

Education Data Mining, Visualization and Sentiment Analysis of Coursera Course Review

Dhaval Bhoi¹, Amit Thakkar²

¹U & P U. Patel Department of Computer Engineering,

²Department of Computer Science & Engineering,

Chandubhai S. Patel Institute of Technology, Faculty of Technology & Engineering,

Charotar University of Science and Technology,

Changa – 388421, Gujarat, India.

Abstract: Objective: No Decisions are good or bad they are taken based on the available data. It is very much essential to represent the data in the right form to the right people and at the right time. Higher Engineering Institutes (HEI) is having a plethora of information available to them. Most of the available data are not used properly and remain just as dead storage.

Methods: In this study, we have shown the importance of data visualization using a case study on Coursera review dataset. Different useful tools that support improving an Education System are summarized. Sentiment analysis is performed for coursera course review dataset using deep learning method. At the end, dashboard is also created to visualize student data using powerBI tool.

Results: Uses of different visualization tools can help to improve the education system and its performance. The Sentiment expressed by students

will help to improve the teaching-learning process and research contribution significantly as they are the major components for evaluation when any HEI wants to receive NAAC [National Assessment and Accreditation Council] approval for benefitting all stakeholders of the HEI.

Conclusions: Proper analysis of available data and their proper visualization can help us to improve the education system to a great extent in terms of improving the most important factors like student teaching- learning and their placement to make their future bright. Students expressed sentiments are also key features to analyze the success of the teaching-learning process for both teachers and students as well. We have also used our institute students' data to generate a dashboard that contains student information from a different perspective that can help higher authorities to make better fruitful decisions.

Keywords: Education Data Mining, Dashboard, Data Visualization, Sentiment.

Dhaval Bhoi

U & P U. Patel Department of Computer Engineering,
Chandubhai S. Patel Institute of Technology, Faculty of
Technology & Engineering, CHARUSAT, Gujarat, India
dhavalbhoi.ce@charusat.ac.in

1. Introduction

Information and technology play a vital role in many aspects including education (Cabada et al., n.d.), as education is an integral part of one's life (Rajarapolu et al., 2022). Information about various stakeholders in higher education institutions, courses and technological applications aids in the

improvement of the educational system as a whole. More than big data is required to lead and build any organization, especially an educational institution. You might be missing out on opportunities to improve results if you can't effectively evaluate the implications of numerous data sets. The main purpose of education system is to enable people to alter themselves and society as a whole (Jha, 2020). Data mining is the process of extracting hidden patterns from given data (Sapountzi & Psannis, 2018). Combinations of this mined data through various filters make analysis better. There is a famous proverb saying "A picture is worth a thousand words," (A Picture Is Worth a Thousand Words - Wikipedia, n.d.), which means a person can represent the thing visually better instead of writing about. Hence, using the appropriate tools and technology, we can represent the education data in some meaningful and appealing form. In the current era, it is now becoming inevitable to use data visualization and analytics techniques in education in order to turn huge amounts of data into actionable insight for making smart and better decisions.

Effective data visualization and analysis of stockholders' opinions are the two most important parameters to be considered while we are in the process of making an effective education system. The primary objective of this research is to use effective data visualization techniques to find meaningful correlations between parameters of interest. The other prime objective is to extract sentiments from the feedback of various stockholders for effective decision making in education (Zentner et al., 2019).

The rest of the paper is laid out as follows: Section 2 discusses numerous tools, as well as a case study on how to create various data analysis and visualizations, as well as their sentiment analysis significance in the education system. Section 3 demonstrates how a single Power BI Dashboard can be used to handle and depict many types of analysis and visualizations, which can aid in the improvement of the educational system.

2. Data Analytics, Visualization And Sentiment Analysis

Online education becomes unavoidable due to current pandemic situation (Bhadri & Patil, 2022). Education data analytics can be used to strengthen decision-making and provide actionable insights. Due to large availability of education data, it is required to

find out different ways to deal with and find out insight from the available data. Data Analytics and visualization plays an important role for decision makers. Education data visualization and infographics tools are an excellent method to keep students and readers interested. Different tools are identified for creating various data visualization elements like infographics, maps, graphs, charts, methodology. These tools along with their uses are described in the below Table 1.

To show the effectiveness of various data visualization technique a dataset about various courses on coursera is selected for experimental analysis from Kaggle platform, which is a subsidiary of Google Limited Library Company (LLC) and is an

Table. 1: Supportive Tools to improve Education

Tools	Uses
Animoto	It is a computer-generated web tool that enables users to quickly make high- quality educational videos.
ClassDojo	It aids instructors in the creation of online collaborative groups, the administration and distribution of educational resources, the assessment of student performance, and communication with parents.
ck-12	This platform offers an open-source interface that allows users to create and distribute content related to education through the internet that can be custom-made.
Edmodo	Teachers can create online collaborative groups to provide educational materials. It can also measure pupil performance, and share with parents.
eduClipper	Teachers and students can use this platform to share and investigate references and instructional materials.
kahoot	Teachers may use this tool to generate quizzes, conversations, and surveys to supplement academic teaching.
Projeqt	It can help educators to share academic presentations with students.
TED-Ed	People can take an active role in other people's learning processes.
Thinklink	Educators may use it to create interactive graphics using music, audio, words, and photos.
Socrative	It allows professors to construct exercises or instructional activities that students may do on their cellphones, computers, or tablets. Teachers may view the outcomes of the exercises.
Storybird	Teachers may generate audiovisual material that conforms to instructional needs using this interface, which is user-friendly and practical.

online open community of data scientists and machine learning experts (Kaggle: Your Machine Learning and Data Science Community, n.d.). We have analyzed the data from different perspective to extract and highlight meaningful information from the data. Below we have discussed data analyzed using various data visualization technique to extract meaningful highlight which may useful for effective decision making related to grow education system.

A. Gender wise Analysis:

As demonstrated in Figure 1, this type of analysis aids in determining the number of male and female students who use the coursera platform. In comparison to male students, we can see how much female students have progressed, and vice versa. We discovered that 81.40% of men and 16.81% of women registered in coursera. 1.76% of people did not want to reveal their gender, which is an interesting fact. This information can be used by decision-makers to encourage female students more to enroll in these worthwhile courses.

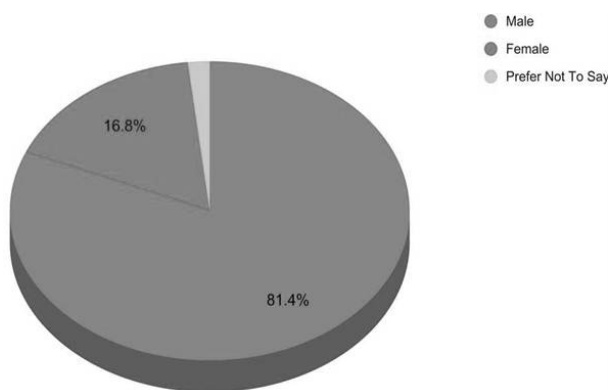


Fig. 1: Students using Coursera Platform

B. Student Education Level Analysis

Secondly, we have also analyzed the education level of students those are using Coursera Platform as shown in the Figure 2. The majority of them have a Master's or Bachelor's degree. People with a PhD degree are ranked third. The information is assessed based on the educational level of students who have utilized the Coursera platform to learn. The degree holders are represented, as well as their preference for online learning. According to the findings, students with a Master's degree are more inclined to choose online learning through platforms such as Coursera.

Many students at the undergraduate level have also chosen to learn online using the Coursera platform. Due to COVID-19 and the National Education Policy 2020, this number may rise in the near future. Due to the requirement of completing coursework related to their research work and receiving widely accepted and trusted course certification, a large number of Ph. D students also choose online courses. Despite this, the analysis shows that these online platforms are still ineffectively used by school- aged children. The reason for this could be that appropriate courses for school students are not available on the coursera platform. On the coursera platform, even professional degree courses are not promoted.

Education Level of Students on Coursera

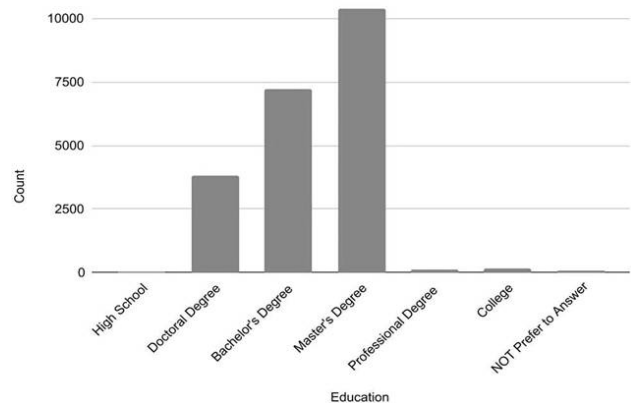


Fig. 2: Education Level of Students on Coursera

C. Coursera courses country wise popularity

When we looked at the popularity of courses across countries as part of our next analysis, we came up with the following result, as shown in Figure 3. The United States of America was first, India was second, and China was third. This research will aid in determining the popularity of Coursera courses in various countries. This type of analysis could aid in

