Project

On

Analysis and Prediction of CSRIC Best Practices

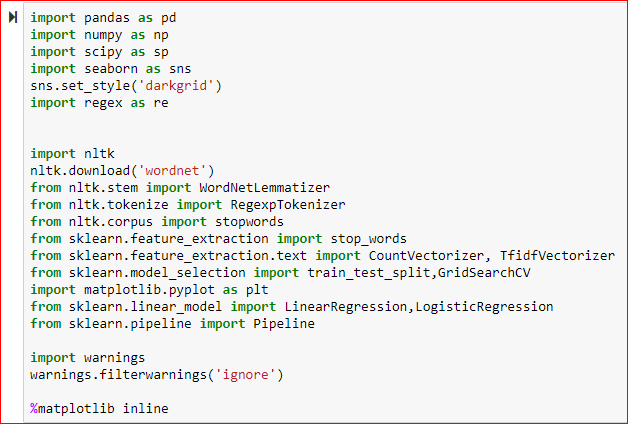
**Introduction**

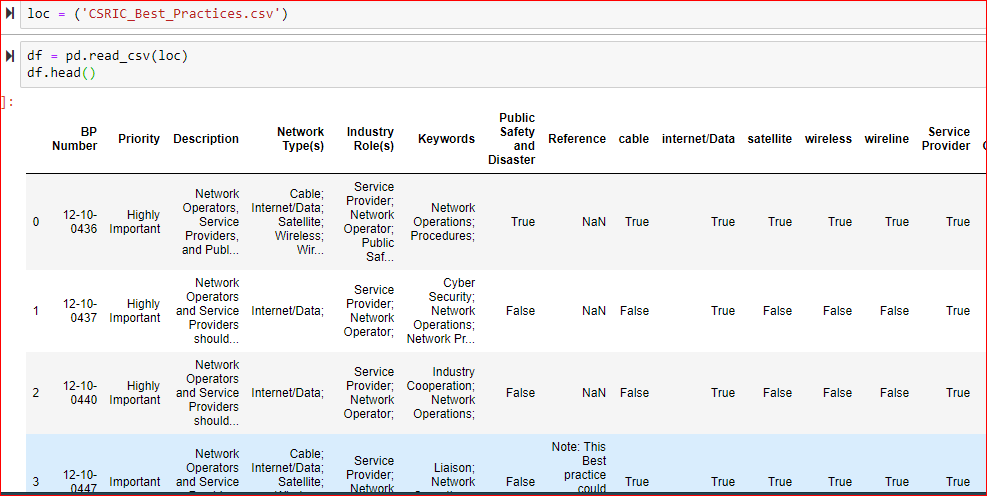
This is the data regarding the Federal Communications Commission(FCC).FCC follows the CSRIC Best Practices Search Tool which allows us to search CSRIC’s collection of Best Practices using variety of criteria including Network Type , Industry Role , Keywords , Priority Levels and BP number.

The approach is to apply **natural language processing** techniques on the description of the event and see how that correlates to the priority of the event.

**Steps involved**

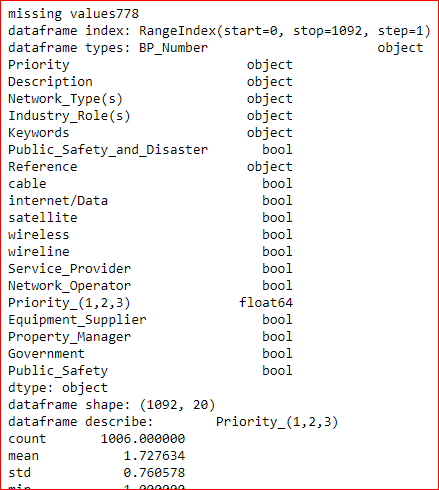
1. Importing the modules needed for the project



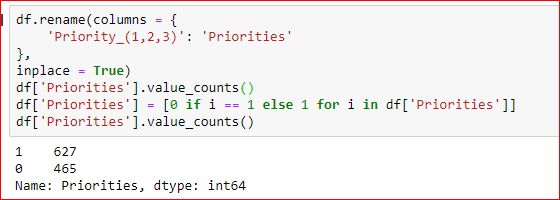
1. Load the data from csv file 
2. For better understanding the data,’eda’ is a function that view any null values, replace any blank spaces with an underscore, reformat the data frame index, see the data types for each column, display any duplicated data, describes the statistical analysis of the data & checks the shape.

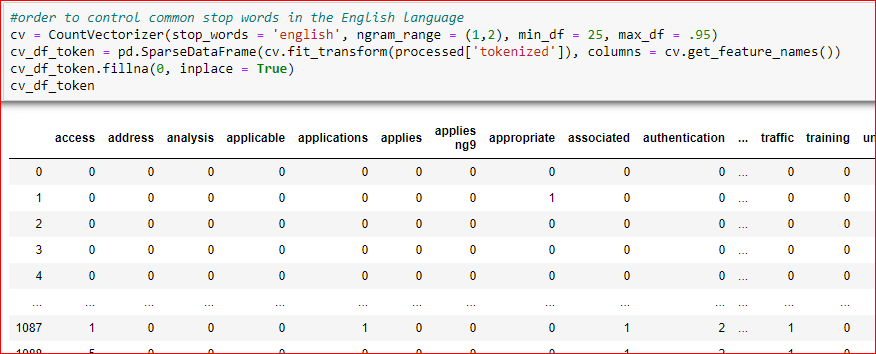


Output



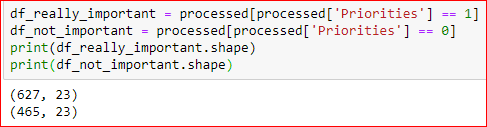
1. Priority columns are the object related column that ranks the severity of the event to important, highly important & critical.



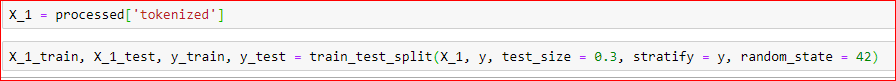
1. **Pre-processing** is the technique of converting raw data into a clean data set.We will preprocess the data and clean out the common words.
2. Apply ‘CountVectorize’ on our stemmed, lemmatized & tokenized description words in order to control common stop words in the English language. 

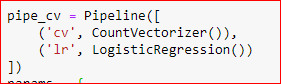
Observation: most of the words that popped up are network or security related.

1. Group the descriptions based on the urgency or severity of the importance of the event. iIt is ranked as 0 (Not Important) or really bad ranked as 1 (Really Important).

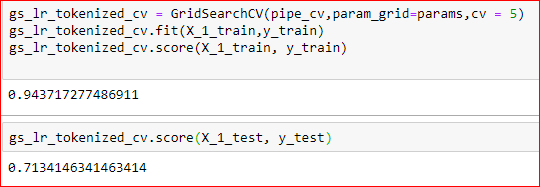


1. Modeling regression & classification metrics on the tokenized data

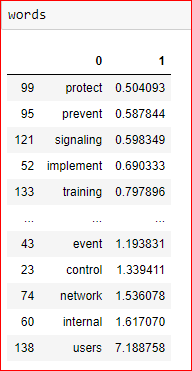




1. Apply the pipeline on our logistic regression model within a grid search object



11. predict the outcomes of the best y variable.



Conclusion: This model accuracy for a training data as well as 10% increase in accuracy for our testing data.

Complex events are:

* Secure networks
* encrypted systems

Simple events are:

* Virus protection
* Providing secure transfer of information.