Durgesh Samariya

Tweelyzer

An Approach to Sentiment Analysis of Tweets

Anchor Academic Publishing

disseminate knowledge

Samariya, Durgesh: Tweelyzer. An Approach to Sentiment Analysis of Tweets, Hamburg, Anchor Academic Publishing 2016

PDF-eBook-ISBN: 978-3-96067-590-7
Druck/Herstellung: Anchor Academic Publishing, Hamburg, 2016

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Bibliographical Information of the German National Library:
The German National Library lists this publication in the German National Bibliography. Detailed bibliographic data can be found at: http://dnb.d-nb.de

© Anchor Academic Publishing, Imprint der Diplomica Verlag GmbH
Hermannstal 119k, 22119 Hamburg
http://www.diplomica-verlag.de, Hamburg 2016
Printed in Germany

ABSTRACT

The underlying trend of people using microblogging to express their thoughts on various topic has increased the need for developing computerised techniques for automatic sentiment analysis on texts that do not excessed 200 characters. Twitter is a "micro-blogging" social networking site that has a large and rapidly growing base on users. Twitter's tweets or messages are limited to 140 characters. Because of limitation, it is more difficult to express sentiment and the classification of the tweets difficult as well. The sentiment analysis can be done by two types: emotion and opinion. This research completely focus on sentiment analysis of opinions. These opinions can be divide in three different classes: positive, negative and neutral (Between positive and negative).

The main goal of this project is to build a model that predict election movement and provide sentiment score from Twitter message (which can not exceed 140 characters). In this project, I apply the novel approach that classify sentiment and emotions of Twitter tweets automatically. After that message is categorised in classes (positive, negative, neutral). For the sentiment first of all, I retrieved tweets from twitter and the convert in to dataset. Therefore applied pre-processing (Data Cleaning) to dataset. After pre-processing applied proposed algorithm namely : TWEELYZER to dataset. At the end I measured performance of TWEELYZER in term of accuracy and recall.

In this project, all tweets of people regarding to movie, brand, actor, actress was collected form twitter and then cleaned and analysed according to proposed algorithm. These tweets were collected using R Studio software. Different process took place in pre-processing tweets, I clean tweets in order to make better for analysis. In process of pre-processing I add some stopping words, removing Hash(#) tag, removing punctuation marks, removing of URLs (ex: www.abcd.com/xyz/abc), etc. After pre-processing, using R Studio developed different insights.

Keywords: Sentiment Analysis, Twitter, Data Analysis, Classification, Big Data, Social Media, R Studio

ACKNOWLEDGEMENT

I wish to express our heartfelt gratitude to **Dr.G.Viswanathan**, Chancellor, VIT University,Vellore for providing facilities for the fourth semester project.

I am highly grateful to our Vice Presidents, **Shri. Sankar Viswanathan, Shri. Sekar Viswanathan and Shri. G.V.Selvam**, Vice Chancellor **Dr. Anand A. Samuel**, Pro-Vice Chancellors **Dr.V.Raju and Dr. S. Narayanan** for providing necessary resources.

My sincere gratitude to **Dr. Aswani Kumar Cherukuri,**, Dean, School of Information Technology and Engineering for giving us the opportunity to undertake the project.

I wish to express my sincere gratitude to **Dr.P.M.Durai Raj Vincent**, HOD , Department of Digital Communications and the Project Coordinator **Prof.M.Deepa**, Assistant Professor(Senior), School of Information Technology and Engineering for providing me an opportunity to do my project work in the **industry/VIT University.**

I would like to express my special gratitude and thanks to my internal guide **Dr. Dhinesh Babu L.D,** Associate Professor, School of Information Technology and Engineering whose esteemed guidance and immense support encouraged to complete the project successfully.

I thank the Management of VIT University for permitting me to use the library resources. I also thank all the faculty members of VIT University for giving me the courage and strength I needed to complete my goals. This acknowledgement would be incomplete without expressing my whole hearted thanks to my family and friends who motivated me during the course of the work.

SAMARIYA DURGESH OMPRAKASH

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES