Business Objective: Classification of US publicly listed companies based on business description

Files Attached

- **Company Business Description**

  This file has the list of all US publicly listed companies. It has the company name, stock ticker and business description.

- **Training Data**

  This file has 2,000 US publicly listed companies which are categorized by certain tags mentioned in the column G. You will see lot of tags for one company. You need to worry about the first tag only. Ignore the rest of the tags.

  Note: You need to create new variable with first tag from column G variable it will be your target variable and business description will be feature variables.

Analysis

You will have to classify the companies mentioned in the "Company Business Description" file as per the tags mentioned in the "Training Data" file. As mentioned above please consider the first tags only and ignore the rest of the tags.

Note: You need to build model using training data, classify the Companies in 'Company Business description file.

Key expectations (these steps need to be included as part of the processing):

- Be systematic and give some reasons behind every task that you perform
- Perform detailed text processing steps including tokenization, conversion of words to lowercase, remove stop words, remove unnecessary words, take care of abbreviations/short names, Named entity recognition, Parts of speech tagging, stemming, lemmatization etc.)
- Perform detailed exploratory data analysis (example: word clouds, word frequencies, word clustering, identify key topics using topic mining etc.)
- Incorporate some visualization may be using some simple plots
- Build classification model with multiple techniques including decision trees, Random Forest, XGboost, ANN, SVM etc.
- Build the models with different vectorization methods including count vectorizer, TF-IDF vectorizer, word embedding's etc.
- Perform other variable reduction steps.
- Calculate all the goodness of fit metrics including precision, recall, f1-score etc.
- Validate the model on test data.
- Create simple app (robot) with User Input and output. If user provides input of business description, the app should reacts with output of classification