Top Countries - As per Coursera Popularity

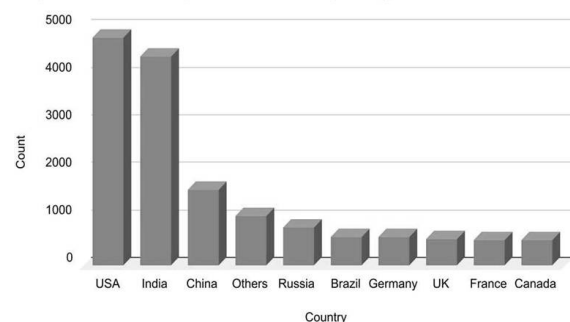


Fig. 3: Coursera courses popularity country wise

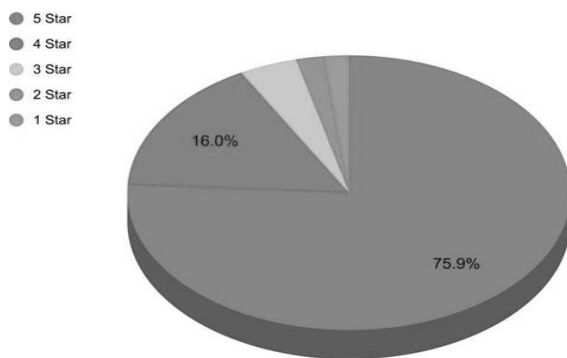


Fig.7: Sentiment Analysis of Machine Learning Course

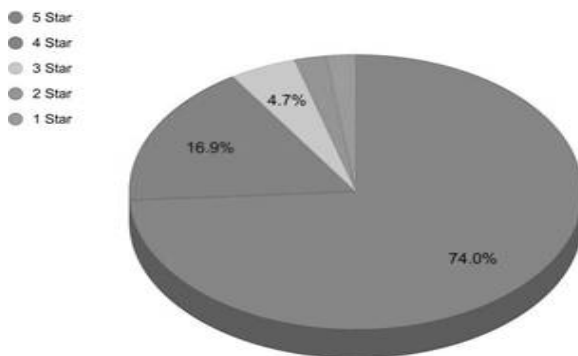


Fig. 8: Sentiment Analysis of Python Course

F. Labels of Reviews by Course

As part of our next analysis, we looked at Machine Learning and Python Course reviews based on star ratings. A course's rating ranges from 5 to 1, with 5 being the highest and 1 being the lowest. For the Machine Learning and Python Programming courses, we received the results shown in the below Figures 7 and 8, respectively. 1.93% more users gave five-star ratings to machine learning course compare to python

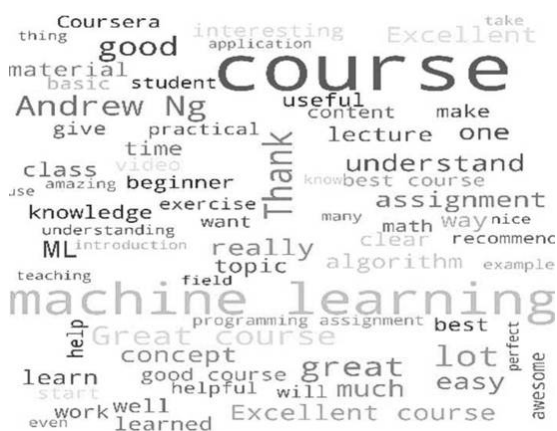


Fig. 9: Word Cloud for Machine Learning Course



Fig. 10: Word Cloud for Python Programming Course

course. They are more positive and preferred for taking machine learning course in compare to python course. This type of analysis can aid in comparing the popularity of different subjects among students. This can be used to identify trends among students at the department, institute, or university level.

G. Review comparison of courses using Word Cloud

As a part of our sentiment visual experiment, we have also generated word cloud from the comments or sentiment expressed by different coursera users as shown in below Figures 9 and 10 for machine learning and python course respectively. More frequent words appearing in the user comments are highlighted in large fonts while less frequent words emerge as fonts with small appearance. Emerging patterns for both courses can be seen clearly. This analysis will help to

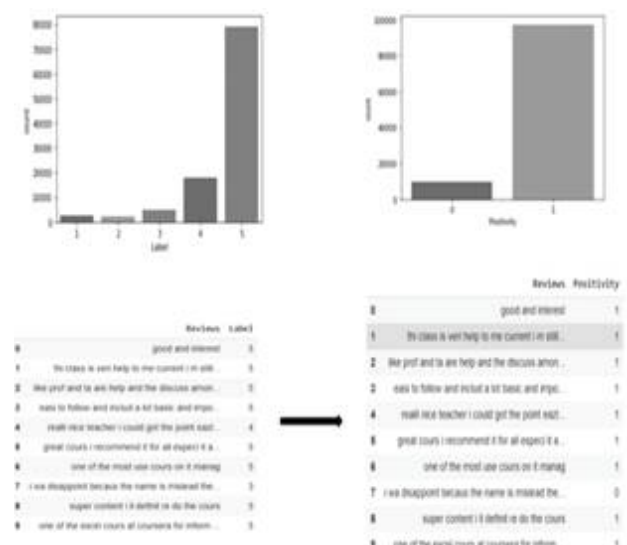


Fig. 11: Dataset Details and Preprocessing

identify feelings of users for a particular course. For example, Andrew Ng, an expert of machine learning course emerges as one pattern in the below Figure 9.

We have used deep learning models to perform sentiment classification in a subsequent education data sentiment analysis experiment. We chose the Coursera course review dataset, which contains 1, 07,016 review samples, for this. Each review has three attributes: a review id, a review text, and a star rating ranging from 1 to 5. Where 5 denote a very positive review and 1 denotes a very negative review. To turn the problem into a binary classification problem, we considered reviews with a rating of 5 and 4 as positive, and reviews with a rating of 1, 2, or 3 as negative. We converted the class label of positive reviews to 1 and the class label of negative reviews to 0. The LSTM

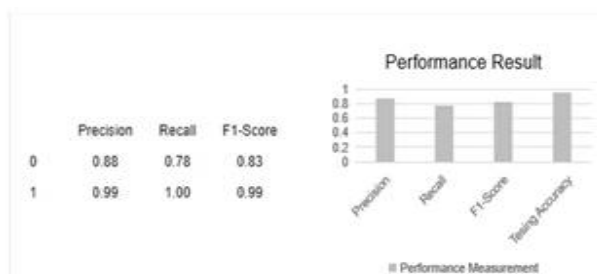


Fig. 12: Sentiment Classification Performance Analysis

deep learning model is used because of its ability to retain textual review information over time. Data Preprocessing is crucial for achieving better sentiment classification performance. We used basic review preprocessing tasks like removing stop words, stemming, and lemmatization to achieve this. The details of the dataset are shown in Figure 11 below. It

Class	Precision [%]	Recall [%]	F1- Score [%]
0	88.00	78.00	83.00
1	99.00	100.00	99.00

Table. 2: Performance Parameters and Result

shows that our data set has a higher number of positive reviews than negative reviews.

To improve feature representation, we used the Glove (glove.twitter.27B.100d.txt) word embedding technique to perform sentiment classification. Because of its ability to retain long-term review text sequences, the LSTM deep learning model is used to achieve better classification results. Figure 12 below shows the achieved performance results.

For the proposed methodology, we were able to achieve a testing accuracy of 95.47%. As accuracy is not the only parameter in order to measure the performance of proposed approach. As shown in Table 2, we also received better precision, recall, and F1-Score for negative class (0) and Positive Class Review(1).

3. Student Dashboard Creation Using Power Bi

Here in this section, we have explained the importance of Dashboard, its uses for the betterment of the education system as a whole. To demonstrate the effectiveness of our approach we have used and focused mainly on student data, but can be extended in future for other users, for an example, faculty, parents, industry.

There are two forms of student information.

A. Student General Information

Students' generic information includes their ID number, name, address, phone number, parent information, cast information, blood group, and many more.

B. Student Academic Information

Academic information about a student comprises his previous performance prior to enrolling in a higher education institution. It also includes the student's current status in terms of CGPA (Cumulative Grade Point Average), as well as his performance improvement in each semester during his academic career.

Dashboard helps us to keep all the relevant information on a single page. We can filter data on one visualization and it gets reflected on other related visualization. This dashboard enables educational professionals to track and verify performance, report generation and set assessments and objectives for forthcoming tasks (Nadj et al., 2020).

