

Student Performance Analysis using Machine Learning Algorithm

Shivangi Rawat¹, Himanshu Khosla²

^{1, 2}Information Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, India

Abstract: Student Performance Analysis is used to predict the performance of a student based on various factors like marks, attendance, staff rating, ragging, social media usage, extra-curricular and stress factor using the following machine learning algorithms.

a) Naïve Bayes

b) K Nearest Neighbour

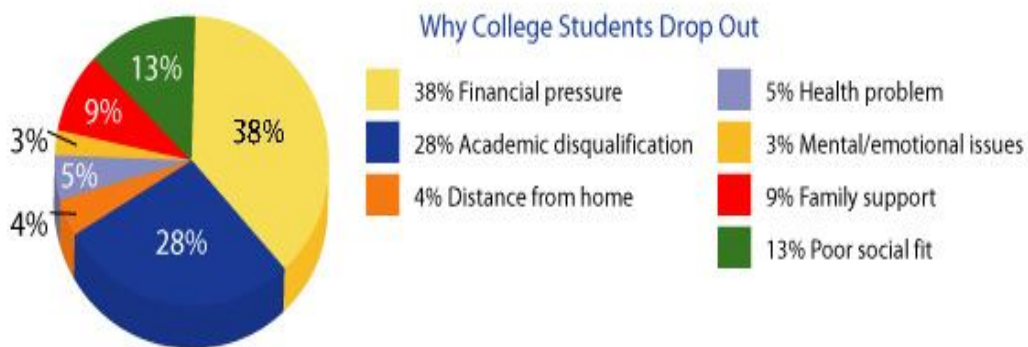
A comparative analysis is also performed on both the algorithms in terms of accuracy.

Keywords: Performance, Machine Learning, Naïve Bayes, K Nearest Neighbour

I. INTRODUCTION

A. The Issues Faced By The Educational System Today Are

- 1) Increase in drop-out rates at higher educational level
- 2) Negligence of non-academic factors on the performance of a student.



Survey conducted 4th quarter 2007 in a survey of 14,500 students at fifteen colleges by Duck9. Error Margin +/- 2%

Fig. 1: Reasons for college drop out

Machine learning algorithms can be used to predict the future performance of the student and a comparative analysis can be performed on both the algorithm to gauge the efficiency of both the algorithms.

II. OBJECTIVE

The objective of this model is to analyse the performance of a student using machine learning and data mining based on multiple academic and non-academic factors which would enable a student to improve his/her performance.

III. METHODOLOGY

The workflow of the data is as shown in the following diagram which allows to predict the performance of the student. All the fields (marks, attendance, staff opinion, extra-curricular activities participation, ragging, stress) are obtained from the database fed in by the user. Naïve Bayes and KNN algorithms are applied on the data to obtain the results.

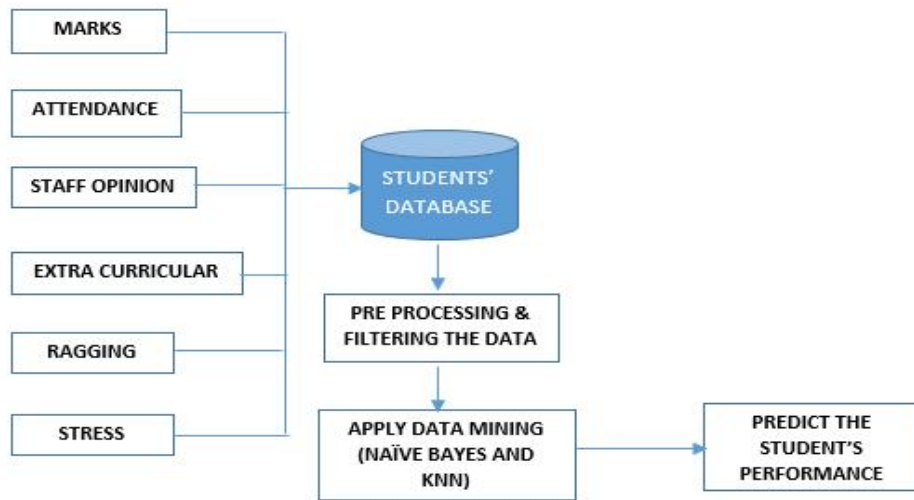


Fig. 2: Workflow

The following figures depicts the user interface which depicts the fields to be entered by the student.

