

PREDICTING COURSES BASED ON TIME DURATION USING MAP-REDUCE

Bhavishya Puttagunta
Department of CSE
V R Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
bhavishyaputtagunta@gmail.com

Alaparathi Jitendra
Faculty at Department of CSE
V R Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
alparthijitendra@vrsiddhartha.ac.in

Charitha Sree Chanamolu
Department of CSE
V R Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
charithasreechanamolu@gmail.com

Vasundhara Devarapu
Department of CSE
V R Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
vasundharadevarapu20@gmail.com

Abstract -With the upcoming of new innovations, new developed patterns are brought into the Educational framework which brings about enormous information which is unordered and it is likewise, a test for understudies to allude to those scholastic courses which are useful in their modern preparation and expands their vocation possibilities. Another test will help to change over the unregulated information into organized and significant data there is a requirement for Data Mining Techniques. The appropriate file system is utilized to hold a large amount of information and also allows to reduce the over storage of unregulated data. Information extricated utilizing MapReduce is useful in decision making for understudies for deciding courses picked to undergo modern preparation. In this paper, with the help of MapReduce and Kmeans algorithms we sort the data into their respective time durations from which the users can choose the their courses based on the required time duration in which they are willing to complete the courses. Here, utilizing DFS, errands run over MapReduce, and the result is gotten after a conglomeration of results. Our implementation allows users to choose their subject of interest with the best possible time duration they require to complete the course.

Keywords—Data Mining Tools, Distributed File System, Map Reduce.

I. INTRODUCTION

Information is perhaps the most conspicuous region in current innovation for recovering significant data from immense measures of unstructured and appropriated information utilizing equal handling of information. There is a gigantic benefit to the Educational areas for following data mining techniques to analyze information collected from understudies, criticisms, most recent scholastic patterns, and so on which helps in giving quality schooling and dynamic methodology for understudies to build their vocation possibilities and the right determination of courses for modern preparation to satisfy the expertise hole relates between essential instruction and industry recruiting understudies. Information mining has an extraordinary effect in scholastic frameworks where instruction is weighed as the essential contribution for cultural advancement. Huge information is arising in the field of information mining. It is a term for datasets that are extremely enormous or difficult that conventional information handling the application's programming is difficult for managing. Huge information incorporates gathering information for investigation purposes which deal with activities like searching, sharing, representing information, query handling, refreshing, along with keeping up with the protection of data. In Big information, there is a very huge dataset that is dissected computationally to bring out a lot of examples, patterns, and affiliations. So it organises and maintains unstructured data, such as MS Office documents, PDFs, and texts, as well as structured data, such as social data. Hadoop is one method of enormous information and replies to issues connected with the treatment of unordered and huge information. Hadoop is widely available source used for distributed processing system for big data applications. They help in controlling data processing and storage. This approach can allow schools, establishments, universities to get a complete part of the understudies. It coordinates in addressing questions connected with the learning practices, better arrangement and educational plan patterns, and future course choices for understudies which assist with making charming growth opportunities for understudies. With the help of Map Reduce techniques the issue of big datasets can resolved easily. Hadoop clusters run over Map Reduce occupations by dividing the enormous information into little lumps along with handling the information by running in equal amounts on conveyed bunches.

II. LITERATURE SURVEY

Sonali Agarwal et al., [1] analyses the admission academic performance and, ultimately, placement, they used a Data Mining approach. It has been proven beyond a shadow of a doubt that their placement, which is an important component of the educational process, might be based on their performance in both the qualifying examination and the test. This is the unmistakable outcome of the investigation.

Woo, Jongwook et al., [2] states that study offers the temporal complexity of the Apriori-Mapreduce method is demonstrated, which theoretically shows that as the number of nodes increases, it achieves significantly better performance than the sequential algorithm. The item sets computed and output by the technique can be used to compute and output Association Rules for market analysis.

Katrina Sin et al., [4] studied that the principles of applying Data Mining techniques in Education and Prediction or assessment of Student Performance were recognised as the two significant themes. Because the previous trend included papers on performance prediction, we can deduce that "Performance Prediction Using Data Mining" is the main topic.

Manpreet Singh et al., [12] proposed methodology, which identifies them. It also enables the teacher to intervene before a student drops out or prepare for resource allocation with the assurance that knowing how many students are going to pass or fail will provide. The proposed system also graphically displays data according to the need or organisation, assisting them in making crucial decisions. We also use clustering for future work, as it allows us to see the domain and interests of students in a certain topic.

R. B., S. S. Thorat et al., [13] In this study, employed a data mining technique in a student database to forecast students' results using the K-means clustering algorithm. We expect that the data gathered as a result of the data mining technique will be beneficial to both instructors and students.

Shakeel, Khawar et al., [14] used classification algorithms to predict student performance using educational data. We discovered how useful educational data mining may be for institutional purposes. This research will aid students and administrators in identifying gaps and areas for improvement, as well as identifying poorer pupils and providing them with special materials to assist them better.

Kumar, Varun et al., [15] the research looked at how one data mining technique, association rule mining, could be used to improve the quality of students' Post-Graduation performances. The discovered association rules highlight a number of elements, including student engagement, curriculum design, teaching, and assessment procedures, that can have an impact on students who have not achieved sufficient performance at the post-secondary level.

Vikesh Kumar et al., [16] this project aims to employ Data Mining techniques to examine academic data from students and improve the quality of technical education. We used BayesNet, Nave Bayes, Multilayer Perceptron, IB1, Decision Table, and PART Classification methods to classify student data in this study.

III. METHODOLOGY

Each line is treated as an exchange, with each line carrying a mixture of factors or object sets. Following the use of the MAP reduce Hadoop Data Mining Technique, understudies' favourite example for modern instructional class mixes is expected. The information dataset received from understudies is shown and saved in the HDFS for Map reduction. The information is then divided into groups and sent to the mapper, who connects the data to the outcome. The mapper's output is addressed as a couple. The mapper's output is then combined in the combiner before being sent to the minimizer. Hadoop divides the errand into Map and Reduce Tasks in order to sort out the labour. The components of the Hadoop Distributed File System are discussed below. The Map-Reduce software converts a list of input data components into a list of output data components, and it is completed using Map and Reduce. The team uses Hadoop to execute apps and identify hub data as Web Interfaces. first sections, then planned, and finally rearranged. Following the reorganization, the unstructured data is sifted to obtain a result, which is also known as the "ReducePhase."

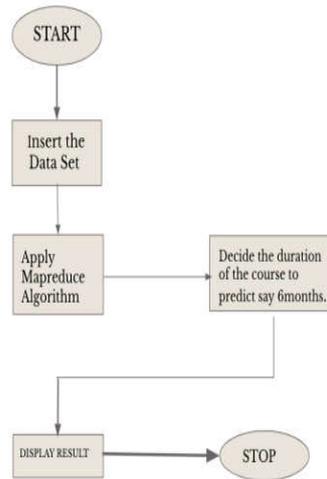


Fig 1. Snapshot of the Flow

IV. PROPOSED SYSTEM

In our proposed framework we are dissecting the courses that are being favored year-wise utilizing MapReduce and giving the outcomes. This gives the client data in regards to the course inclination for different years. This would uncover whether there is progress or relapse in the number of understudies inclining toward the course and the understudy could choose whether to decide on a course or not.

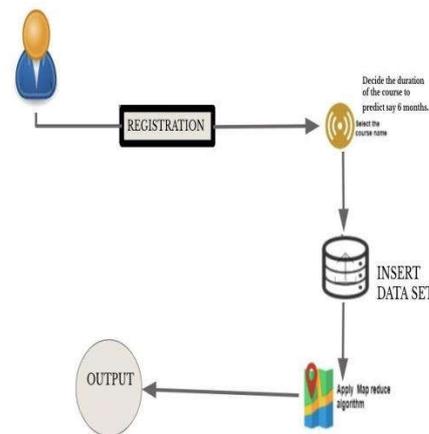


Fig 2. Snapshot of Proposed System

1. A. Advantages to the Proposed System

- Helps in giving quality instruction and dynamic methodology for understudies to build their professional possibilities
- For the right choice of courses for modern preparation to satisfy the expertise, the hole relates between essential instruction and industry recruiting understudies
- Keeps away from vagueness.
- Simple course finding for preparing.

B. Implementation

In the above screen select and transfer the 'data.txt' document and afterward click on the 'Open' button to stack the dataset and afterward select the ideal subject starting from the dropbox and afterward click on the 'Apply MapReduce Algorithm' button to apply MapReduce and to get course inclination %

```

d.3984 NSM CSBC KNCD - Voice Over 5G - Options for Deployment | GS02-2020KT-A08
d.4001 NSM CSBC KNCD - APJ Graneenphone Cloud Core Network Design | GS02-2020KT
-G58
d.4004 NSM CSBC KNCD - 3rd Party Onboarding | GS02-2020KT-G61
d.4007 NSM CSBC KNCD - Radsys MRF 15.0 training - Day 1 | GS02-2020KT-G63
d.4011 NSM CSBC KNCD - Learning Index Mastery - GS02-2020KT-LM01
d.4014 NSM CSBC KNCD - IMS services into the cloud | GS02-2020KT-PH002
d.4017 customer Service Confrontation and Conflict | GS02-2020KT-PH002
d.4019 Improving Customer Service in Technical Support| TBUS4039W_v1.0
d.4022 InfoView - Enterprise
d.4035 Automatically create budgetary private wireless offers to port customers
with InstaQuote (K+)
d.4038 Manufacturing & Logistics
d.4040 NSM CSBC KNCD L4 CFX-5000 Design Engineer | GS02-CFX-0400-F
d.4043 ESU xDA Curriculum 1.0 | GS00297-K-2005
d.4051 5G to NDC Solution (K+)
d.4054 Change Map Performance Drivers
d.4071 GS Deploy
d.4074 Deploy & CNBD Contractual "Cookbook" | GS05-100901-M-01
d.4077 GS Deploy Contract Template Test | GS00260-F-2002
d.4081 MN Backlog Management for FOTs | MN00893-M-0020
d.4084 CDFE Stack Deployer (CSFP 20 Deep Dive Session) | NS00301-K-2000
d.4087 RCP CI Gerrit code review tool training (WBT)| MB0248END08HZ7_SPL
d.4089 Jenkins (Skillsoft)| CUSD0915-01A
d.4092 Kafka Essential Training
d.4102 Learning Jira Software
d.4105 Digital Insights Suite introduction (NS00501)
d.4108 Ethical Leadership Through Giving Voice to Values
d.4110 NSM: HM Demand Planning Refresher (S18)
d.4113 FOT Leader Videos - The FOT Approach | MN00927-M-0620
d.4121 NSM CSBC KNCD Docker Associate Curriculum | GS02-DOCK-21000-K
d.4124 NSM CSBC KNCD Kubernetes Associate Curriculum | GS02-KUBE-21000-K
d.4141 NSM CSBC KNCD Kubernetes Foundation Curriculum | GS02-KUBE-G11000-K
d.4144 NSM CSBC KNCD DevOps Foundation Curriculum | GS02-DVOP-G11000-K
d.4147 Nokia Paris-Saclay Master Classes
d.4151 Building effective teams | SMAACOD-M-1904
d.4154 Managing Organizational Change for Managers
d.4157 Holding Yourself Accountable
d.4159 Accountability
d.4162 DevOps Introduction for NSM | IT00004-K-1811
d.4172 Peer and Frustration - Performance Driver
d.4175 Process Improvement - Quick Read
d.4178 Unlock Passion & Future
d.4180 NSM: (NAM) Quote Administration Process outsourced to HiPro (S19)
d.4183 Business and Leadership
d.4191 Building Peer Relationships
d.4194 Global Collaboration | LSHMM1-W-1806
d.4211 Using Strategic Thinking to Consider the Big Picture
d.4214 Influencing Key Decision Makers
d.4217 The Fruits of Integrity: Building Trust at Work
d.4221 Leveraging Emotional Intelligence

```

```

j.5977 Clone of Template for Training Solution
j.5979 Network Resiliency and Assurance
j.5982 IP Optimization
j.5992 Smart Plan Suite (SPS) R. 20.3
j.5995 EdenNet
j.5998 Home Device Manager
d.6000 HDM Operations
d.6003 Traffica
d.6011 Service Routing Troubleshooting
d.6014 CMSB200 Provisioning and Administration for 19 release
d.6031 5G RAN
d.6034 LTE for Technical Managers
d.6037 LTE & SRAN BTS Mediator
j.6041 IMPACT
j.6044 Netguard Certificate Manager
j.6047 Airscale Cloud BTS
j.6049 IMPACT Fundamentals
j.6052 WCDMA Delta Operation and Maintenance
j.6065 Passive Optical LAN Deployment
j.6068 FlexIMS
d.6070 FlexIMS 18.0
d.6073 EdenNet Operation
d.6081 Pre-Sales & Solution Engineers Learning Pathway: Technical Content
d.6084 NCA (Nokia Certified Associate) LTE Networks Fundamentals
d.6101 Introduction to Virtualisation and Telco Cloud
d.6104 MDM Technology Overview
d.6107 Public Safety Solutions
d.6111 Optical Self-Paced Learning
d.6114 Data Reflery 19.6
j.6117 NPP-P Troubleshooting
j.6122 CW MRE and SGSN (ILT)
j.6132 CBRS CPT Certification
j.6138 Nokia Bell Labs End-to-End 5G Certification Program
d.6140 ONAP Service Engineering & Maintenance
j.6143 Session Border Controller (SBC) for Carrier R. 20.0
d.6151 5G RAN Network Professional (NRNP)
d.6154 5G RAN Network Operation Expert (NRNE)
d.6171 IP & Optical Network Automation for Enterprises
d.6174 1830 PSS Introduction
d.6177 Introduction to E2E Cloud
d.6181 NSP Classic Management for SAEGW (S/P GW) OAMMP
d.6184 Cloud Signalling Director (CSD) 4G R. 20.2
d.6187 FlowOne Release 20 Solution Developer
d.6189 Digital Operations Center
j.6192 Introduction to Modular Private Wireless -
j.6202 Nokia MIFI
j.6205 Nokia Automated Analytics Solution (NAAS) for Access Control
j.6208 Introduction to 1830 Products
j.6210 NetAct 20 Cloud Administration
Time taken: 1.87 seconds, Fetched: 1293 row(s)
hive>

```

Fig 3. Snapshot of Implementation

In the above screen, MapReduce starts processing and then will give the below output.

V. OUTPUT SCREENSHOTS

The below output screenshots represent different types of graphs, say pie chart, Bar graph, Line graph, and Donut chart.

