VisuaLizer: Data Visualization and Analytics

Kalpesh N Raundal, Jay Rawlani, Ambuj Mishra, Shivam Kamble, Shubhangi Kale

School of Computer Engineering and Technology, MIT Academy of Engineering

ABSTRACT

Data viewing involves presenting data in a graphical way that makes information easier to understand. Processes large datasets in the form of charts and graphs or other viewing strategies. It is a very effective way to communicate information or data details about our data. It helps us determine the course of action and the required facts and information about the data. The visual system should perform data reduction, modify the data and produce the original data and its details on the screen, actually making the data come alive. It should employ charts and graphs to display the data and provide them in an easy-to-understand format. Data visibility should be easy to understand unless a user can find incorrect interpretations of the data. Data View has become one of the fastest-accepted technologies in the world as it has become a fast, easy way to transfer ideas in a universal format. Information Identification not only provides us with basic and simple answers to our questions, it also conveys complex ideas clearly, accurately, and effectively. There are several data detection tools available such as power bi, google analytics, tableau, Domo, etc.

I. INTRODUCTION

Data viewing involves presenting data by image or image which helps us understand the image easily. It helps to explain the facts and decide what to do. The main motive of the detection system is, it performs data reduction, and modifies and generates original on-screen data. It should employ charts and graphs to display the data and provide them in an easy-to-understand format. It has become a fast, easy way to convey ideas in a universal format. It must communicate complex ideas clearly, accurately, and effectively. This representation of data in form of graphs and charts will be used for analyzing the data, and these insights can help in taking business decisions and finding the trends and patterns of the data.

The Paper consists of ten major sections.

1) Abstract
2) Introduction
3) Study of Related Work
4) Methodology
5) System Design and Modelling
6) Results
7) Discussion
8) Conclusions
9) Acknowledgement
10) References

II. STUDY OF RELATED WORK

[1] The system solves the first challenge by training a binary classifier for deciding whether the particular visualization is good or not. It discusses Machine learning techniques for suggesting graphs for the users.[2] The paper mainly focuses on different visualization techniques, applications, and challenges faced in visualizing data. Also talks about issues related to performance, size of operability, and diversity of data.

[3] To achieve the most understandable information of data, visualization has been made. It discusses the information about auto charting, Network diagrams, Sankey diagrams, and word cloud. Also, about the data visualization platforms.[4] In this paper, fourteen distinguished scientists have expressed their thoughts regarding the top future research challenges in Big Data visualization and analytics and emerging applications in Big Data.

[5] This handbook covers computational statistics in the field of big data. This paper discusses the history of data visualization, graphics, Data views, and methodologies of data visualization.[6] The paper is based on the machine learning library 'Seaborn'. It discusses statistical graphics and their functions and different APIs. Seaborn automatically maps the data values for visual attributes like color, size, and style.

[7] They focus on Data Visualization, Graph Drawing, Data Filtering, Clustering, Force-directed, and Stock Investment Analysis in the field of the stock exchange market.[8] They discussed different criteria for data visualization, Comprehension, and Embellishment. Also does the comparative detailed analysis on Data Visualization.

[9] They talked about various improvements in data Visualization. They discuss the arguments between industry and governments for adopting the Data Visualization in Big Data.[10] We got to learn about emerging technologies which contribute to the analytics worlds of business intelligence.[11] The Maturity Model of business Intelligence was mentioned in this paper. Also, it describes types of business intelligence models like strategic BI, Tactical BI and operational or Real-time BI.

III. METHODOLOGY

There are two algorithms used in the VisuaLizar: -

Algorithm for Data Cleaning and Transformation

1. Start
2. Defining parser arguments in main function
3. Input dataset
4. Applying Auto-clean()
a. Replacing NANs with Medians if possible
b. Else Replacing NANs with Mode
c. Else if mode can't be computed, using the nearest valid value
d. If string is there, then encoding all strings with numerical equivalents and applying same process
5. Returning dataframe
6. Inputting the input file as training data and above output generated file as
    testing data to Autoclean_cv() to perform a series of automated data cleaning

    transformations on the provided training and testing data

7. Get the final dataframe as cleaned dataset
8. Stop

Algorithm for Questions Recommendation
Name :- Comaparing and Extracting Data
Input :- Column Data
Output :- Questions Recommendation
Start
 1. For i in Questions_Dataset
     1.1 For j in Questions_column
         1.1.1   count = 0
         1.1.2   if value in j in columns_name
                 1.1.2.1 count++
     1.1.3 If count equals to rows in Number_of_rows_Question_dataset
                 1.1.3.1 count =0
     1.1.4 store value of count in arr
 2. Create an array with 0's and name it as Row_Array
 3. For i in range(length(Arr))
         3.1   Row_Array[i%Number_of_Rows]   = Row_Array[i%Number_of_Rows] + arr[i]
 4. for i in Row array
     4.1 if value if I greater than or equal to 2
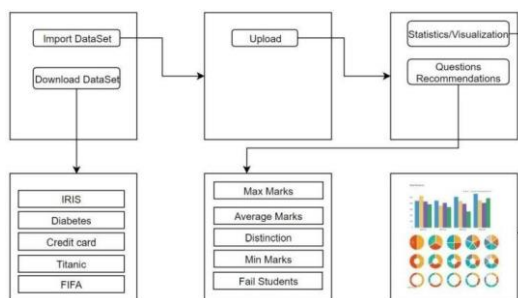   4.1.1 Extract value from Question Column from ith Row
End

Fig 1. Flow of Visualizer

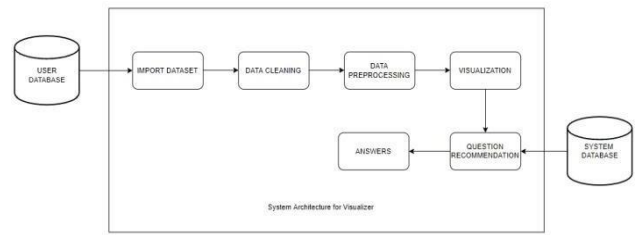## IV. SYSTEM DESING AND MODELLING

Fig 2. System Architecture of Visualizer

Import Dataset - Importing the dataset user wants to visualize and analyze.

Data Cleaning - Data cleaning involves the process of fixing or removing missing values, NaN values and incorrect values within a dataset.

Data Transformation - Data transformation involves the process of converting data from one format to another, typically replacing the string values by their numerical equivalent and then applying the Data Cleaning Algorithm again on it.

Data Visualization - The Model Visualizes Data related to Results. Our system would Visualize Data in the form of Bar Graphs, Pie Charts, Line Graphs, Histograms, ScatterPlot.

Questions Recommendation - If the Columns consisting the name of the selected attribute has a single attribute then the Questions that we have been recommending will give all the questions consisting that attribute.

V. RESULT

The dataset on which the system is currently working is 'student result' data. The data of 100 students is recorded and each student has 10 attributes like student_id, Name, Department and marks of five subjects like M1, EGR, Chemistry, Python, EEE and Total of the marks and percentage. The dataset contains numeric as well as string data.
The results in form of graphs and charts will be displayed on the webpage designed for results. In which, we are recommending question to user which are stored in different dataset file. Based on that questions user will get graphs as a answer to those questions.
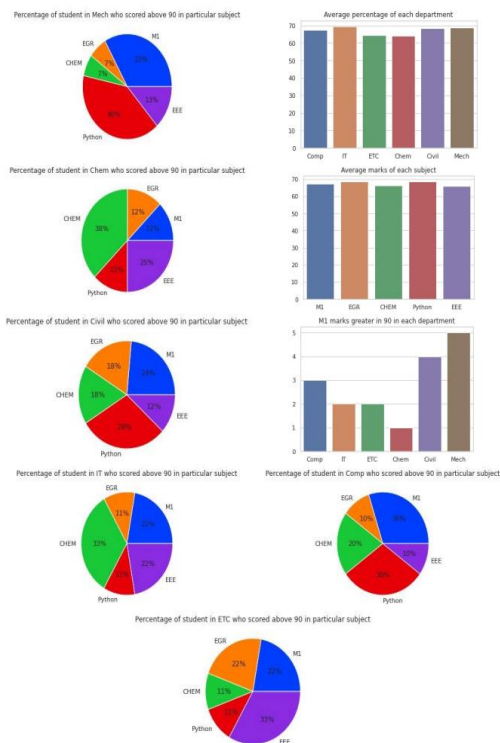
Fig 3. Results displayed with Graphs and Labels

