

Sentiment Analysis using twitter api





Hello!

Zia Khan

zia@thedevmasters.com



CRISP-DM

1

Business Objective

2

Data Understanding

3

Data Preparation

4

Modeling

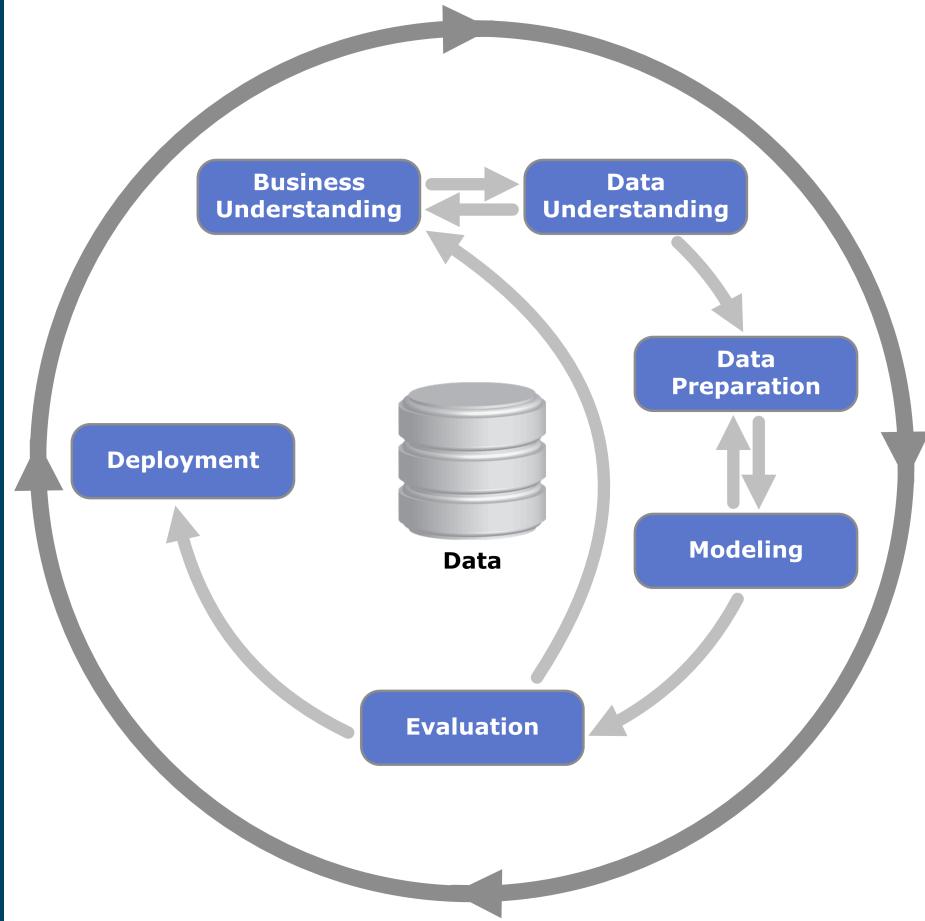
5



CRISP-DM

Overview

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment



CRISP-DM

1

Business Objective

2

Data Understanding

3

Data Preparation

4

Modeling

5



2. Classify tweets

The business objective of the project is to gather tweets using both search and streaming Twitter API. Save the tweets from the streaming API into a file, read the file, analyze and classify them as positive or negative.



Gathering data

Read twitter API (search and stream)

- Create a twitter account
- Create an app
- Get the keys



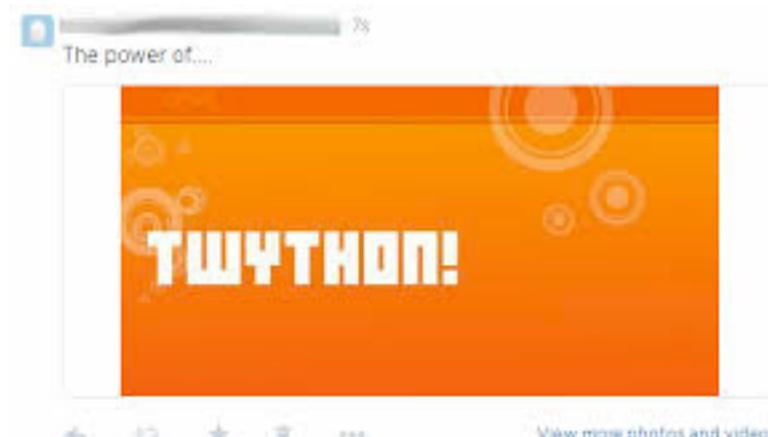
A screenshot of the Twitter Apps interface. It features a blue header bar with the Twitter logo and the text "Twitter Apps". Below the header, there's a single app entry card for "Sentiment Analysis Project Zia". The card includes a small profile picture, the app name, and a brief description: "Sentiment Analysis using Python or R". There's also a "Create New App" button.



Gathering data

Read twitter API (search and stream)

- Read twitter API (using Tweepy or Twython)
- !pip install tweepy



CRISP-DM

1

Business Objective

2

Data Understanding

3

Data Preparation

4

Modeling

5



Tweet

```
{u'search_metadata': {u'completed_in': 0.1,
    u'count': 100,
    u'max_id': 963505361629032453,
    u'max_id_str': u'963505361629032453',
    u'next_results': u'?max_id=962952109082648575&q=data%20science&geocode=33.6845673%2C-117.82650490000003%2C50mi&count=100&include_entities=1',
    u'query': u'data+science',
    u'refresh_url': u'?since_id=963505361629032453&q=data%20science&geocode=33.6845673%2C-117.82650490000003%2C50mi&include_entities=1',
    u'since_id': 0,
    u'since_id_str': u'0'},
  u'statuses': [{u'contributors': None,
      u'coordinates': None,
      u'created_at': u'Tue Feb 13 20:09:13 +0000 2018',
      u'entities': {u'hashtags': [{u'indices': [22, 27], u'text': u'data'},
          {u'indices': [29, 37], u'text': u'science'}],
        u'media': [{u'display_url': u'pic.twitter.com/KRgH6gn9b0',
          u'expanded_url': u'https://twitter.com/ProfBaird/status/963505361629032453/photo/1'},
          {u'id': 963301183229116417,
          u'id_str': u'963301183229116417',
          u'indices': [63, 86]}]}]}
```



CRISP-DM

1

Business Objective

2

Data Understanding

3

Data Preparation

4

Modeling

5



Data Understanding

- Get the tweets
- Tweets are in a dict format
- Use the statuses key of dict to create DataFrame
- User key has many attributes within
- Create a separate DataFrame of user key
- Combine the two DataFrames



Analyze tweets

- Analyze regions that are tweeting the most
- Count vectorize the tweets
- What type of words are being used



Classify tweets

- Get tweets ready for the Algorithms
- Classify
- Rate the tweets as positive and negative





Thanks!

Any questions?

zia@thedevmasters.com

