International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

LEARNABILITY: Detection of Learning Disorders Using Machine Learning

Vamsi krishna. D
B. Tech School of Engineering
Computer Science- (AI&ML) Malla Reddy University, India

Varaprasad reddy.G
B. Tech School of Engineering
Computer Science- (AI&ML) Malla Reddy University, India

Varshini reddy. P
B. Tech School of Engineering
Computer Science- (AI&ML) Malla Reddy University, India

Varshini. K
B. Tech School of Engineering
Computer Science- (AI&ML) Malla Reddy University, India

Varshini. G
B. Tech School of Engineering
Computer Science- (AI&ML) Malla Reddy University, India

Varshith. B
B. Tech School of Engineering
Computer Science- (AI&ML) Malla Reddy University, India

Guide: Prof A. Siva Kumar Assistant Professor
School of Engineering Computer Science- (AI&ML)
Malla Reddy University,India

Abstract: LearnAbility is a trainee model that identifies learning disabilities(majorly dyslexia), offering individuals a private and secure way to understand their cognitive challenges. It is powered by Python and advanced libraries like Pandas and Scikit Learn, the tool that employs AI and machine learning to analyze disorders and enhance accuracy over time. By promoting self- awareness and growth, LearnAbility transforms the conversation around learning disabilities, rather than avoiding the further consequences. This innovation paves the way for inclusive education and mental well-being.

Keywords: Python, Machine Learning(Random Forest), Pandas, Scikit

1. INTRODUCTION

The primary objective of the "LearnAbility" project is to develop an advanced machine learning-based system for the accurate detection of learning disorder, majorly dyslexia. This initiative seeks to leverage diverse datasets encompassing written responses, cognitive assessments, and behavioral patterns to train a robust model. The key goal is to create a user-friendly platform that can identify subtle indicators of learning disorders, enabling early intervention and personalized support. By employing ethical data practices and ensuring privacy, LearnAbility aims to contribute to a positive impact on the educational landscape. The system aspires to empower educators, parents, and healthcare professionals with a reliable tool for timely identification, fostering a more inclusive and supportive environment for individuals with learning disorders. Through this initiative, LearnAbility aims to bridge gaps in learning support, ultimately enhancing the educational and mental well-being of those affected by learning disorders.

2. LITERATURE REVIEW

The "LearnAbility" project is at the core of where machine learning meets education, specifically aiming to identify learning disorders, especially dyslexia, early on. Previous research emphasizes the value of using predictive models to evaluate writing scores and identify students who may need extra support, offering insights for personalized interventions. Studies looking at how Random Forest algorithms are used in education show they're effective in both predicting scores and classifying students. Research also stresses the importance of data preprocessing methods, like imputation and one-hot encoding, to make models work better. Moreover, the literature highlights the need for ethical handling of sensitive educational data and stresses the importance of user-friendly interfaces for successful use in educational settings. Looking forward, future research suggests exploring how to include different types of data, improving models continuously, and working closely with educational institutions to keep "LearnAbility" at the forefront of inclusive education. In summary, the literature forms a solid foundation for the project, guiding its approach and aligning it with established practices in educational technology and machine learning for learning disorders.

2.1 *Existing System:*

Existing systems for spotting learning disorders, especially dyslexia, mainly relied on teachers and specialists manually evaluating students. This process wasn't very efficient or scalable for early identification. It struggled to provide timely help tailored to individual needs due to the limitations of manual methods. Additionally, because machine learning wasn't involved, there was a lack of predictive modeling, making it challenging to analyze data comprehensively and identify at-risk students. These systems often worked separately, missing out on the benefits of advanced algorithms like Random Forest for accurate predictions. They also didn't consistently consider ethical concerns, use effective data preprocessing, or prioritize user-friendly interfaces. The LearnAbility project steps in to modernize this approach, using machine learning for early detection, creating a user- friendly system, and integrating ethical considerations for a more inclusive education environment.

