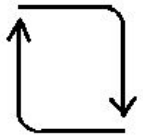
Twitter Sentiment Analysis
@tweetanalysis

Exploration in Data Science - Project Presentation by

@LarrisaCarvalho
@SiriKoduru
@RuchaKulkarni

Objective

- To classify Tweets as positive or negative.

- Obtain high accuracy on classifying sentiment in Twitter messages using machine learning techniques.

Twitter

- Microblogging and social networking service
- Users post and interact with "tweets"
- Allows businesses to engage personally with consumers

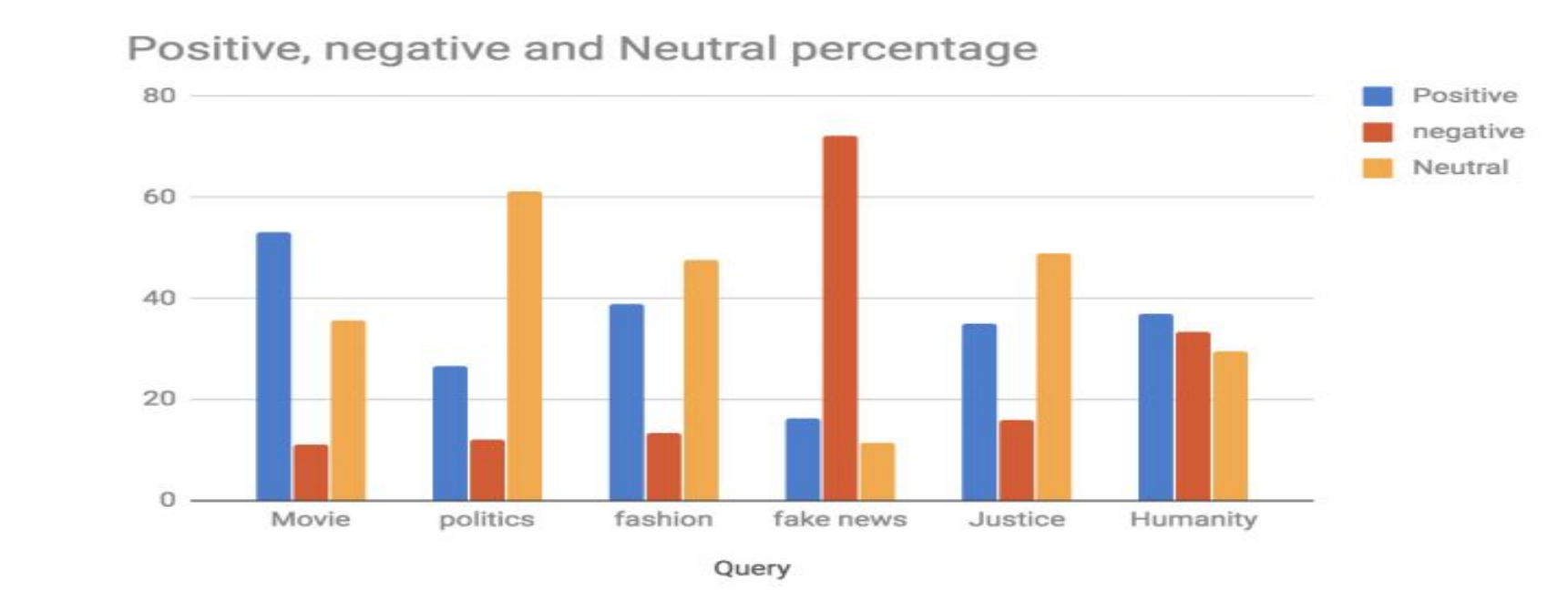
What is Sentiment Analysis?

- Sentiment analysis is a natural language processing technique
- Used to determine whether data is positive, negative or neutral.
- Often performed on textual data

Example

Tweet	Parameter	Sentiment
“Loving the reels feature on Instagram”	Instagram	Positive
“Bad service by ABC auto, not recommended”	ABC auto	Negative
“Landed at Portland”	Portland	Neutral

Statistics


Positive, negative and Neutral percentage
80
60
40
20
0
Positive
negative
Neutral
Movie
politics
fashion
fake news
Justice
Humanity
Query

Why is Sentiment Analysis Important?

