**MACHINE LEARNING ASSIGNMENT**

These are the steps used to solve this problem.

**STEP 01:**

Reading a .csv file.

**STEP 02:**

*DATA PREPROCESSING*

- Checking missing values in the dataset.

- Data Cleaning (Dropping special character, Removing whitespace from the beginning and end of a string)

**STEP 03:**

Using nltk (Natural Language Toolkit) library to perform these tasks:

- Tokenization (Tokenization is a way of separating a piece of text into smaller units called tokens.)

- Removing stopwords (eliminate words that are so commonly used that they carry very little useful information.)

- Stemming (Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma)

**STEP 04:**

Creating a new column (skills). This column contained Technical skills.

Python libraries that are used in this assignment are:

* Pandas
* nltk (Natural Language Toolkit)
* re (Regular Expression)