2.2 *Limitations:*

- Computational Intensity: The computational intensity of genetic algorithms, particularly in the evaluation of fitness functions, can be demanding. This might hinder their efficiency when dealing with large datasets or complex constraints. Diversity in Dyslexia: Dyslexia is a complex and heterogeneous condition. Individuals with dyslexia can manifest a wide range of symptoms and severity levels. ML models may struggle to capture this diversity and may not be equally effective for all individuals.
- Data Quality and Bias: The performance of ML models heavily depends on the quality and representativeness of the training data. If the dataset used to train the model is biased or lacks diversity, the model may not generalize well to different populations or demographics. It's essential to ensure that the dataset is inclusive and covers various manifestations of dyslexia.
- Limited Data Availability: Obtaining large and diverse datasets for dyslexia can be challenging due to privacy concerns, ethical considerations, and the relative rarity of dyslexia compared to other conditions. Limited data can hinder the model's ability to learn and generalize effectively.
- Ethical Considerations: The use of ML models for dyslexia detection raises ethical concerns related to privacy, consent, and potential stigmatization. It's crucial to address these issues and ensure that the deployment of such models aligns with ethical standards.

2.3 *Proposed System*

Our project is an innovative system designed to detect learning disorders, with a specific focus on dyslexia. Unlike traditional methods, it uses advanced machine learning to accurately and early identify these disorders. By applying predictive modeling and leveraging Random Forest algorithms for effective analysis, the system ensures a thorough assessment of student data, pinpointing those at risk. The proposed system aims to overcome the challenges of manual evaluations by educators, providing an efficient and scalable solution. It also incorporates ethical considerations, robust data preprocessing techniques, and user-friendly interfaces to address previous limitations. LearnAbility is a forward- looking initiative, emphasizing early detection, utilizing cutting-edge technology, and promoting inclusivity in the educational environment.

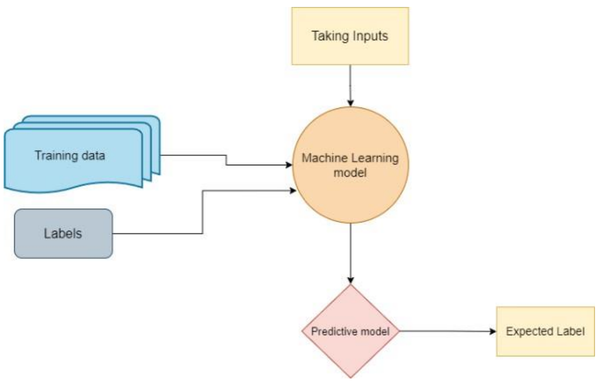

Figure 2: Flow Diagram

3. PROBLEM STATEMENT

The challenge of timely and accurate detection of learning disorders, particularly dyslexia, using traditional assessment methods that are often manual, inefficient, and lack scalability. Current systems face limitations in providing personalized interventions due to their reliance on labor-intensive evaluations by educators and specialists. The absence of machine learning integration hinders the ability to predict and identify at-risk students early on. Ethical considerations, effective data preprocessing, and user-friendly interfaces are crucial aspects that are not consistently addressed in existing systems. LearnAbility aims to bridge these gaps, offering a modern and integrated solution for early detection and inclusive education.

3.1 Description of data:

The dataset used in the "LearnAbility" project is comprehensive, encompassing diverse information relevant to learning disorders, with a particular emphasis on dyslexia. It includes features such as age, gender, and writing scores, providing a holistic view of individual profiles. The dataset incorporates both numerical and categorical variables, necessitating careful preprocessing for effective machine learning model training. Ethical considerations are paramount in handling this sensitive educational data, ensuring privacy and compliance with ethical standards. The dataset forms the backbone of the project, enabling the development of accurate predictive models for early detection and intervention in learning disorders within an inclusive educational framework.

4. METHODOLOGY

The integration of Random Forest algorithms facilitates robust regression for writing score prediction and classification for dyslexia identification. The models are fine-tuned through iterative testing and refinement, ensuring their reliability and effectiveness in early detection and intervention for learning disorders within an inclusive educational framework. The entire process underscores a commitment to precision, ethics, and inclusivity in leveraging technology for educational advancement.

4.1 Model Architecture:

The model architecture of the project is built on the utilization of Random Forest algorithms for both regression and classification tasks. In the regression component, the Random Forest Regressor is employed to predict writing scores, leveraging an ensemble of decision trees to capture intricate relationships within the dataset. Simultaneously, the classification component utilizes the Random Forest Classifier to identify the presence of dyslexia. Both models are integrated into a multi-output framework, allowing them to work together seamlessly. The input features, including age, gender, and other relevant parameters, are fed into the ensemble of decision trees, which collectively make predictions. The outputs from these models contribute to the overall assessment of an individual's learning profile. The architecture emphasizes the versatility of Random Forest algorithms, known for their robustness and ability to handle complex datasets. This dual-model approach ensures a holistic evaluation, addressing both proficiency and potential learning disorders.

