**Learning Styles Classification Using natural language processing**

**Abstract**

Exploring learning styles from early years is important for students in result the best practice in lesson delivery that match their style such as visual, aural, Read/write, and Kinesthetic (VARK). the conventional approaches to learning style detection use a questionnaire, but most of the students will choose the answer randomly and that leads to the wrong result, to solve this problem I am going to develop a model that predicts learning styles using sentences written by the students.

**About Dataset**

The dataset consists of 15,451 samples with 2 columns as below :

1. Feature: English sentence
2. Target: Learning Style ( Visual, Auditory, Kinesthetic )

**Methodology**

To get the final model I will move through these Stages:

* **Exploratory Data Analysis** : I will use Term Frequency Analysis which means looking at the most frequent words in each class sentences.
* **Data Preprocessing :** There are various tasks that need to perform in data processing like:

1. Convert the entire text into lower case
2. removing stop words
3. removing digits
4. removing duplicated values
5. Removes all punctuation from a string
6. Replace the text with numeric value using different techniques
7. convert the target column into numeric form using Label Encoder technique
8. Split the data into train and test

* **Model Selection:** The main deep learning architecture for text classification is Recurrent Neural Networks (RNN), most likely I will use Long Short-Term Memory (LSTM) architectures as it is a type of Recurrent Neural Network, but it is better than traditional recurrent neural networks.
* **Train the Model:** After building the neural network, I will feed the model with the training dataset. At the first the weights of the model will be initialized randomly then in each epoch the model will improve the weights using the backpropagation technique.
* **Evaluation:** to check the performance of our model I will apply the testing data to the model and compare the prediction with the real target using these metrics:
  1. **Accuracy:** is defined as the percentage of correct predictions for the test data.
  2. **Precision:** is defined as the fraction of relevant examples (true positives) among all the examples which were predicted to belong in a certain class.
  3. **Recall:** is defined as the fraction of examples that were predicted to belong to a class with respect to all the examples that truly belong in the class.
  4. **F-score:** is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model’s precision and recall.

**Outcomes:**

The result of this project is a trained model with good accuracy that receives sentences from the student and the output will be the Learning Style of that student.

**Timeline:**