

**SCHOOL OF COMPUTER SCIENCE AND APPLICATIONS**

A Project Synopsis

on

DYSLEXIA PREDICTION USING VARIOUS MACHINE LEARNING ALGORITHMS IN CLOUD

Bachelor of Science (Honors) in Computer Science -

Cloud Computing and Big Data

Submitted by

Keerthivasan R.V

R22DB044

Nathish R

R22DB049

Under the guidance of

Internal Guide

Mrs. Jesla Joseph

May 2024

Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru-560064

[www.reva.edu.in](http://www.reva.edu.in)

1. Title of the project
2. Introduction
3. Abstract
4. Existing and Proposed System
5. Advantages of Proposed systems
6. Module Description
7. Architecture Diagram
8. Result

**Introduction**

In the Medical Industry, Reading Disability (Dyslexia) identification is a major problem to detect, because there are multiple reasons which causes this reading disability to the people.

The reading disability leads to lack of reading words, color co-ordination complexity, mismatching and assumptions of words. Dyslexia is also associated with the human eyes (vision power), majority of the people with low vision power are affected mainly and majority of the school Students from the age group of 5-12 are affected by reading disability, which can lead to lack of interest in studies and unhealthy lifestyle among the people.

In this digital era with the help of Machine Learning, predicting the accuracy of Dyslexia in the people has become possible. Identifying the causes and symptoms which lead to dyslexia helps medical professionals and people to predict and provide the necessary medications and treatments in the earlier stage and helps them to lead a better life.

Various algorithms are used in machine learning to fetch accuracy to predict the person affected by reading disability. Data are collected in different forms like structured, unstructured and semi structured data, these data can be used for training and testing the data. To process, analyze, preprocess and get the output from the collected data and there are different types of algorithms applied to these data, while using these algorithms there are time complexity and space complexity involved, which involves the time and memory that is taken by the algorithm to produce the result in a faster way and store the result.

There is various research under process to get a higher accuracy by reducing the time and space complexity with the various algorithms. But there is still a lack of best results among the different algorithms and data.

By opting the Decision Tree and KNN algorithms to the different types of data, the highest accuracy can be achieved for prediction. To store the data which is collected will be stored in the cloud to reduce the physical storage space and to maximize the availability of the resources.

**Abstract**

Reading disability (known as Dyslexia) is a common problem faced by many children and younger people around the globe. There are various ways used to predict Dyslexia symptoms and habits with the help of using Machine Learning algorithms and Artificial Intelligence. The data generated to predict the Dyslexia problem from various people are stored using Cloud storage services. By applying various Machine Learning algorithms, the accuracy rate varies, determining and applying the best algorithm is one of the key aspects in analyzing and predicting the result with less time and space complexity.

By using Machine Learning with cloud computing the results can be generated in a faster way with less physical resources. Security is also improved by using the Cloud compared to the on-premises infrastructure for critical and sensitive data in the medical industry. The best practices are done for getting high-quality results from the various types of data with different functionalities.

**Existing System and Proposed System**

In the Existing System there were different ways data collected and processed for getting the highest accuracy. The eye co-ordination of the people is gathered and used for research purposes. The researchers have used Python as a major software for processing and presenting the graphical representation of result. There are google forms created for gathering data.

There were certain points which are used for predicting the Dyslexia. Ex: if the eye co-ordination of the people less than 0.85% is considered as the first level of Dyslexia and the points more than 0.85% are considered as non-Dyslexia.

In the Existing System the major algorithms used are

|  |
| --- |
| Random Forest |
| ID3 |
| Decision Tree |
| logistic regression |
| SVM |
| KNN |

There are Time and Space complexity involved in the existing system.

The data which has collected and analysed did not provide the better results and some data were not useful and there were lot of noisy data were also presented while preprocessing the data.

**Proposed System**

In the Proposed system, there is a website that will be created to gather the data from the people. The real-time data is gathered from the people.

For each text there will be timer given for calculating the speed that is taken for the user to predict the correct words and images.

The algorithms used for getting the prediction result are:

KNN (K-nearest neighbor)

Decision Tree

Every data which is entered by the user will be stored in the AWS RDS as rows and columns and those data will be changed to csv format, to predict the eye-coordination of the people used and to get higher accuracy with less time and space complexity.

The Dyslexia symptoms vary depending on the age group.

|  |
| --- |
| 5-10 |
| 11-15 |
| 16-25 |
| 25 and above |

The format of the data will be given according to the age, when a particular age is selected the data will appear to fill in the box with the timer.

**Advantages of the Proposed System:**

**Easy Data Gathering** – The data can be collected in an efficient way from the people by deploying the website in the search engine.

**Faster Processing** – There will be less time taken to process the data and find the result compared to the previous methods.

**Better Accuracy** – By using the effective Machine Learning algorithm better accuracy can be achieved.

**Less Time Complexity** – The time taken to analyze the n numbers of data will be reduced by micro or nano seconds.

**Less Space Complexity** – The space for storing the results will be in the cloud, there is no need for any physical storage to store the data.

**Security** – The security of the data will be taken care by the cloud service provider and regular backups and replication of the data will be done to eliminate the data loss.

**Module Description**

The purpose of the project is to find the best accuracy by using Machine Learning with less time and space complexity so that the medical industry can identify the Dyslexia issue from the children, young and elder people in early stages and provide the right medication and treatment to lead their life better.

**Data Collection** – Data is collected from the website which will be created and given to the people for use. The format of the data will be in various formats like text, image and color. The data for different categories of people will be taken with various formats.

**Data Preprocessing** – Preprocessing refers to the cleaning or removing the irrelevant data such as the noise, color, etc.., form the gathered data

**Data Transformation** – Data Transformation refers to the converting the various forms of data to a particular format for further analysis of the data. By doing this process, there will be a huge amount of time that can be saved.

**Testing the Data with the previous data** – Comparing the test data with the real-time data to find the similarity helps in prediction easier and combining both the data to find the speed and accuracy of the data.

**Predicting accuracy**- By using the Machine Learning algorithms the accuracy can be found in terms of percentages (%).

**Representation of Result** – In the statistical format like the Bar chart, Pie chart or Line graph the results will be presented for better understanding of the results.

**ARCHITECTURE DIAGRAM**

Accuracy

90% and above

Combining Test and processed Data

Real Time Data

Preprocessing of the data

**RESULT**

Prediction Accuracy with more than 90%.

The Less time complexity with less than 5 seconds.

Space complexity with less than 1gb of space for storing the data physically.

Graphical method to visualize the result.