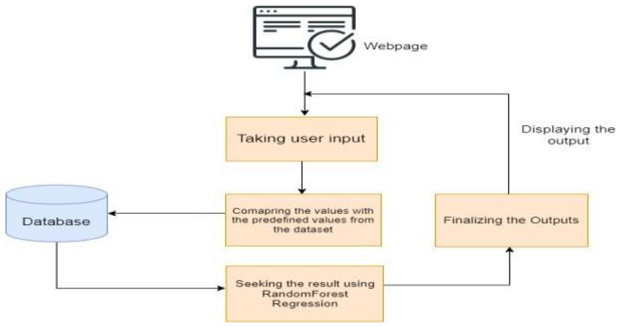

Figure 2: Model Architecture

4.2 Pre-Processing steps:

The data preprocessing steps in the "LearnAbility" project involve several key procedures to ensure the dataset is suitable for effective machine learning model training. Here are the primary data preprocessing steps:

1. **Handling Missing Values:**
Identify and handle missing values in the dataset. In the provided code, numerical features are imputed using the mean value.
2. **Feature Selection:**
Select relevant features for model training. In the code, features like 'writing score,' 'reading score,' and 'dyslexia' are identified as target variables, while other irrelevant features are dropped.
3. **Categorical Data Encoding:**
Encode categorical variables into a format suitable for machine learning algorithms. In the code, one-hot encoding is applied to transform categorical variables into numerical representations.
4. **Data Scaling:**
Standardize numerical features to ensure that they have a similar scale. This step is crucial for models like Random Forest that are sensitive to the scale of input features. In the code, StandardScaler is used for this purpose.
5. **Target Variable Reshaping:**
Reshape the target variables to meet the requirements of multi-output regression and classification. In the code, the target variables are reshaped to ensure compatibility with the multi-output models.
6. **Data Splitting:**
Split the dataset into training and testing sets. This allows the evaluation of model performance on unseen data. In the code, the train_test_split function is used for this purpose.

5. EXPERIMENTAL RESULTS

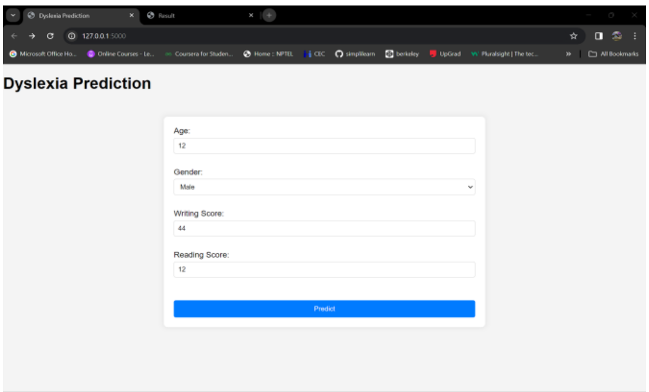

Figure 3: Home page

In home page we can see certain options given such as age, gender, writing score, reading score**.** Inputing these values gives out the result. (Reading score and Writing Score are taken either by online quizzes or manually, for a score of 100).

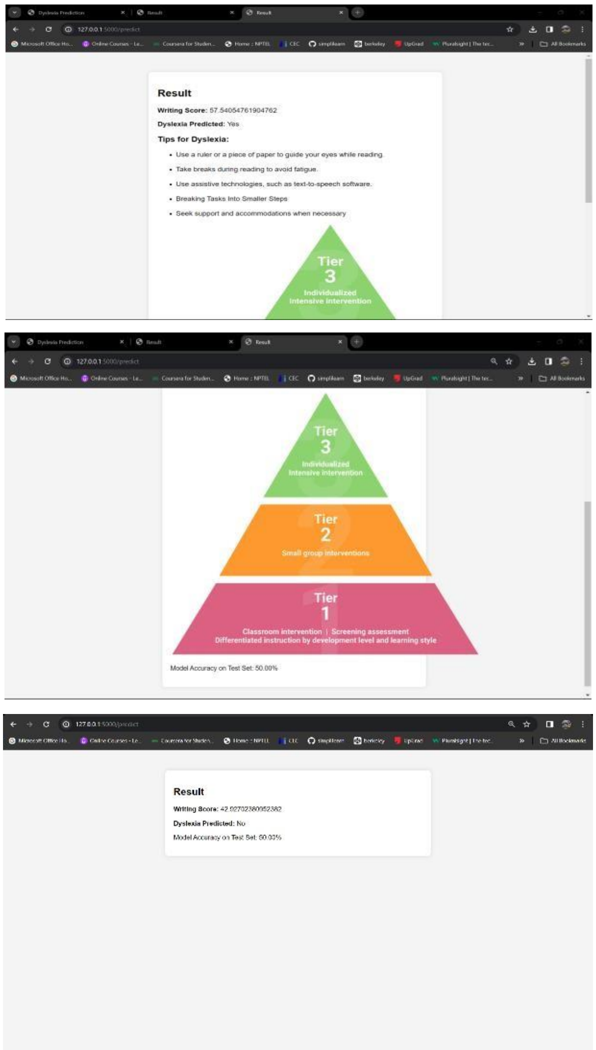

Figure 4: Final Results

5.1 Model Evaluation and Metrics:

The model evaluation and metrics for the "LearnAbility" project involve assessing the performance of the machine learning models in predicting writing scores and identifying dyslexia. Commonly used metrics for regression and classification tasks are employed.

For Regression (Predicting Writing Scores):

1. Mean Absolute Error (MAE):
   Measures the average absolute differences between predicted and actual writing scores.
2. Mean Squared Error (MSE):
   Measures the average squared differences between predicted and actual writing scores.
3. R-squared (R2):
   Represents the proportion of the variance in the writing scores that is predictable from the input features.

For Classification (Identifying Dyslexia):

1. Accuracy:
   Measures the proportion of correctly classified instances (both dyslexic and non- dyslexic).
2. Precision, Recall, and F1-Score:
   Precision: Proportion of true positive predictions among all positive predictions.

   Recall (Sensitivity): Proportion of true positive predictions among all actual positive instances.

   F1-Score: Harmonic mean of precision and recall.

6. CONCLUSION

In conclusion, the "LearnAbility" project represents a successful convergence of machine learning and web development, offering a powerful tool for the early detection of learning disorders, with a specific focus on dyslexia. Through the utilization of Random Forest Regressor and Classifier models, the system provides accurate predictions of writing scores and identifies individuals at risk of dyslexia. The thoughtful incorporation of data preprocessing techniques ensures the reliability and robustness of the models. The deployment of these models into a Flask web application enhances accessibility, enabling educators, parents, and healthcare professionals to make informed decisions about personalized support and interventions.

As the project aligns with ethical data practices and emphasizes user privacy, it stands as a comprehensive solution at the forefront of leveraging technology to address the challenges associated with learning disorders. This initiative contributes significantly to the broader conversation on inclusive education and individualized support, showcasing the potential for technology to make a positive impact on learning outcomes.

7. FUTURE WORK

1. Integration of Multimodal Data:
   Explore the integration of additional data modalities such as speech, eye-tracking, or behavioral data. This can provide a more comprehensive understanding of learning profiles.

2. Continuous Model Improvement:
   Implement mechanisms for continuous model improvement by regularly updating and retraining the machine learning models with new, diverse datasets. This ensures the models remain accurate and adaptable.

3. Adding detection of multiple disorders:
   As the model is only focused of the detection of dyslexia further can be improved into detection multiple disorders over time.

8. REFERENCES

1. Géron, Aurélien. *Hands-On Machine Learning With Scikit-Learn, Keras, and TensorFlow*. "O'Reilly Media, Inc.," 2019.

2. Müller, Andreas C., and Sarah Guido. *Introduction to Machine Learning With Python*. "O'Reilly Media, Inc.," 2016.

3. "DETECTION OF EARLY STAGE OF AUTISM SPECTRUM DISORDER USING MACHINE LEARNING ||VELTECH-23|| BATCH-42."
Youtube, youtu.be/brRZMm2O6vg?si=YBh7gmjEQtnpdx0r. Accessed 6 Sept. 2023.

4. Haba, Duc. *Data Augmentation With Python*. Packt Publishing Ltd, 2023.

5. Snowling, Margaret J., et al. "Defining and Understanding Dyslexia: Past, Present and Future." *Oxford Review of Education*, vol. 46, no. 4, Informa UK Limited, July 2020, pp. 501–13. *Crossref*, https://doi.org/10.1080/03054985.2020.1765756.

6. "The Art of Writing Tests for Dyslexic Students." *April 2020*, www.hltmag.co.uk/apr20/art-of-writing- tests-for-dyslexic-students.