Other key advantages of Dashboard:

- Charts and graphs provide a graphical depiction of performance [Performance Analysis]
- Provide ability to recognize patterns or trends [Student Progression]

- A simple method of determining efficiency [Effectiveness Analysis]
- With a single click, you may produce extensive reports [Report Generation]
- The ability to make better decisions based on more information [Decision Making]
- Education systems, campaigns, and acts are all fully visible [Actionable Insight]
- Outliers and correlations in data can be identified quickly. [Correlation Analysis and Outlier Detection]

We have utilized Power BI, a tool for better visualization viewpoint to construct the dashboard. It is a set of software services, applications, and connections that work together to transform heterogeneous data into logical, visually engaging, and interactive insights. Data can have many different forms, such as a simple Microsoft Excel spreadsheet or a collection of data warehouses (Krishnan, 2017). We can connect to various data sources with Power BI, see and analyze what matters, and share our findings with everyone interested and attached with the analysis.

Data visualizations can enhance the impact of our communications for our target audiences and makes data analysis results more appealing. It can connect the data of education system across all departments

and divisions. Visualization helps you to see and interpret enormous amounts of data quickly and efficiently. It facilitates in the better comprehension of data so that its influence on the education system can be assessed, and it graphically conveys the knowledge to external and internal viewers.

It is nearer to impossible to make decisions in without proper representation of data and its visualization. Decision-makers can use available data and their insights to make better decision and actions.

Different people need data to view from different perspective. Say for an example, higher authority may be interested in overall performance of students studying at the university, while individual teacher may be interested in knowing how students are performing in his or her individual course. Student counselor may be more interested in analyzing the progress of the students assigned to him or her while head of the departments will be interested in result of their own department. As a case study, we have utilized Power BI to create a dashboard to visualize the student data insights. Dashboard can be configured as per the need and requirement of educational institute. Figure 13 shows a demo sample screenshot of the same. We have used different filters to select particular department, semester or counselor to view data on the dashboard.

Data visualizations may help with a variety of educational issues. All firms should employ data visualizations technology to reap revolutionary

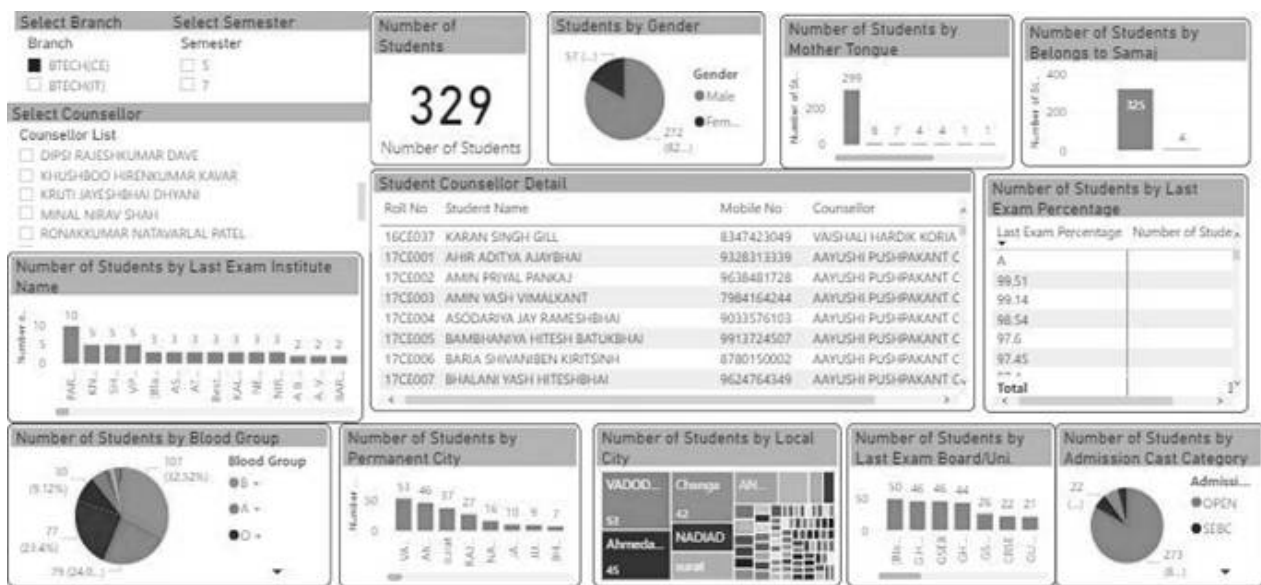


Fig. 13: Student Dashboard

benefits in their critical areas of operations. Furthermore, education dashboards can be used for the board of directors, parents, and the general public. With a greater emphasis on openness in budgeting, student evaluation, and higher engineering institute site performance, districts are increasingly using public-facing dashboards to demonstrate student accomplishment.

