**Python Certification Project Report -Edureka**

Sentiment Analysis for Birthday Celebrities

Manoj K V

05/03/2017

**Index:**

# **Problem statement**

# **Solution Flow Diagram**

# **Tools and Packages Used**

# **Solution: Python Script with proper commenting**

# **Output**

# **Challenges Faced during the project**

# **Appendix**

**Problem statement**

IMDB provides a list of celebrities born on the current date. Below is the link:

<http://m.imdb.com/feature/bornondate>

Get the list of these celebrities from this webpage using web scraping (the ones that are

Displayed i.e. top 10). You have to extract the below information:

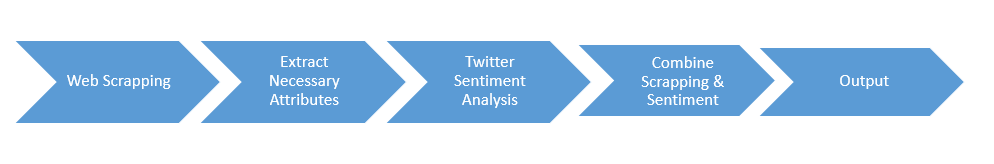
* Name of the celebrity
* Celebrity Image
* Profession
* Best Work

Once you have this list, run a sentiment analysis on twitter for each celebrity and finally the output should be in the below format

* Name of the celebrity:
* Celebrity Image:
* Profession:
* Best Work:
* Overall Sentiment on Twitter: Positive, Negative or Neutral

Hint: Use IMDB scrapping sample example as reference for scraping the mentioned web page. For sentiment analysis, use the Twitter sentiment code as reference.

**Solution Flow Diagram**



**Tools and Packages Used**

Tools:

* PyCharm Community Edition 2017.1

Packages:

* selenium
* bs4
* re
* csv
* tweepy
* codecs
* string

**Solution: Python Script with proper commenting**

* **imdbWebscrapping.py**

This is the main Python Script, which will extract web components from the given URLs first, and then calling the child Python script for sentiment analysis on twitter.

****

* **twitterSentimentAnalysis.py**

This script is used to publish the tweets and do sentiment analysis for the given Celebrities

****

**Output:**

* Final output of Python Scripts Execution:



**Challenges Faced during the project**

* Dynamic URL identification
* Best Film attribute has comma (,) in its values
* Output file writer
* Plugin Twitter script in IMDB web scrapping scripts

**Appendix:**

Below are the keywords files used for Sentiment Analysis on Twitter

**Negative Keywords:** This file contain all the negative keywords



**Positive Keywords:** This file contain all the positive keywords