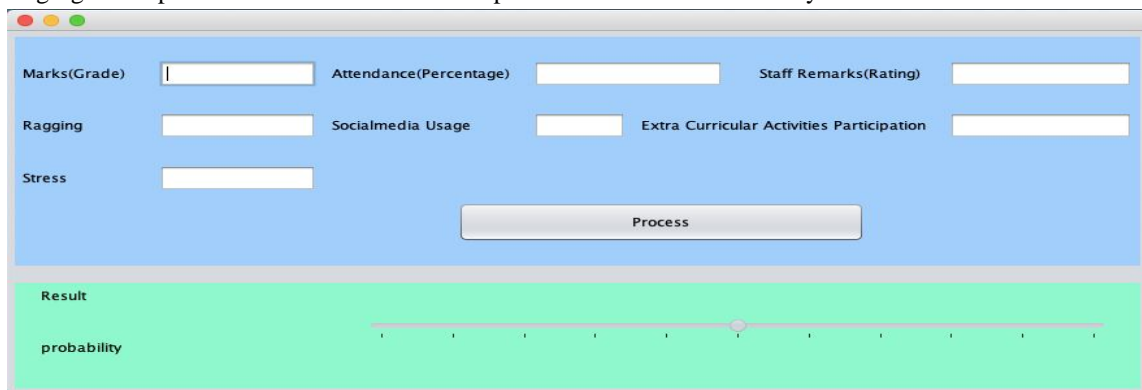


Fig. 3: User Interface

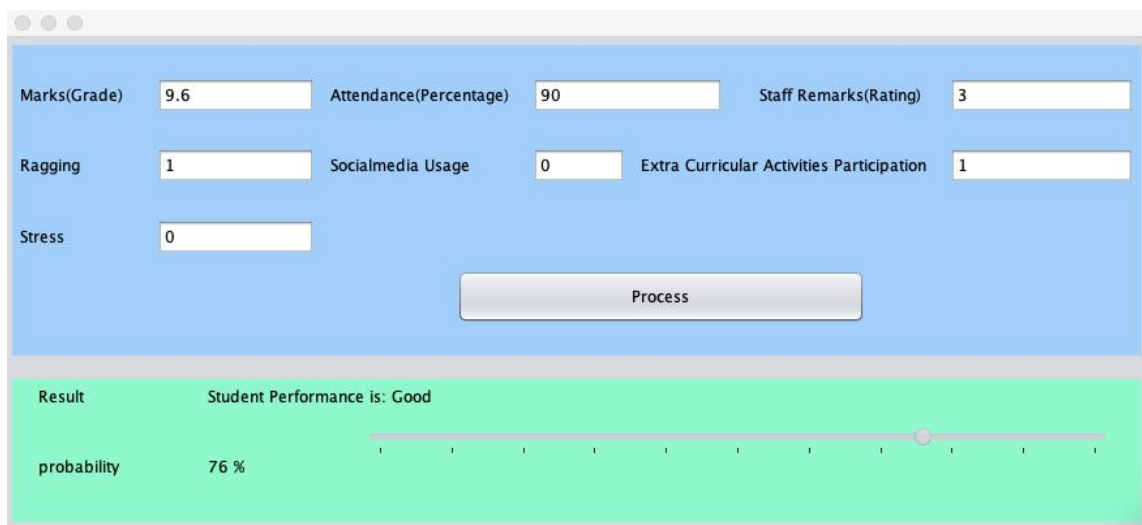


Fig. 4: Sample Input 1

```

Output - student (run) x
99
Naive Bayes classifier Result
no
.....
K Nearest Neighbours classifier Result
yes
.....
Accuracy in: Naive Bayes 1ms
Accuracy in: K Nearest Neighbours 0ms
.....
BUILD SUCCESSFUL (total time: 1 minute 2 seconds)
  
```

Fig. 5: Sample Output 1

The accuracy of both the algorithms is compared in terms of time taken by both the algorithm to execute and produce results.

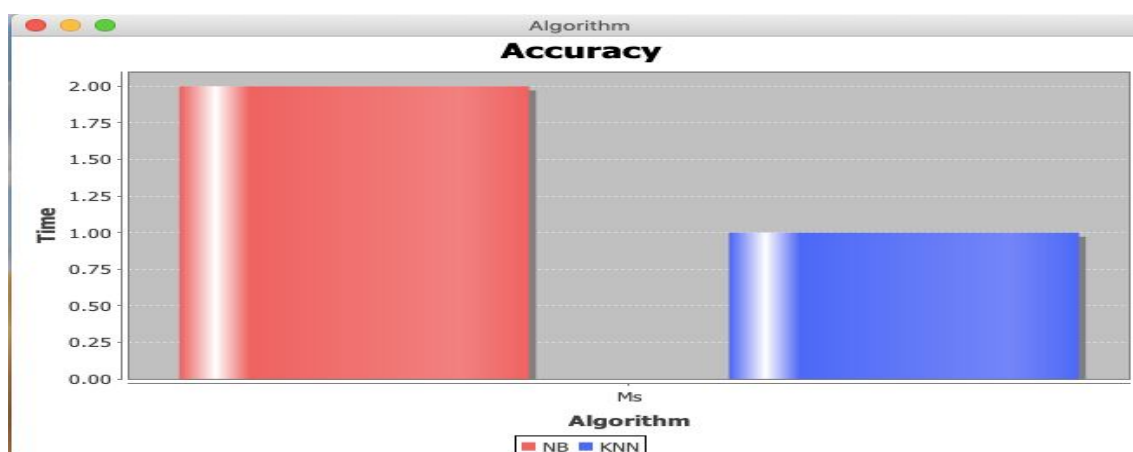


Fig. 6: Algorithm Accuracy

IV. CONCLUSION

With education playing a major role in the development of a country, this project would play a major role in the overall development of its students by:

- A. Predicting the performance of the students.
- B. Taking into consideration both academic and non-academic factors.

REFERENCES

- [1] Sharma,P.,Vaghela,D., and Parmar,K., "Performance Prediction of Students using Distributed Data Mining," IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIECS' 15,2015.
- [2] Osmanbegovic,E., and Suljic,M., "Data mining approach for predicting student's performance," Economic Review – Journal of Economics and Business, Vol. X, Issue 1 ,2012.
- [3] Banumathi,A., and Pethalakshmi,A., "A Novel Approach for Upgrading Indian Education by Using Data Mining Techniques," IEEE, DOI: 10.1109/ICTEE.2012.6208603,2012.
- [4] Anoopkumar,M., and Rahman,A., "A Review on Data Mining Techniques and Factor Used in Educational Data Mining to Predict Student Amelioration," IEEE, DOI: 10.1109/SAPIENCE.2016.7684113,2016.