FLIP ROBO

Malignant Comment Classifier.

Submitted by:

Sagnik Das

ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to (Datatrained) also to my SME (Mr Keshav Bansal). Who gave me the golden opportunity to do this wonderful project on the topic (Micro Credit Defaulter Project), which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to tall of them.

INTRODUCTION

- ## Business Problem Framing

  The proliferation of social media enables people to express their opinions widely online. However, at the same time, this has resulted in the emergence of conflict and hate, making online environments uninviting for users. There has been a remarkable increase in the cases of cyberbullying and trolls on various social media platforms. This can take a toll on anyone and affect them mentally leading to depression, mental illness, self-hatred and suicidal thoughts. So, the goal of this project is to identify and classify hate comments.

- ## Conceptual Background of the Domain Problem

  Our goal is to build a prototype of online hate and abuse comment classifier which can used to classify hate and offensive comments so that it can be controlled and restricted from spreading hatred and cyberbullying.

- ## Review of Literature

  This is a comprehensive summary of the research done on the topic cyberbullying through the spread of hate comments and threats.

- ## Motivation for the Problem Undertaken

  Our goal is to build a prototype of online hate and abuse comment classifier which can used to classify hate and offensive comments so that it can be controlled and restricted from spreading hatred and cyberbullying.

Analytical Problem Framing

- Mathematical/ Analytical Modeling of the Problem

  Covariance Matrix computation

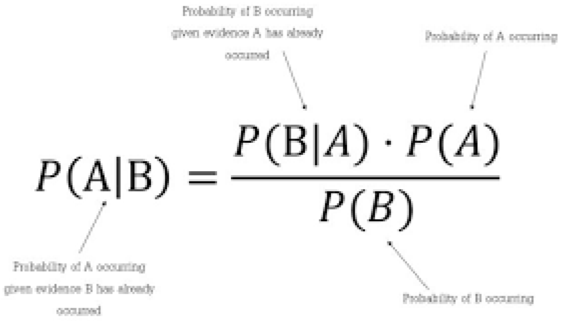
$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

Covariance Matrix for 3-Dimensional Data

Linear Regresion

$$Y = a + bX$$

Naïve Bayes Classifier


Probability of B occurring given evidence A has already occurred
Probability of A occurring
$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$
Probability of A occurring given evidence B has already occurred
Probability of B occurring

Support Vector Machine.

- Data Sources and their formats

  Data Provided by Flip Robo Technologies.

- Data Preprocessing Done

  Classifying the data into the headings Malignant, highly_malignant, rude, threat, abuse, loathe.

- Data Inputs- Logic- Output Relationships

  Classifying the data into the headings Malignant, highly_malignant, rude, threat, abuse, loathe and put them in catagories 0 and 1

- State the set of assumptions (if any) related to the problem under consideration

  Here, you can describe any presumptions taken by you.

- Hardware and Software Requirements and Tools Used

  Hardware – Laptop
  Software – import pandas as pd, import os, import csv, import sklearn, logistic regression, svm, naïve bayes algorithm.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  Confusion Matrix, Support Vector Machine, Logistic Regression, Naïve Bayes Algorithm.

- Testing of Identified Approaches (Algorithms)

  Confusion Matrix, Support Vector Machine, Logistic Regression, Naïve Bayes Algorithm.

- Run and Evaluate selected models

  Confusion Matrix, Support Vector Machine, Logistic Regression, Naïve Bayes Algorithm.

- Key Metrics for success in solving problem under consideration

  Confusion Matrix

- Visualizations

  Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

  If different platforms were used, mention that as well.

- Interpretation of the Results

  So based on this dataset, Support Vector Machine appears to be a superior predictor of hate speech. Logistic regression also produced excellent results. This dataset appears to be an artificial intelligence product used to classify hate and abusive speech.

CONCLUSION

- Key Findings and Conclusions of the Study

  So based on this dataset, Support Vector Machine appears to be a superior predictor of hate speech. Logistic regression also produced excellent results. This dataset appears to be an artificial intelligence product used to classify hate and abusive speech.

- Learning Outcomes of the Study in respect of Data Science

  This dataset appears to be an artificial intelligence product used to classify hate and abusive speech.

- Limitations of this work and Scope for Future Work

  There is a high scope of work on this field as the usage of social media is increasing.