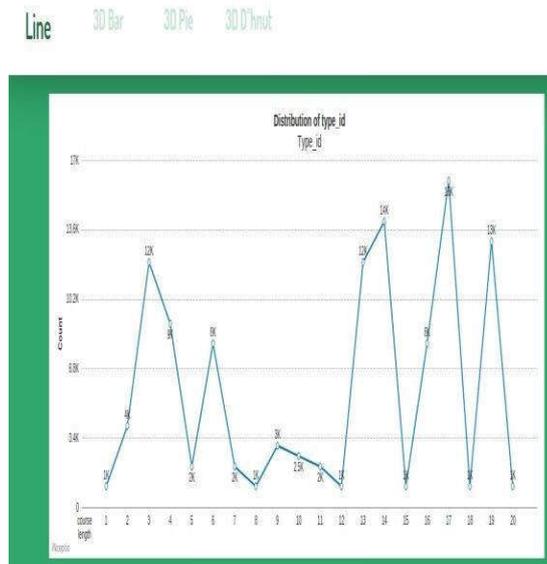


Fig 4. Snapshot of Line graph representing Distribution of type-id



Fig 5. Snapshot of Bar graph representing Distribution of type-id

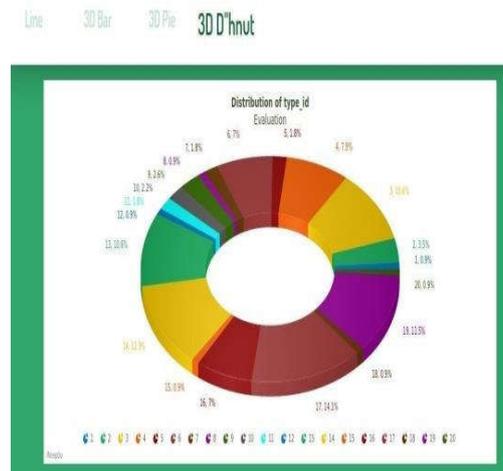


Fig 6. Snapshot of Donut chart representing Distribution of type-id

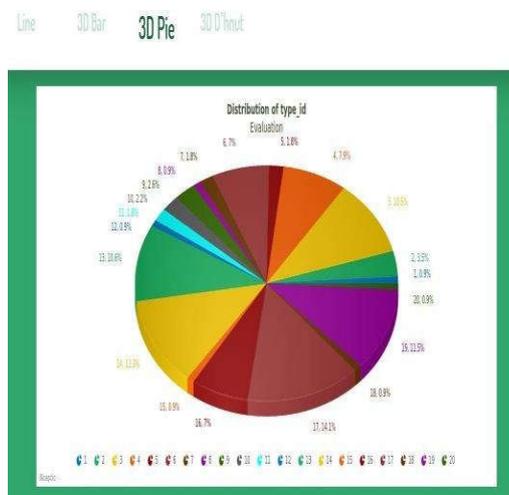


Fig 7. Snapshot of Pie chart representing Distribution of type-id

VI. CONCLUSION

Since the data created from various educational institutions are huge in size, it is difficult for a researcher to investigate the dataset and find the required course to take according to his time duration. So here comes the response of MapReduce that is utilized to handle huge data and with the guide of utilizing this, the researcher can inspect the data and find the suitable courses offered by the institution which is close to his/her required time period. The entered dataset is taken care of into the mapper and in the wake of going through the mix stage, the minimizer presents the result subsequent to arranging the tuples got from a mapper and are looking like pair. The researcher is even outfitted with graphical presentations for each course which makes the researcher research the data successfully and precisely. The information dataset is taken care of into the mapper and in the wake of going through the mix stage, the minimizer shows the result following the collection of tuples obtained from a mapper. The data demonstrates that students have chosen a variety of course combinations for modern preparation, and the information has become unstructured and erroneous for students to choose courses for preparation.

REFERENCES

- [1] Sonali Agarwal, G. N. Pandey, M. D. Tiwari, "Data Mining in Education: Data Classification and Decision Tree Approach", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2012.
- [2] Woo, Jongwook. "Apriori-map/reduce algorithm." Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012.
- [3] Yang,XinYue, Zhen Liu, and Yan Fu. "MapReduce is a programming model for association rules algorithm on Hadoop." *The 3rd International Conference on Information Sciences and Interaction Sciences*. IEEE, 2010.
- [4] Katrina Sin, Loganathan Muthu, "Application of Big Data in Education Data Mining and Learning Analytics – A Literature Review", ICTACT Journal on Soft Computing, ISSN: 2229-6956 (online), Vol5, Issue 4, July 2015.
- [5] Manjulatha, B., Ambica Venna, and K. Soumya. "Implementation of Hadoop operations for big data processing in educational institutions." *Int. J. Innov. Res. Comput. Commun. Eng* 4 (2016).
- [6] N.Tajunisha,M.Anjali,"PredictingStudentPerformanceUsingMapReduce",IJECS,Vol.4,Issue 1, Jan 2015, p. 9971-9976.
- [7] Mallikarjuna Reddy, V. Venkata Krishna, L. Sumalatha," Face Recognition based on Cross Diagonal Complete Motif Matrix", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.10, No.3, pp. 59-66, 2018.DOI: 10.5815/ijigsp.2018.03.07
- [8] Ayaluri MR, K. SR, Konda SR, Chidrala SR. 2021. Efficient steganalysis using convolutional autoencoder network to ensure original image quality.PeerJComputerScience7:e356<https://doi.org/10.7717/peerj-cs.356>
- [9] Shankar M.Patil, Praveen Kumar, "Data Mining Model for Effective Data Analysis of Higher Education Students UsingMapReduce", IJERMT, ISSN:2278-9359, Vol.6, Issue 4, April 2017.
- [10] Jeffrey Dean, Sanjay Ghemawat, " MapReduce: Simplified Data Processing on Large Clusters", Google, Inc, OSDI 2010.
- [11] Sin,Katrina, and Loganathan Muthu. "APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS--A LITERATURE REVIEW." *ICTACT Journal on soft computing* 5.4 (2015).
- [12] Bhullar, Manpreet Singh, and Amritpal Kaur. "Use of data mining in education sector." *Proceedings of the World Congress on Engineering and Computer Science*. Vol. 1. 2012.
- [13] Bhise, R. B., S. S. Thorat, and A. K. Supekar. "Importance of data mining in higher education system." *IOSR Journal Of Humanities And Social Science (IOSR-JHSS)* 6.6 (2013): 18-21.
- [14] Shakeel, Khawar, and N. Anwer Butt. "Educational data mining to reduce student dropout rate by using classification." *253rd OMICS international conference on big data analysis & data mining*. 2015.
- [15] Kumar, Varun, and Anupama Chadha. "Mining association rules in student's assessment data." *International Journal of Computer Science Issues (IJCSI)* 9.5 (2012): 211.
- [16] Singh, Samrat, and Vikesh Kumar. "Performance analysis of engineering students for recruitment using classification data mining techniques." *International Journal of Science, Engineering and Computer Technology* 3.2 (2013): 31.