- Customers express their thoughts and feelings more openly
- Opinions in survey responses and social media conversations,
- It allows businesses to understand the sentiment of their customers
- Allows brands to learn what makes customers happy or frustrated
- Businesses can make better and more informed decisions.

Twitter Sentiment Analysis Use Cases

- Customer Service
- Market Research
- Brand Monitoring
- Social Media Monitoring
- Political Campaigns

Social Media Marketing

- Online reputation
- Bad review can be costly
- Help you detect angry customers or negative mentions before they escalate.
- Get valuable insights that drive decisions

Political Campaigns

- Twitter is excellent place to measure public opinion

- How people feel about a specific candidate

Data Gathering

- Kaggle

- Consists of some tweets along with their sentiment

Train the Data

- We use tweets from data set to train our model.

- In order to train the model we need to represent each of the tweet in the form of a feature vector using Word2Vec.

How we are going to classify the data?

We are going to classify the Tweets from Twitter as positive, negative or neutral tweets.

Few Positive Tweets Examples:

1. @Msdebramaye I heard about that contest! Congrats girl!!
2. UNC!!! NCAA Champs!! Franklin St.: I WAS THERE!! WILD AND CRAZY!!!!!! Nothing like it...EVER http://tinyurl.com/49955t3

How we are going to classify the data?

Few Negative Tweets Examples:

1. no more taking Irish car bombs with strange Australian women who can drink like rockstars...my head hurts.
2. Just had some blood work done. My arm hurts

Why Word2Vec model?

1. Deep learning model developed by Google.
2. Capturing the context of words..
3. Takes an input of a large corpus of documents like tweets or news articles and generates a vector space of typically several hundred dimensions.
4. Important concept behind word2vec is that word vectors that are close to each other in the vector space represent words that are not only of the same meaning but of the same context as well.
5. Word2vec delivers enhanced feature engineering for raw text data.

Training a classifier for classifying tweets

We are planning to use these 3 machine learning algorithms in order to classify the feature vectors i.e. the tweets and then classify them accordingly.

Algorithms:

1. SVM
2. Random Forest
3. Multi Layer Perceptron

SVM

1. SVM - a supervised learning algorithm.
2. most robust prediction methods.
3. Non probabilistic binary classifier
4. Maps training examples to points in space to maximize width between categories.
5. New examples are then mapped into the same space and classified accordingly.
6. Non linear classification also can be done.

MLP

1. A class of feed forward neural networks.
2. 3 layers - input, hidden, output layers.
3. Activation function for all nodes except input nodes.
4. Backpropagation.
5. MLP is distinguished from a linear perceptron by its numerous layers and non-linear activation.
6. It can tell the difference between data that isn't linearly separable and data that is.

Random Forest

1. Ensemble learning method for classification.
2. The output of the random forest is the class selected by most trees.
3. For regression tasks, the mean or average prediction of the individual trees is returned.
4. Random decision forests correct for decision trees' habit of overfitting to their training set.
5. Performance is better than decision trees.

What to do after we train the classifier with train data?

After training we need to test the data to check the performance of each algorithm.

Now we test the data with the real time twitter data.

How do we do this?

Tweepy

Tweepy is the python client for the official Twitter API.

- The tweets need to be gathered so as to perform Sentiment analysis on those tweets. They can be fetched from Twitter using the Twitter API.
- In order to fetch tweets through Twitter API, one needs to register an App through their twitter account.
- Get the 'Consumer Key', 'Consumer Secret', 'Access token' and 'Access Token Secret'
- Establish connection with twitter get the tweets.

What next to do after getting data from twitter

We use Tweepy Cursor object in order to get the tweets from twitter based on a keyword.

For example we need to get the tweets for a movie to see whether the tweets have been positive, negative or neutral regarding the movie.

So the key word movie name “Fast and Furious” for example is passed as keyword for Tweepy cursor and then we predict from the tweets.

Thank you