Students Attendance Visualization

SURYANARAYANAMURTHY BOGGAVARAPU[1], K. MANIKANTA [2],

[1] Assistant professor, MCA DEPT, **Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh

[2]PG Student of MCA, Dantuluri **Narayana Raju College**, Bhimavaram, Andharapradesh

Abstract Understanding the academic performance of students in colleges is an essential topic in Education research field. Educators, program coordinators and professors are interested in understanding how students are learning specific topics, how specific topics may influence the learning of other topics, how students' grades/attendances in each course may represent important indicators to measure their performance, among other tasks. The use of data visualization and analytics is expanding in education institutions to perform a variety of tasks related to data processing and gaining into data-informed insights. In this paper, we present a visual analytic tool that combines data visualization and machine learning techniques to perform some visual analysis of students' data from program courses. Two educational data collections were used to guide the creation of i) predictive models employing a variety of well-known machine learning strategies, attempting to predict students' future grade based on grade and attendance previous semesters and ii) a set interactive layout that highlight the relationship between grades and attendance, also including additional variables such as gender, parents' education level, among others. We performed several experiments, also using these data collections, to evaluate the layouts' ability of highlighting interesting patterns, and we obtained promising results, demonstrating that such analysis may help the education experts to understand deficiencies on course structures.

Index Terms: - data visualization, machine learning, attendance, layouts

I Introduction

The importance of analytics and predictive methods in higher education, as well as the determining factors that contribute to academic performance are discussed in many research studies in order to improve the achievement of education goals, offer new modern opportunities for improving education system effectiveness and provide learning personalization.

Although the impact of learning strategies and gender differences on academic performance are addressed by Ruffing et al. (2015), the factors underlying the prediction of academic performance are still of great interest in educational psychology. One reliable task is to ask the school for some anonymized information about previous students to obtain useful information to perform analytic tasks. In this sense, novel analysis strategies are useful in comprehending educational scenarios involving students performance and related factors, guiding decision making by educational experts.

In this paper we present a computational system for educational data analysis. We intend to demonstrate how predictive analytics and data visualization techniques can provide powerful decision-making aids for educators and school administrators, by providing means to identify and explore trends and patterns on these data, as well as to comprehend the real situation regarding a specific education scenario. We believe that a visual analysis tool employing machine learning and information visualization techniques improves the comprehension, by educational experts, of student's behavior on subjects over the semesters, guiding them in defining effective strategies to mitigate related deficiencies. Our main goal is to investigate the ability of the system's tools in addressing the following research questions:

Is it possible to predict the students' grades using grades and attendance in subjects from past semesters? Does students' attendance impact on their grades in the semesters? Research shows that attendance is one of the

most important factor in students' academic performance and achievement (Jones 2006; Kassarnig et al. 2017). Considering students' attendance, we employed a variety of machine learning models to predict students' data trends over several semesters and we compared this prediction to the real data, in order to measure their accuracy.

Is there any correlation between students' gender and their performance on different subjects?

We considered gender in a performance statistical evaluation. As this factor as well as race, ethnicity, educational, and psychological factors are also addressed in some studies (Dee 2005; King et al. 2002; Wilson and Shrock 2001). Do external factors, such as parents' education level impact on students' performance in school? The influence of external factors is also discussed by Gooding (2001). We employed a "Multidimensional Projection" technique to explore the structure of the relationship among students in terms of similarities. The idea is to identify profiles and/or outliers that may explain their performance in the courses over the semesters.

In order to evaluate our system, we considered two educational data collections, one of them containing students records from Exact science programs of the Faculty of Computing of Federal University of Uberlandia, Brazil, and the other one containing students records from two public schools of Portugal.

2 Literature survey

Althrough, there are several works addressing the use of data visualization by instructors to enhance learning (Klerkx et al. 2014; Anaya et al. 2016; Thompson et al. 2013), not many studies are found that use visualization approaches to analyze educational data.

The selection, processing, visualization, and analysis of multiple learning and learning environments elements, as well as the links between them have been discussed in Thompson et al. (2013) to provide a comprehension of the learning processes in complex learning environments. Performing limited studies, Lacefield et al. (2018) explore applications of machine learning, predictive analytics, and data visualization to student information available to educational decision makers. They concentrated more on demonstrating individual academic performance histories to identify "at-risk" students in real time for advising, academic coaching, and other support services.

DeCotes (2014) used a Heatmap visualization to represent students' grade performance for different course pairs. In this work three different course pairs are compared and the results have shown that different course pairs yield different behaviors. For instance, the study indicates that students who did poorly in level one course will receive similar poor grades in the level two course. Using this visualization, they could see the transition of grades from one level to another level more clearly.

