November 15, 2019

ML Business Plan

FigureEight Disaster Recovery

Company: Figure Eight (https://www.figure-eight.com)

Mission: Figure Eight (formerly Crowdflower) is a human-in-the-loop machine learning and artificial intelligence company based in San Francisco. The company raised $58 million in venture capital and was acquired by Appen in March 2019 for $300 million. Figure Eight crowdsourced the tagging and translation of messages to apply artificial intelligence to disaster response relief.

Business Issue: Figure Eight needed a better way to classify incoming text messages in order to send them to an appropriate disaster relief agency and eventually supply the most efficient resources. If resources are

Project Outline: This project was outlined in the following way:

Objective: build a model for an API that classifies disaster messages

- Cleaning data set containing real messages that were sent during disaster events

ETL Pipeline:

- Loads the ‘messages' and ‘categories' datasets
- Merges the two datasets
- Cleans the data
  - Splits the categories column into separate
  - Clearly named columns
  - Converts values to binary
  - Drops duplicates
- Stores it in a SQLite database

- Once data is cleaned use NLP techniques to extract out word relations

NLP Processing:

- Loads data from the SQLite database
- Splits the dataset into training and test sets
- Builds a text processing and machine learning pipeline
  - Use tokenization function NLTK to case normalize, lemmatize, and tokenize text to vectorize and then apply TF-IDF to the text

Once word relations are established use a machine learning model to classify data

Machine learning Pipeline:

- Trains and tunes a model
  - Performs multi-output classification on the 36 categories in the dataset
  - Use GridSearchCV to find the best parameters for the model
- Outputs results on the test set
  - F1 score, precision and recall for the test set is outputted for each category
- Exports the final model as a pickle file

Deployment:

- Web application where an emergency worker can input a new message and get classification results in several categories

- When a user inputs a message into the app, the app returns classification results for all 36 categories.

- Web application that will also display visualizations of the data using Plotly

Technology Stack used: Python, SK-Learn, Pandas, SQLite database, flask, html, css and javascript, Plotly, NLTK, TF-IDF

Video Of Final Product: