

## LITERACY RATE ANALYTICAL STUDY ON CRISP-DM ALGORITHM USING PYTHON

<sup>1</sup> S. Amar Nadh, <sup>2</sup> Utiri Gopinath, <sup>3</sup> N Swathi, <sup>4</sup> Chintapalli Jayachandru  
<sup>1,2,3</sup> Assistant Professor, <sup>4</sup> Student, Dept. of Computer Science Engineering, Newton's Institute of Engineering, Macherla, Andhra Pradesh, India.

### Abstract

Education is the foremost important aspect for development of any country. Spread of literacy is largely connected with developments like urbanization, industrialization, trade and modernization. Better education and literacy enhances social and economic conditions of any country. Ministry of Human Resource Development releases a data on literacy rate each year which can be exceptionally valuable in examining different elements influencing education rate of a state or an area. Data to be analysed is handled and cleaned to draw out the most significant features. The data at that point analysed gives the last outcome which is presented, making it easy to understand.

### Keywords:

CRISPDM, Business understanding, Data understanding, Data preparation, Data modelling, Analysis, Feature extraction.

### INTRODUCTION

LITERACY is defined as the ability to read, write and think rationally. It represents the lifelong, intellectual process of gaining meaning from print. Key to all literacy is reading development, which involves a progression of skills that begins with the ability to understand spoken words and decode written words, and culminates in the deep understanding of text. Reading development involves a range of complex language underpinnings including awareness of speech sounds (phonology), spelling patterns (orthography), word meaning (semantics), grammar (syntax) and patterns of word formation (morphology), all of which provide a necessary platform for reading fluency and comprehension. Once these skills are acquired the reader can attain full language literacy, which includes the abilities to approach printed material with critical analysis, inference and synthesis; to write with accuracy and coherence; and to use information and insights from text as the basis for informed decisions and creative thought. Literacy has always been an issue for the world. Every country aims to achieve full literacy rate. Although literacy rate has increased up to a great extent now but still there is a need to know the areas that are still lagging behind. So, study and analysis of literacy data of the world is required to provide a timely and informed basis for helping planning and management of education services and to establish or contribute to an education system for collection, organization and utilization of education data.

Problems with traditional teaching Classroom environment which plays a crucial factor in motivating the children to learn is missing in many schools. Children face many challenges to engage in learning; the most intimidating of them all is not having a role to participate in the process. Children need to be encouraged to speak, to discuss, to express their opinions and to solve problems together. It has become very difficult to teach students lessons in many subjects esp. Science, Social studies and Technology due to the lack of learning tools (apart from the conventional classroom teaching). This leads to student disengagement and discontent which results in frequent absenteeism in schools. In the past, students had to adapt to the teaching styles of their teachers but now there is a need for the education system to modify its delivery process to suit the learning styles of the students.

There is a severe shortage of qualified teachers in many schools in India. Also, teacher absenteeism in schools is a major issue. On the teachers' front, as much as the other nonteaching duties come in the way of good preparation, lack of subject expertise and teaching resources and over-crowded classes are few of the reasons that limit the teachers to a delivery mode of lessons, rather than engaging children in partnership for learning. Lack of resources Many schools do not have adequate resources to invest in teaching and learning aids.

There is a clear evidence to suggest that achievement levels tend to decline as the children move along the educational hierarchy. This shows that schools are not able to cope with the teaching learning load as the pupils' progress through various grades. There is a need for transformation in the way people find, learn and consume educational content. Economic Divide The schools in urban schools are relatively in a better position when compared to most of the rural schools in terms of access to quality teaching, availability of teachers, access to quality education. But the situation of many low-income government and private schools in urban areas is almost similar to their counter parts in rural areas. This economic divide causes education inequity. This leads to disengagement of the neglected students, dropouts and lack of opportunities for higher studies. In the long term, this education inequity decides and affects the life paths of the students, unemployment and lack of means of living.

## 2. RELATED WORK

### PROBLEM STATEMENT

Education is the foremost important aspect for development of any country. Spread of literacy is largely connected with developments like urbanization, industrialization, trade and modernization. Better education and literacy enhance social and economic conditions of any country. Ministry of Human Resource Development (DISE) releases a data on literacy rate each year which can be exceptionally valuable in examining different elements influencing education rate of a state or an area. Data to be analyzed is handled and cleaned to draw out the most significant features. The data at that point analyzed gives the last outcome which is presented, making it easy to understand.

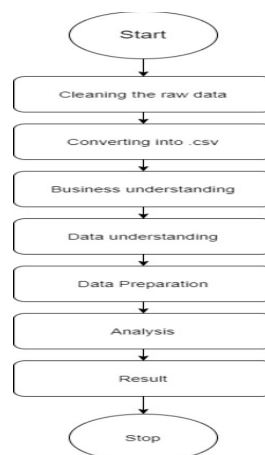


Fig1.1. Flow Diagram of Proposed System

Problems with traditional teaching Classroom environment which plays a crucial factor in motivating the children to learn is missing in many schools. Children face many challenges to engage in learning; the most intimidating of them all is not having a role to participate in the process. Children need to be encouraged to speak, to discuss, to express their opinions and to solve problems together. It has become very difficult to teach students lessons in many subjects esp. Science, Social studies and Technology due to the lack of learning tools

(apart from the conventional classroom teaching). This leads to student disengagement and discontent which results in frequent absenteeism in schools. In the past, students had to adapt to the teaching styles of their teachers but now there is a need for the education system to modify its delivery process to suit the learning styles of the students.

There is a severe shortage of qualified teachers in many schools in India. Also, teacher absenteeism in schools is a major issue. On the teachers' front, as much as the other nonteaching duties come in the way of good preparation, lack of subject expertise and teaching resources and over-crowded classes are few of the reasons that limit the teachers to a delivery mode of lessons, rather than engaging children in partnership for learning.

Lack of resources Many schools do not have adequate resources to invest in teaching and learning aids. There is a clear evidence to suggest that achievement levels tend to decline as the children move along the educational hierarchy. This shows that schools are not able to cope with the teaching learning load as the pupils' progress through various grades. There is a need for transformation in the way people find, learn and consume educational content.

Economic Divide The schools in urban schools are relatively in a better position when compared to most of the rural schools in terms of access to quality teaching, availability of teachers, access to quality education. But the situation of many low-income government and private schools in urban areas is almost similar to their counter parts in rural areas. This economic divide causes education inequity. This leads to disengagement of the neglected students, dropouts and lack of opportunities for higher studies. In the long term, this education inequity decides and affects the lifepaths of the students, unemployment and lack of means of living.

Issues in the usage of technology in general has transformed our larger society. It has become central to people's daily lives. And yet technology has been kept in the periphery of schools, used for the most part only in specialized courses( in a few cases esp urban schools). Technology in form of digital content delivered through devices such as projectors, computers, DVD players, handheld tablets can help the students to get quality content and uniform instruction without much dependence on parameters such as availability of qualified teachers and location of schools. Students and teachers must be encouraged use technology. For this, the teachers must be trained to use technology effectively and integrate it in their classroom instruction. However, the frequent power outages in schools or complete lack of access to grid supply of electricity in schools esp. in remote rural areas makes it very challenging to implement technology solutions in classrooms. This is the case also for schools that can afford to invest in e-learning equipment in urban areas. The table below shows the state wise percentage of schools which have electricity connection. However, the power supply is very erratic and unreliable.

#### **FEATURES OF EXISTING SYSTEM**

1. Literacy rate of each year which can be exceptionally valuable in examining different elements is processed and noise is removed. Data mining is used which takes a lot of time to obtain results.
2. Raw data is being cleaned and unwanted and noisy data is removed.

3. Data is pre-processed and transformed into required format.
4. Large data sets are processed linearly as a whole unit.

#### LIMITATIONS OF EXSITING SYSTEM

1. High cost hardware is required for storing and processing of such huge data.
2. It consumes lot of time to give results.
3. It didn't show any factors which when improved leads to improve in literacy rate.

#### SOLUTION TO THE PROBLEM

##### PROPOSEDSYSTEM

Proposed system will be following CRISP-DM and will be marking my steps throughout the notebook. The steps for CRISPDM are as follows:

1. Business understanding: outlining the questions we will answer along with why they are relevant and important
2. Data understanding: understanding the data and providing stats about it
3. Data preparation: showing how we clean and prepare the data
4. Data modelling: if we created a model, this is the section we would put it in of the analysis
5. Results evaluation: conclusion and discussion

We are making use of python programming language for instant conclusions and analysis. It is easy to obtain graphs in python as there are libraries predefined which can be used to. NumPy allows developers to do the operations like Shape manipulation and Fourier transformation, logical as well as mathematical operations, linear algebra operations using inbuilt functions. NumPy contains a multi-dimensional array and matrix data structures.

#### ADVANTAGES OF PROPOSED SYSTEM

In our proposed system we are performing analysis using python. In all the other base papers and researches, they are not concentrated on the factors why a specific state is having less literacy rate and how to improve the literacy rate. Our proposed system works on finding the trends in all the factors and how it is related to literacy. Advantages of system are as follows

- Python has many libraries for data science, this makes it as a best choice.
- The graphs and outputs are readable and easy to draw conclusions
- High cost hardware and software is not needed, all are open source and freely available
- It takes less time to produce outputs.

#### CONCLUSION

We started out with a comparison of the literacy rates in the different states and union territories in India. We compared the Male and Female literacy rates in these states and saw that there was a huge difference in the states with a low overall literacy rate. The top 3 states were Kerala, Lakshadweep and Mizoram and the bottom 3 were Bihar, Telngana and Arunachal Pradesh. Next we saw that difference in male and female literacy rates, rural population proportion and dropout rates from 8th to 9th class played a huge role in separating the top 3 and bottom 3 states.

Finally, the dropout rates in different classes were explored and while the dropout rate for 6th class was really high, more students had enrolled in class 4 than had dropped out.

The literacy rate in India has been improving but there are some key issues that need to be tackled aggressively in order to improve the state of education in India. This is not just the job of the govt, but it is the duty of each and everyone living in the country. Hope to see the stats showing a much better India in the future.

**REFERENCES**

1. <https://jupyterlab.readthedocs.io/en/latest/>
2. Pratham, Annual Status of Education Report (Rural)2011, <http://www.pratham.org/M-20-3-ASER.aspx>
3. Matthew N. Anyanwu, Sajjan G. Shiva, “Comparative Analysis of Serial Decision Tree Classification Algorithms”, International Journal of Computer Science and Security (IJCSS), Volume 3, Issue 3, June 2009, ISSN (Online): 1985- 1553
4. Wikipedia, “List of countries by literacy rate”, [http://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_literacy\\_rate.html](http://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate.html)
5. Wikipedia, “Literacy”, <http://wikipedia.org/wiki/literacy>
6. <https://www.kaggle.com/datasets>
7. <https://www.ijser.org/researchpaper/Literacy-Rate-Analysis.pdf>
8. <https://www.papermasters.com/standards-literacy.html>
9. <http://documents.worldbank.org/curated/en/313651468741665795/What-factors-influence-world-literacy-is-Africa-different>
10. <http://documents.worldbank.org/curated/en/313651468741665795/pdf/wps3496.pdf>