In terms of predictive analytics, Urrutia-Aguilar et al. (2016) employed also a logistic regression model for the prediction of variables that have impact on the academic performance of first year biomedical students. Soule (2017) also employed multiple logistic regressions to improve prediction techniques regarding the future performance of students in selected university courses and his study showed that in all cases, logistic prediction models matched or exceeded the performance of current prediction methods while using an equal or lesser number of explanatory variables.

Fernandes et al. (2019) performed a descriptive statistical analysis to gain insight from the academic performance of students. Gutierrez et al. () used Random Forest models to predict

students' academic performance in different engineering subjects. We employed some of these predictive analytics techniques in our system. Fernandes et al. (2019) used Classification models based on the Gradient Boosting Machine (GBM) to predict academic outcomes of student performance at the end of the school year and they showed that 'grades' and 'absences' attributes were the most relevant for predicting the end of the year academic outcomes of student performance. The analysis of demographic attributes revealed that 'neighborhood', 'school' and 'age' are also potential indicators of a student's academic success or failure. We also investigated 'grades', 'absences' and 'age' factors in our analysis employing visualization and machine learning approaches. We also performed descriptive statistical analysis to gain insight from data and used visualization techniques to show the significant differences.

Academic visual analysis system

We developed a web-based system[Footnote1] that provides a set of interactive visualization tools to perform an academic data visual analysis. An overview of our web-page can be found in Fig. 1. The layouts and interaction tools were developed using Javascript D3.js[Footnote2] and Python Scikit-learn[Footnote3] libraries. Two educational data collection were used to guide the creation of the layouts, and also to evaluate their capability of highlighting interesting patterns. This section describes these collections, and the layouts developed for the analysis.

3 Implementation Study

Existing system is a manual entry for the students. Here the attendance will be carried out in the hand writing registers. It will be tedious job to maintain the record for the user. The human effort is more here. The retrieval of the information is not as easy as the records are maintained in the hand written registers.

This application requires correct feed input into the respective field. Suppose the wrong inputs are entered, the application resists to work. So the user find it difficult to use.

3.1proposed methodology

In this project we are reading attendance data from dataset and then displaying pie chat for present, absent, authorised and non-authorise absent. In histogram chat we are calculating attendance students per semester week wise.

3.2 Methodology

1. Upload Attendance Dataset

Using this module we can upload 'sample.csv' file and then click on 'Open' button to load dataset

2. Pie Chart Attended, Not Attended

In this Module screen calculate attended and not attended students and then plot in graph

3. Week Wise Attendance Histogram Chart

calculate week and semester wise attended students.

wearing.

4 Results and Evolution Metrics

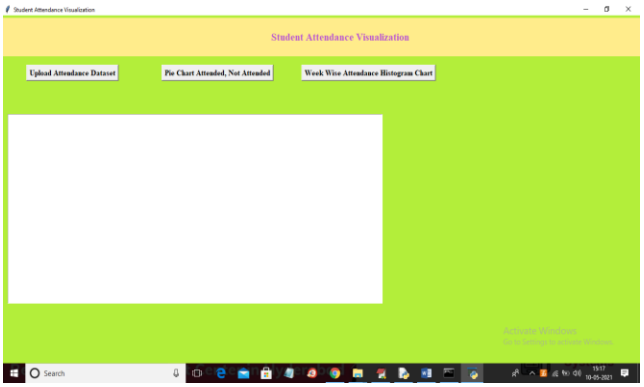

Fig 1: In above screen click on 'Upload Attendance Dataset' button to upload dataset.

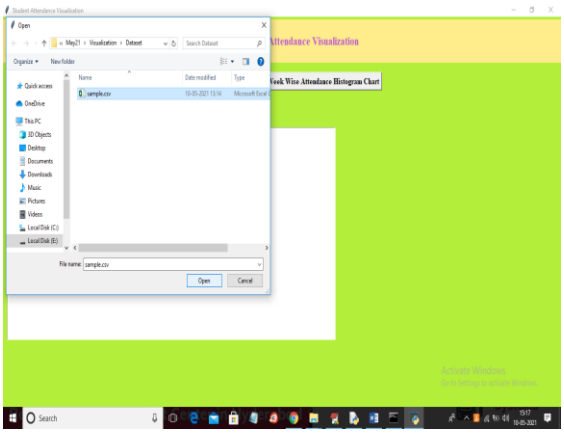

Fig 2:- In above screen selecting and uploading 'sample.csv' file and then click on 'Open' button to load dataset and to get below screen

