The problem statement is to Identify Learning Disabilities in Children from various Informal Educational Assessment Tools. This project tackles a real-world problem under the hood of special education. If a child faces any sort of difficulty with one or more skills such as reading writing spelling in comparison to their peers or in processing certain information like sequencing, auditory and visual discrimination, memory, then he or she is said to have learning disabilities which is also often used interchangeably with learning difficulties. These are issues are independent of the child’s IQ. In order to quantify these difficulties there are tools such as the informal educational assessment tools like Wepman’s Auditory discrimination test, Burt’s reading test, Schonell’s spelling test which give out a certain score after taking each test, which when we perform inference with a customized model or algorithm predicts the child’s approximate reading, writing spelling age. Reading age or spelling age here is a term that defines the reading ability of a child against their biological age. The model also predicts if the child has any disability in reading spelling or anything else at all and generates a report stating these facts. We are going to deploy this model using a web interface which we started working on and perform the inference with a suitable model after experimenting a handful of models. It is necessary to identify these disabilities in the early years of the child otherwise it may lead to further complications or they might turn out to be the symptoms of higher disorders like dyslexia, dysgraphia, etc. The assessment report basically helps in conveying what kind of disabilities the child is facing. Such children who have been identified can later be dealt with in the special education units opened up in their own schools or any form or additional support.

## We can also use techniques such as PRINCIPAL COMPONENT ANALYSIS? which help in dimensionality reduction of large datasets by recuding them to variables that contain the most information in them

The first innovation is automating the entire assessment process, which traditionally is a manual process with the help of a trained remedial practioner or a special educator. This provides the advantage of making no prior assumptions about the nature of the data obtained from the subject who is the child here and eliminates natural bias when carried out by humans.

The second innovation is training and performing inference with the help of Informal Educational assessment tools.

The other innovations include identifying potential tools for diagnosis, so in the process of doing the assessment we can determine which tools are more precise in predicting LD. Plus we can also study the new classes and combination of occurrence of LD, since it is not necessary that it occurs singly. By applying algorithms such as SOM, we can identify clusters in the population of subjects each corresponding to a new class of LD. Using the clusters identified by the SOM as class labels for the evaluation of Bayesian classifiers using the dataset, and enabling the identification of the tests that can be considered the best for that cluster identification.

Phase 1:

1. PCA
2. Clustering SOM
3. Reliable Tests Shortlist

Phase 2:

1. From Test Scores Predict – Reading/Writing/Spelling Age (for the purpose of a report)
2. Classify if the Child has LD or not
3. Deploy Model and Perform Inference using a Web Interface

1. Utilising PCA to reduce the number of variables contained in the original dataset to a smaller set of combined variables in the projected dataset while retaining most of the information contained within it. The projected dataset makes the application of the SOM unsupervised clustering algorithm both easier and more effective. It becomes easier as the smaller number of variables reduces the computational requirements of the algorithm, and more effective as the first principal components (which form the new set of variables) encode the largest differences between subjects.

2. Applying the SOM algorithm to identify clusters in the population of subjects. The identified clusters however do little in determining which of the original variables to assign a subject into a cluster, since they are computed based on the first principal components.

3. Using the clusters identified by the SOM as class labels for the development and evaluation of Bayesian classifiers using the original dataset, and thus enabling the identification of the tests (giving rise to the original variables) that can b

1. A **self**-**organizing map** (SOM) is a grid of neurons which adapt **to** the topological shape of a dataset, allowing us **to** visualize large datasets and identify potential clusters. They are like a rectangualar lattice of hexagonal structures.

Naive **Bayes** is a kind of **classifier** which uses the **Bayes** Theorem. It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is considered as the most likely class.

1. Mental Attributes Profiling System (MAPS) identification and classifying to a certain type.