**TCS ION**

**RIO-210: Automate Identification of Semantics Errors for Enabling Errorless Proof Reading**

**Project Objective and Brief**

To develop machine learning algorithms with an aim to first detect the grammatical errors and contextual errors from a given sentence or paragraph and then recognize the same.

**Project Guidelines**

1. Prepare or collect a reasonable set of sample sentences and paragraphs with grammatical and contextual errors.

2. Some examples of contextual errors are conjunction errors, punctuation errors, subject-verb errors, etc., in a sentence or paragraph.

3. Develop a machine learning or deep learning algorithm for detection and recognition of grammatical mistakes from a sentence or paragraph.

4. Enhance the previous algorithm with the capability of detection and recognition of contextual errors from a sentence or paragraph.

5. Test the application for reasonable accuracy.

**Algorithms used:**

1. **Data Preprocessing**:
   * Text cleaning and tokenization.
   * Stopword removal and lemmatization.
2. **Feature Extraction**:
   * TF-IDF vectorization of the cleaned text.
3. **Clustering**:
   * Application of K-Means clustering to the TF-IDF vectors.
4. **Error Type Classification**:
   * Use of Random Forest to classify the type of grammatical error based on features.
5. **Grammar Correction**:
   * Utilization of a pre-trained T5 model to correct grammatical errors in sentences.
6. **Model Evaluation**:
   * Evaluation of the Random Forest classifier using accuracy and classification reports to measure performance.