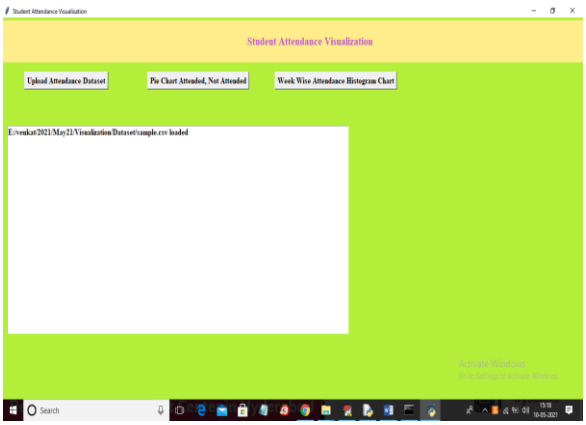

Fig 3: In above screen dataset loaded and now click on 'Pie Chart Attended, Not Attended' button to calculate attended and not attended students and then plot in graph

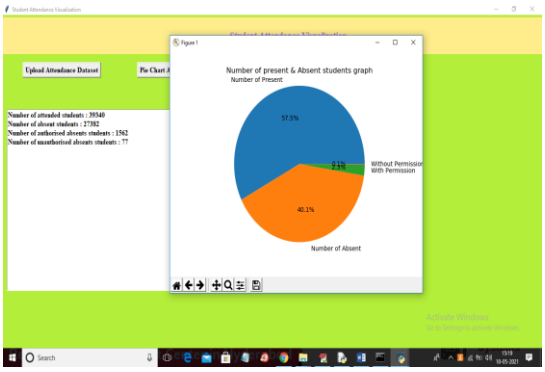

Fig 4:- In above screen in text area and in graph we can see attended, not attended, with and without permission and now close above graph and then click on 'Week Wise Attendance Histogram Chart' button to calculate week and semester wise attended students.

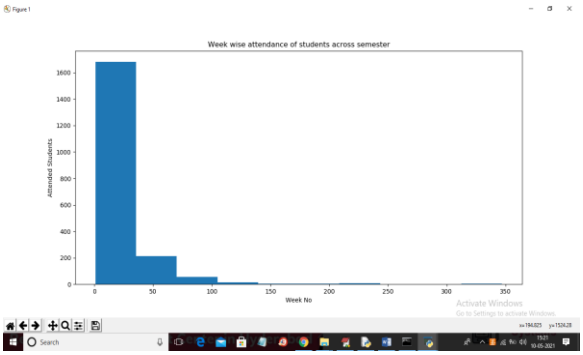

Fig 5:- In above graph x-axis represents week no and y-axis represents attended students

5 Conclusion

We developed a web-based visual analytics system that includes several visualization approaches to analyze academic data. Our initial results are reasonable enough to make sense out of this type of multivariate data that helped us to gain insight for better future decision making in the academic environment. We were able to show the relationships between grades and attendance, grades and genders, grades and parents' educational level and students populations and genders. Additionally, we implemented a variety of machine learning models that predict the performance of students based on their attendance (absence rate), and we employed a visualization technique to check the accuracy of different models.

6 References

1. A. R. Anaya, M. Luque, M. Peinado, A visual recommender tool in a collaborative learning experience. Expert Systems with Applications. **45:**, 248–259 (2016).
2. L. Breiman, Random forests. Springer. **1**(45), 5–32 (2001).
3. P. Cortez, A. M. G. Silva, *Using data mining to predict secondary school student performance* (EUROSIS-ETI, Ostend, 2008).

4. M. B. DeCotes, Data analytics of university student records (2014). Master's Thesis, University of Tennessee.
5. T. S. Dee, A teacher like me: Does race, ethnicity, or gender matter?American Economic Review. **95**(2), 158–165 (2005).
6. H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, V. Vapnik, in *Advances in Neural Information Processing Systems*. Support vector regression machines (MIT PressCambridge, 1997), pp. 155–161.
7. R. Etemadpour, R. Motta, J. G. de Souza Paiva, R. Minghim, M. C. F. de Oliveira, L. Linsen, Perception-based evaluation of projection methods for multidimensional data visualization. IEEE Transactions on Visualization and Computer Graphics. **21**(1), 81–94 (2015).
8. E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, G. V. Erven, Educational data mining: Predictive analysis of academic performance of public school students in the capital of brazil. Journal of Business Research. **94:**, 335–343 (2019).
9. E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, G. Van Erven, Educational data mining: Predictive analysis of academic performance of public school students in the capital of brazil. Journal of Business Research. **94:**, 335–343 (2019).

E. Fix, J. Hodges Jr, Significance probabilities of the wilcoxon test. The Annals of Mathematical Statistics, 301–312 (1955). https://doi.org/10.1214/aoms/1177728547