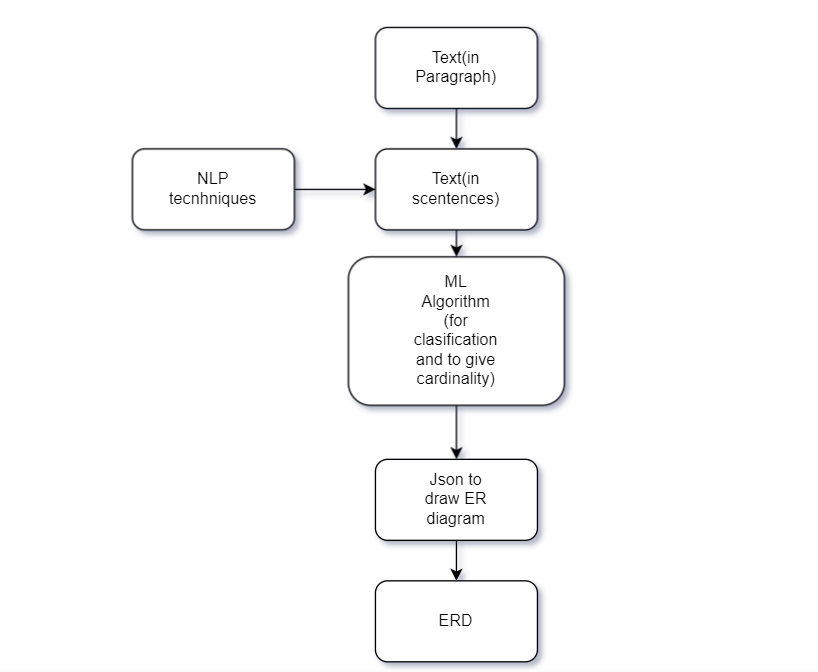
Text to ER

Aim: My main aim is to take text paragraph from the user and convert it into ER diagram.

Flow chart:



Steps involved:

Step 1: Take the paragraph from the user and do sentence segmentation by using nltk library.

Step 2: Take each sentence and do word segmentation by using spaCy library.

Remove unwanted characters in the given text if any.

Tagging the parts of speech with the words by using nltk library.

Step 3: By using classification algorithms like SVM, we can classify the Parts of speech tagged word to Entity, attribute, or relation. This can be done by taking a dataset and training the model or writing our own dataset and training the model.so that ML algorithm can classify the Tagged words of a sentence to entity, attribute, or relation.

SVM mean support vector machine is an ml algorithm used in classification problems it draws a hyperplane which separates the data given and helps in classification based on the dataset in which it got trained.

Step 4: use JSON to convert the final classified text (Entity, attribute, or a relation) into ER diagram.

Detailed Design:

Step 1:

Input from user, i.e. in paragraph and do sentence segmentation by using nltk library. Stores in new variable “Sentences”.

Diagram, text

Description automatically generated

Step 2:

Takes variable “Sentences” and remove unwanted special characters and stores in the same variable.

Take each sentence and do word segmentation by using spaCy or nltk libraries.

And do tagging POS to words using nltk and store in a variable named “tagged\_words”.

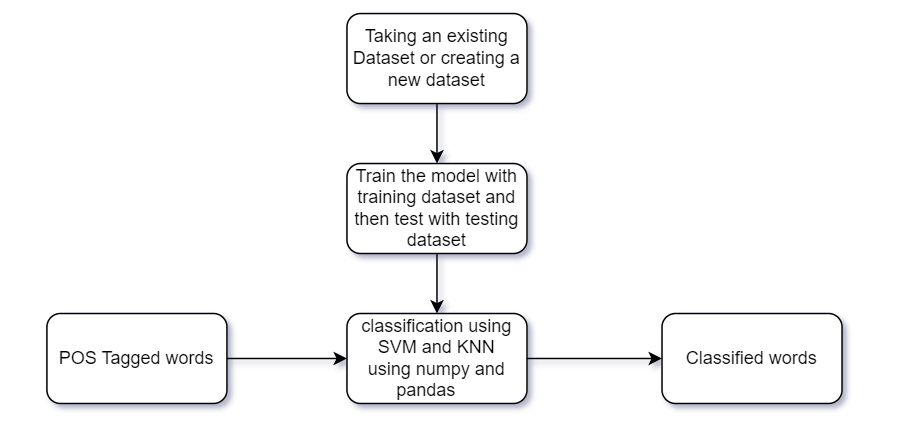
Diagram

Description automatically generated with low confidence

Step 3: Take an existing Dataset or create a new Dataset to train the ML model.

Use numpy package and pandas library for writing SVM and KNN algorithms.

The tool will first get trained by the dataset given, and then read the variable “tagged\_words” and will classify it as an entity, attribute, or a relation.



Step 4: JSON

Here we use JSON to draw E-R diagram. This module take the input from the ML module, ix`.e. the entity, attributes, relation, and the cardinality of it. And will draw the final E-R diagram.

Diagram, text

Description automatically generated