VI. Discussion

The paper explains working of our prototype but the question arises how it differs from existing tools available in market? If we take an example of Tableau, we need to manually select measures and columns and then it visualizes the data. But what our tools does is, the user has to upload the dataset and will have to select what kind of output he/she wants and the graphs with the label is displayed on next page as shown above.

VII. Conclusion

This paper presents a brief overview of Data Visibility, Data Viewing involves presenting data in the form of an image or image that highlights the information easier to understand. It helps to point out the facts and decide what to do. It will also be an incentive for researchers, engineers, and students to consider the technological benefits that will benefit society. So, we completed our research successfully and started using VisuaLizer: Data Viewing and Analytics. The Proposed Model is an addition to intelligent systems that promote the use of smart technologies in day-to-day institutional processes.

VIII. Acknowledgment

References

[1] Xuedi Qin, Yuyu Luo, Nan Tang, and Guoliang Li.DEEPEYE: An Automatic Big Data Visualization Framework.IEEE.March,2018.

[2] Sadiku, Matthew & Shadare, Adebowale & Musa, Sarhan & Akujuobi, Cajetan & Perry,Roy. (2016). DATA VISUALIZATION. International Journal of Engineering Research and Advanced Technology (IJERAT). 12. 2454-6135.

[3] Toasa G, Renato Mauricio & Maximiano, Marisa & Reis, Catarina & Guevara, David.(2018). Data visualization techniques for real-time information — A custom and dynamic dashboard for analyzing surveys' results. 1-7. 10.23919/CISTI.2018.8398641.

[4] Andrienko, Gennady & Andrienko, Natalia & Drucker, Steven & Fekete, Jean-Daniel & Fisher, Danyel & Idreos, Stavros & Kraska, Tim & Li, Guoliang & Ma, Kwan-Liu & Mackinlay, Jock & Oulasvirta, Antti & Schreck, Tobias & Schumann, Heidrun & Stonebraker, Michael & Auber, David & Bikakis, Nikos & Chrysanthis, Panos & Sharaf,Mohamed & Papastefanatos, George. (2020). Big Data Visualization and Analytics:Future Research Challenges and Emerging Applications.

[5] Mouataz Zreika, Maria Estela Varua,Ana Hol,Jie Hua.Applying Data Visualization Methods on Australian Stock Investment Analytics. International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 6, Issue 1, January 2017

[6] Noor Nashriq Ramly, Fazli Mat Nor, Nurul Haszeli Ahmad, and Mohd Haris Aziz. Comparative Analysis on Data Visualization for Operation Dashboard. International Journal of Information and Education Technology, Vol. 2, No. 4, August 2012.

[7] Raineri P., Molinari F. (2021) Innovation in Data Visualization for Public Policy Making. In: Concilio G., Pucci P., Raes L., Mareels G. (eds) The Data Shake. SpringerBriefs in Applied Sciences and Technology. Springer, Cham.

[8] Daniela Ioana SANDU IBM Italy. Operational and real-time Business Intelligence. Revista Informatica Economică.

[9] C.Chen,W.Hardle,A.Unwin.Handbook of Data Visualization. 2008. Springer-Verlag Berlin Heidelberg.

[10] Michael L. Waskom.seaborn: statistical data visualization.The Journal of Open Source Software.2021,April. Azvine, B., Cui, Z. & Nauck, D.D. Towards real-time business intelligence. BT Technol J 23