Headspace Marketing Campaign

By: Seraj Khazei

Overview

In this project, we are looking at facebook statuses, and the neurotic score based on the language used in the status update. The dataset includes more than 500,000 status updates. This dataset is well-suited for the business problem, as the goal is to identify neurotic individuals in order to send them ads for headspace.

For data preparation, I dropped non-relevant columns and assigned neurotic or not neurotic to each status update. Then, i used re to get rid of unimportant characters in the status update. After that, i tokenize the status to split each status into its respective words so that each word is a feature in the model. We also remove stopwords as they provide little to no sentiment value to the status. Lemmatization is performed to group words with the same meaning together as one word. TF-IDF is applied in order to assign each word in each tweet a numeric value based on its importance across all tweets. Finally, SMOTE is applied in order to eliminate the class imbalance observed in the target variable (sentiment). We used pandas to perform data filtering and visualization, nltk to perform text preprocessing, SMOTE for oversampling and sklearn for TF-IDF.

For modeling, i utilized the sklearn's LogisticRegression, DecisionTree and RandomForestClassifier methods. I tuned our models using GridSearchCV also provided by sklearn.

Our final model has an f1 score of 84%, which means that it correctly identifies the sentiment of statues 84% of the time. I used a validation set and the test set f1 scores in order to validate our model performance.

Business Understanding

Headspace among everyone in the world has noticed there has been a spike in depression and anxiety since the Pandemic has started. Headspace as being a mediation app understands they are capable of reducing the depression or anxiety rate at any scale, so in order to fufill the following task Headspace has created a special promotion for those who are already have mental illness or are of risk to mental illness. Although Headspace needs help to reach the individuals who meet the following requirement. This is where I come in to strategize a way to target the ads towards the following individuals and this notebook will be a demonstration for how the objective was tackled.

Data Understanding

```
    The simplest way to use nbconvert is

    > jupyter nbconvert mynotebook.ipynb --to html

    Options include ['asciidoc', 'custom', 'html', 'html_ch', 'html_embed', 'html_to
c', 'html_with_lenvs', 'html_with_toclenvs', 'latex', 'latex_with_lenvs', 'markdown',
'notebook', 'pdf', 'python', 'rst', 'script', 'selectLanguage', 'slides', 'slides_with
_lenvs', 'webpdf'].

    > jupyter nbconvert --to latex mynotebook.ipynb

    Both HTML and LaTeX support multiple output templates. LaTeX includes
    'base', 'article' and 'report'.  HTML includes 'basic' and 'full'. You
    can specify the flavor of the format used.

    > jupyter nbconvert --to html --template lab mynotebook.ipynb

    You can also pipe the output to stdout, rather than a file

    > jupyter nbconvert mynotebook.ipynb --stdout

    PDF is generated via latex

    > jupyter nbconvert mynotebook.ipynb --to pdf

    You can get (and serve) a Reveal.js-powered slideshow

    > jupyter nbconvert myslides.ipynb --to slides --post serve

    Multiple notebooks can be given at the command line in a couple of
    different ways:

    > jupyter nbconvert notebook*.ipynb
    > jupyter nbconvert notebook1.ipynb notebook2.ipynb

    or you can specify the notebooks list in a config file, containing::

        c.NbConvertApp.notebooks = ["my_notebook.ipynb"]

    > jupyter nbconvert --config mycfg.py

To see all available configurables, use `--help-all`.
```

In [ ]: