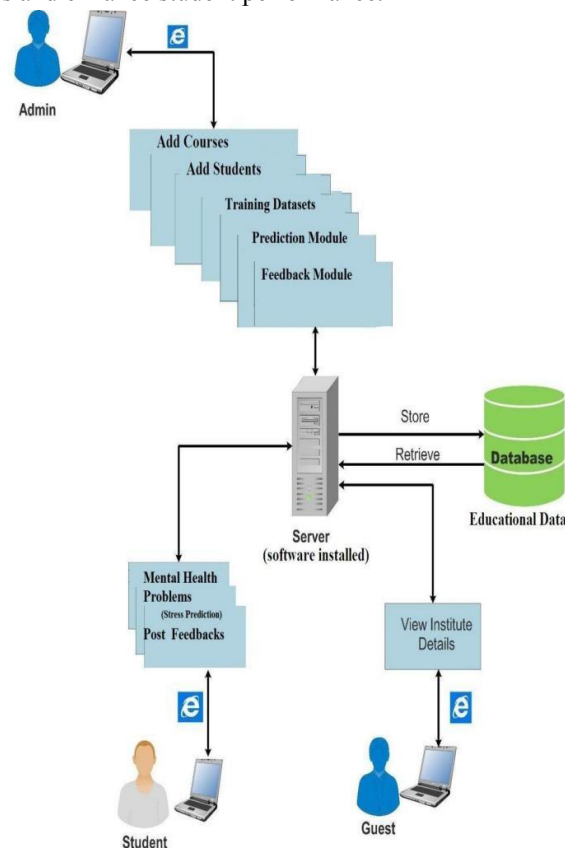
and to reduce the complexity of the study of the healthcare data transactions. Also presents a comparative study of different data mining applications, techniques and different methodologies applied for extracting knowledge from database generated in the healthcare industry. Finally, the existing data mining techniques with data mining algorithms and its application tools which are more valuable for healthcare services are discussed in detail.

Disadvantages

- Requires huge amount data
- Less accurate results

III. PROPOSED WORK

1. Proposed system meant for education sector.
2. The system identifies key factors that strongly influence stress levels.
3. Stress were identified based on gender, family history and availability of health benefits in the education sector.
4. By identifying the stress of students, we can come up with some approaches to reduce stress and create a much comfortable learning place for students.
5. System uses many parameters such as gender, age, family history, e_provided health benefits, share about illness, tech company, tech role, acquiring leave etc..
6. System uses machine learning algorithms or AI algorithms to find stress of an student.
7. The system can be developed as a real-time application, specifically designed for college use. *Visual Studio* and *SQL Server*, known for their strong support for real-time applications, will be utilized for the development of the application.
8. Finding the risk factors which affects the student’s mental health is the major objective of the system.
9. The system provides personalized suggestions to students based on their stress levels.
10. The system's objective is to identify factors influencing student stress levels, predict those levels, and propose solutions to reduce stress and enhance student performance.



IV. METHODOLOGY

ML concerns with construction and study of system that can learn from data. For example, Machine learning can be employed in email messaging systems to learn and differentiate between spam and inbox messages.

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

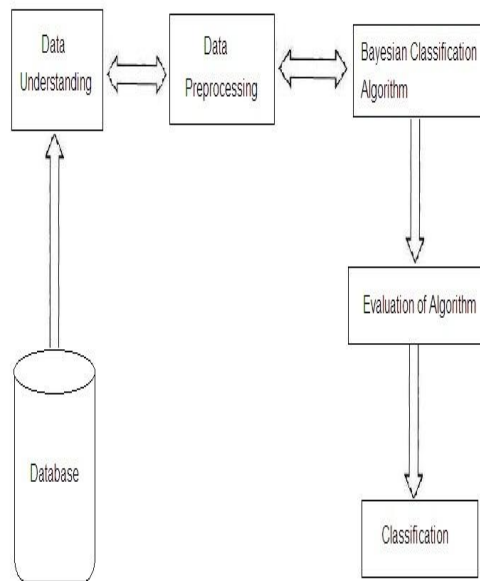
iii. Semi-Supervised Machine Learning

Mixture of both Supervised and Unsupervised Machine Learning

In this project, supervised learning techniques are utilized to process the training dataset. The **Naive Bayes algorithm** is employed to predict student stress levels.

Reasons for selecting Naive Bayes;

1. A majority of the prior medical research papers have employed these algorithms.
2. Surveys indicate that this algorithm is highly efficient for processing medical data.
3. takes less time for data processing.
4. Works fine for n number of parameters. Number of parameters need not to be fixed.