This student dashboard can aid in the visualization and analysis of student information. Students from several semesters can be visualized. Data analysis can be seen, including gender discrimination, geographical analysis, and student counsellor details. Only the students assigned to counselee will be seen on the dashboard when a specific name of the student counsellor is selected. We can look at the data from a variety of angles and perspectives.

4. Conclusions

This research study helps to increase user awareness to understand various support tools for improving the educational system. The use of online tools, education data mining, and education data visualizations techniques could be a huge step forward in resolving the plethora of issues that educational institutes face across the country or around the world. Decision makers can understand learners well to make appropriate decision. Sentiment analysis, as well as the effective correct classification of learners' feelings, can aid in better understanding regarding how they feel about a particular subject, faculty, department, institute or university. This type of analysis can be beneficial to educational institutions in achieving their various short- and long-term objectives. Short term goals could be improving promotional activities based on analysis made from different visualization; Identifying stakeholders' reviews can boost education system to make corrective and necessary changes in the existing system. Long term goal could be to become world class to attract global learners. Different decision makers require data to be viewed from distinct viewpoints. This type of education system innovation will benefit people at all levels because they will have timely access to the right data and reports, and they will be able to generate trusted knowledge and understandings using a single and simple Dashboard, that will help to transform programs, curriculums, and student outcomes, and they will be able to deliver desired fruitful results faster.

Acknowledgments

The authors thank the Principal and Dean of the Faculty of Technology and Engineering, as well as the Head of the U & P U. Patel Department of Computer Engineering at CSPIT, Faculty of Technology and Engineering, Charotar University of Science and Technology, Changa, for their continuous suggestions, encouragement, guidance, and support throughout the research. We would like to thank Management for their unwavering moral support and encouragement throughout the years.

References

- [1] A picture is worth a thousand words - Wikipedia. (n.d.). Retrieved February 10, 2022, from https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words
- [2] Bhadri, G. N., & Patil, L. R. (2022). Blended Learning: An effective approach for Online Teaching and Learning. *Journal of Engineering Education Transformations*, 35(Special issue), 53–60.
- [3] Cabada, R. Z., Lucia, M., Estrada, B., & Oramas, R. (n.d.). Mining of educational opinions with deep learning. <https://www.researchgate.net/publication/331877377>
- [4] Jha, S. (2020). A case study of implementation of active - cooperative learning approaches introduced through a faculty development programme and their effects on the pass percentage of undergraduate engineering students. *Journal of Engineering Education Transformations*, 34(1), 7 – 11. <https://doi.org/10.16920/jeet/2020/v34i1/155007>
- [5] KABIR, A. I., KARIM, R., NEWAZ, S., & HOSSAIN,
- [6] M. I. (2018). The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R. *Informatica Economica*, 22(1/2018), 25 – 38. <https://doi.org/10.12948/issn14531305/22.1.2018.03>

- [7] Kaggle: Your Machine Learning and Data Science Community. (n.d.). Retrieved February 10, 2022, from <https://www.kaggle.com/>
- [8] Krishnan, V. (2017). IR @ INFLIBNET: Research Data Analysis with Power BI. <https://ir.inflibnet.ac.in/handle/1944/2116>
- [9] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.
- [10] Nadj, M., Maedche, A., & Schieder, C. (2020). The effect of interactive analytical dashboard features on situation awareness and task performance. *Decision Support Systems*, 135, 111333. <https://doi.org/10.1016/J.DSS.2020.113322>
- [11] Rajarapolu, P., Bansode, N. V., & Katkar, V. (2022). ICT-A Tool to Enhance Teaching Learning Activity in Technical Education. *Journal of Engineering Education Transformations*, 35(Special Issue), 14–18.
- [12] Sapountzi, A., & Psannis, K. E. (2018). Social networking data analysis tools & challenges. *Future Generation Computer Systems*, 86, 893–913. <https://doi.org/10.1016/j.future.2016.10.019>
- [13] Zentner, A., Covit, R., & Guevarra, D. (2019). Exploring Effective Data Visualization Strategies in Higher Education. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3322856>.