

**SCHOOL OF COMPUTER SCIENCE AND APPLICATIONS**

A Project Synopsis

on

DYSLEXIA PREDICTION USING VARIOUS MACHINE LEARNING ALGORITHMS IN CLOUD

Bachelor of Science (CC & BD) in Computer Science -

Cloud Computing and Big Data

Submitted by

Keerthivasan. R.V

Nathish. R

R22DB044

R22DB049

Under the guidance of

Internal Guide

Prof/Dr. Jesla Joseph

Dec 2024

Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru-560064

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

1. Title of the project
2. Existing System and Proposed System
3. Advantages of Proposed System
4. Methodology Used
5. Architecture Diagram
6. Class Diagram
7. Testing Strategy
8. Implementation
9. Result
10. Future Enhancement

**Existing System**

In the Existing System the eye co-ordination data has been collected from the users in real-time by using the Multiple Choice Questions pattern, the user has to select their native language and other languages correctly, by selecting their native language correctly in the first attempt the user will be rewarded with 1 point and for each wrong options with time taken the points will be taken from the point 0 to 1.

The author used the classification algorithm to get the higher results with best accuracy rate, but the accuracy predicted was from the range of 85%-90% and after using the deep-learning algorithm also the accuracy got was 92% with high time complexity.

The Eye-Coordination data has been collected and no accuracy has been predicted from the data collected.

**Proposed System**

By using the Proposed System, the best accuracy rate has been identified by using the Classification algorithm that has the capability of adapting to various sizes of data for prediction provided by the user in a Tabular format with less time-complexity and space-complexity.

The data collected from the user has been processed by various Machine learning algorithms to find error, time, space and accuracy rate by fitting the model into test and training data. In the various algorithms the training data and testing data have been segregated in the 80% and 20% ratio.

Once the model has been trained with 80% data the remaining 20% data is given to the model for getting accurate results from the tested model. By using this type of testing strategy there will be no confusion in predicting the accuracy of the data.

The Data with the Reading Disability has been found and the

**Advantages of Proposed System**

The Time-Complexity has been reduced.

There is Less Space-complexity that has been established.

The data training process is faster than any other algorithm.

Efficiently handling large datasets for processing.

Reduces the overfitting of the data.

The Scalability of the data can be adopted in the proposed system.

Missing values will be handled by replacing a value to it.

Faster execution of the code.

The versatility of various datasets can be adopted.

**Methodology Used**

The dataset has been studied before implementation of any algorithm with the collected data published by the user. The dataset consists of Labelled data so, Classification and Deep Learning algorithms have been used for predicting the accuracy of the data. There were various Machine Learning algorithms like Pandas, NumPy, SciKit Learn and TensorFlow packages for predicting data with Reading Disability (Dyslexia).

**Requirement Analysis** – The requirement of Analyzing what type Dyslexia that must be predicted.

**Data Collection** – The Comma Separated Value (csv) files that contains the data collected from the user are gather from various sources for analyzing the data.

**Classification and Deep-Learning algorithms used are described below:**

**KNN Algorithm** – With the help of K-Nearest Neighbor algorithm, the prediction result was 100% when the data was split into 80% for training the model and 20% for testing the data but, there was a problem with the algorithm for processing the large volume of data and was not able to predict the data with Reading Disability.

**Decision-Tree Algorithm** – By using this algorithm the prediction result found was 85% when the testing data and training data were divided into 50% each, and when the data were split at 80% and 20% respectively the result got was 86.15%. Even though the results were satisfying, the time taken by the algorithm was high and the algorithm did not give a more accurate rate.

**RNN Algorithm** – The Recurrent Neural Networks (RNN) algorithm is a Deep-Learning algorithm used for large sets of data. The RNN algorithm uses epochs for repeatedly testing the data.

XGBoost Algorithm – The XGBoost algorithm uses built-in tools