System architecture

V. DATASETS

Parameters List

- | | | |
|------------------|---|--|
| Gender | - | 1-Male,2-Female |
| Financial_Issues | - | 0-No,1-Yes |
| Family_Issues | - | 0-No,1-Yes |
| Study_Hours | - | numeric (in hrs) |
| Teaching_Method | - | 1-Fair,2-Not Good |
| Health_Issues | - | 1-Spectacles wearer, 2-Migraine Headache |

Partiality_Fix	-	0-marks, 1- knowledge
Exam_Schedule	-	1-Monthly, 2-Half ,3-yearly , 4-Annual, 5-Slip Test
Friends_Issue	-	0-No,1-Yes
Pressure	-	0-No,1-Yes
Regular	-	1-Regular,2-Irregular
Interaction	-	1-Poor,2-Good, 3-Better, 4- Best
Result	-	0-Stress Free, 1- 25% Under Stress, 2-50% under Stress,3-100% Stress

VI. EXPERIMENT RESULTS

Pseudo Code - Stress Prediction

```
//function which contains the algorithm steps
private string _NaiveBayesAlgorithm(string[] values)
{
    ArrayList _Distance = new ArrayList();
    ArrayList _RecordId = new ArrayList();
    ArrayList s = new ArrayList();
    output.Clear();
    //try
    //{
    s = GetSubject();
    int m = 35; //k value
    //finding the distance between the objects
    for (int i = 0; i < dt.Rows.Count; i++)
    {
        double _val = 0.0;
        for (int j = 0; j < values.Length; j++)
        {
            string _valluee = dt.Rows[i][j].ToString();
            if (_valluee.Equals("?") || values[j].ToString().Equals("?") ||
                _valluee.Equals("") || values[j].ToString().Equals(""))
            {
            }
            else
            {
                _val += Math.Pow(double.Parse(dt.Rows[i][j].ToString()) - double.Parse(values[j].ToString()), 2);
            }
        }
        _val = Math.Sqrt(_val);
        _Distance.Add(Math.Round(_val, 1));
        _RecordId.Add(i);
    }
    ArrayList temp = new ArrayList();
    ArrayList arrayRecords = new ArrayList();
    ArrayList arrayExists = new ArrayList();
    int d = 0;
    for (int x = 0; x < _Distance.Count; x++)
    {
        temp.Add(_Distance[x]);
    }
}
```

```

temp.Sort();
for (int y = 0; y < m; y++)
{
d = 0;
for (int z = 0; z < _Distance.Count; z++)
{
if (_Distance[z].Equals(temp[y]))
{
if (d == 0 && !arrayExists.Contains(_RecordId[z]))
{
arrayRecords.Add(_RecordId[z]);
arrayExists.Add(_RecordId[z]);
++d;
}
}
}
}
string _output = null;
if (arrayRecords.Count > 0)
{
int cnt;
ArrayList arrayCnt = new ArrayList();
ArrayList arrayOutcome = new ArrayList();
for (int i = 0; i < s.Count; i++)
{
cnt = 0;
for (int j = 0; j < arrayRecords.Count; j++)
{
if (dt.Rows[int.Parse(arrayRecords[j].ToString())]["Result"].ToString().Equals(s[i]))
{
++cnt;
}
}
arrayCnt.Add(cnt);
arrayOutcome.Add(s[i]);
}
ArrayList temp1 = new ArrayList();
for (int x = 0; x < arrayCnt.Count; x++)
{
temp1.Add(arrayCnt[x]);
}
temp1.Sort();
temp1.Reverse();
for (int y = 0; y < arrayCnt.Count; y++)
{
if (arrayCnt[y].Equals(temp1[0]))
{
_output = s[y].ToString();
//if (_output.Equals("0"))

```



```

//{
//_output = "No";
//}
//else
//{
//_output = "Yes";
//}
return _output;
}
}
}
return _output;
}

```

VII. RESULTS

Constraint	Algorithm
Accuracy	0.935482054929
Time (min:sec)	0:00
Correctly Classified	0.935482054929
Incorrectly Classified	0.064517945070

VIII. CONCLUSION

College students face numerous mental health challenges, encompassing mental stress, somatization, obsession, interpersonal sensitivity, depression, anxiety, hostility, fear, paranoia, and psychosis. These conditions can have significant detrimental effects on their well-being. Machine learning, as a discipline, enables us to predict future outcomes based on past data. By applying machine learning techniques, we can effectively forecast student stress levels. Moreover, our proposed system will provide personalized suggestions tailored to each student's stress levels.

IX. FUTURE ENHANCEMENTS

To evaluate the effectiveness of the model, additional methods such as the Naive Bayes classifier can be employed. This approach helps to assess the efficiency and performance of the model in predicting student stress levels. One can implement deep learning techniques like CNN (Convolved Neural Networks) and verify how the model performs for the given datasets.

A much more specific and vast datasets can be used as a training model since the number of responses is limited in our case.

REFERENCES

[1] Jena S, Tiwari H, "Stress and mental health problems in 1st year medical students: A survey of two medical colleges in Kanpur, India," International Journal of Research in Medical Sciences, vol. 3, pp. 1-5, Jan 2015.



- [2] Pedrelli P, Nyer M, Yeung A, Zulauf C and Wilens T, "College Students: Mental Health Problems and Treatment Considerations," *Academic Psychiatry*, vol. 39, pp. 503-511, Oct 2015.
- [3] Jenny K. Hyun, Brian C. Quinn, Temina Madon and Steve Lusting, "Graduate Student Mental Health: Needs Assessment and Utilization of Counseling Services," *Journal of College Student Development*, vol. 47, pp. 247-266, June 2006.
- [4] Gaihuan M A. "Construction of Fuzzy Evaluation Index System of College Student's Psychological Health Education," *China Public Security*, Feb 2012.
- [5] Agrawal R, Swami A, "Mining association rules between sets of items in large databases," *Acm Sigmod International Conference on Management of Data*, ACM, pp. 207-216, 1993.
- [6] Koh, H. C., and Tan, G., "Data mining applications in healthcare," *Journal of Healthcare Information Management Jhim*, vol. 19, pp. 64-72, 2005.
- [7] Yan, R W. Cong, L I. Lei and Y Y, "Improvement of Apriori Algorithm Based on Association Rules and the Application in the Insurance CRM of China," *Science Technology and Engineering*, vol. 21, pp. 6469-6473, 2009.
- [8] Pradhan, G. N, and Prabhakaran B, "Association rule mining in multiple, multidimensional time series medical data," *Journal of Healthcare Informatics Research*, vol. 1, pp. 92-118, 2017.
- [9] Jiang, H. B, and Yang, D L, "Application Research on Fast Discovery of Association Rules Based on Air Transportation," *International Conference on Service Systems and Service Management*, pp. 1-6, 2007. 723
- [10] Liu G, Jiang H, Geng R, and Li H, "Application of multidimensional association rules in personal financial services," *International Conference on Computer Design and Applications*, vol. 26, pp. 3877-3879, 2010.
- [11] Qi W, Yan J, Huang S, Guo L, and Lu R, "The application of association rule mining in college students' mental health assessment system. *Journal of Hunan University of Technology*," vol. 6, pp. 94-99, June 2013.
- [12] Meng Q, and Sha J, "Tree-based frequent itemsets mining for analysis of life-satisfaction and loneliness of retired athletes," *Cluster Computing*, vol. 2, pp. 1-9, May 2017.
- [13] Huang S. C, Zhong J D, and Wen-Jua Q. I, "Statistical analysis and association rule mining of application in college students' mental health," *Journal of Anyang Institute of Technology*, pp. 108-111, 2014.
- [14] Long Z, Feng H, Xue D, and Xiangjun D, "Positive and Negative Association Rules Mining for Mental Health Analysis of College Students," *Journal of Mathematics Science and Technology Education*, vol. 13, pp. 5577-5587, 2017.
- [15] Herawan T, Vitasari P, and Abdullah Z, "Mining Interesting Association Rules of Students Suffering Study Anxieties Using SLP-Growth Algorithm," *IGI Global*, pp. 24-41, 2012.
- [16] Han J, Pei J, and Yin Y, "Mining frequent patterns without candidate generation," *ACM SIGMOD International Conference on Management of Data*, pp. 1-12, 2000.
- [17] Wu X, Zhang C, and Zhang S, "Efficient mining of both positive and negative association rules," *Acm Transactions on Information Systems*, vol. 22, pp. 381-405, 2004.
- [18] Dong X J, Wang S J, Song H T and YC Lu, "Study on Negative Association Rules," *Journal of Beijing Institute of Technology*, pp. 978-981, 2004.
- [19] Zhang Y F, and Wang C, "Improved algorithm of mining association rules with negative items," *Computer Engineering & Applications*, vol. 44, pp. 169-171, 2008.
- [20] Dong X, Ma L, and Han X, "e-NFIS: Efficient negative frequent itemsets mining only based on positive ones," *IEEE, International Conference on Communication Software and Networks*, pp. 517-519, 2011.