FINAL YEAR PROJECT 1

NUR SYAMIMI BINTI AZIZ
BTBL 17046420
ISM SAINS KOMPUTER (KRK 5)
SV: SIR AHMAD FAISAL FAISAL AMRI BIN ABIDIN
@BHARUN

TITLE: Classifying sentiments in social networks: Integrating text-based affective method with support vector machine algorithm.
ABSTRACT:

Sentiments classification and opinion mining is the field of study that classify people’s opinions, sentiments, evaluations, emotions, and attitudes from written language. The growing importance of classifying sentiment coincides with the growth of social media such as Twitter, blogs, and many other social networks. People interact with each other, share their ideas, opinions, interests and personal information. These user comments are used for finding the sentiments in a social network.

An increase in activity of opinion mining, or also known as classifying sentiment, challenges are getting added every day. There is a need for automated classify techniques to extract sentiments and opinions conveyed in the user-comments. Recognizing emotion is extremely important for a text-based communication tool. This study purposed a support vector machine algorithm-based approach. The main goal of classifying sentiments in a social network is to explore how text classifying techniques can be used to dig into some of the data.
CHAPTER 1:

1.1 INTRODUCTION:

Classifying sentiments in social media is about how to classify emotion on the text-based in social media. It is also the field of study that classify people’s opinions, sentiments, evaluations, emotions, and attitudes from written language. The written text describes one’s emotion or attitude. An effective method had been used to integrate text-based which is a Support Vector Machine Algorithm. Support Vector Machine Algorithm is a machine learning which can be used for both regression and classification tasks. But, it is widely used in classification objectives.

1.2 PROBLEM STATEMENT:

This project was created to classify emotions based on text in social media which is the output could be either a positive, negative or neutral statement. The research on classifying sentiment has been going for a long time ago. However, it becomes the major issue in field of research and technology. Due to day by day, it was an increasing number of users on social media. So, data produces had a huge amount in the form of text, audio, video and others. There is need to do classifying sentiment based on texts in form of post or comment. Although there are many research about sentiment classify before, this project created to integrate text based affective method with Support Vector Machine Algorithm.
1.3 OBJECTIVES:

The aim of this project is to study classifying sentiment in social networks and integrating text based affective method by using Support Vector Machine Algorithm. The objectives are:

i. To predict the text either positive, negative or neutral based on people’s feelings, emotions, and attitudes.

ii. To apply an affective method with Support Vector Machine Algorithm in social media environments.

iii. To test a correlation between Twitter sentiment and events that have occurred.

1.4 SCOP:

The main highlight of this project is to classify sentiment in social media which involves the computer program, target user, and researcher. Firstly for target user, text that had been write by user will be tested by using computer program to classify the sentiment in social media (twitter). The researcher will then classify the data and text from the social media (twitter) either positive, negative or neutral statement.
1.5 LIMITATION:

There are a few method and algorithm can be used to classify sentiment in social media. This project is only focusing on the Support Vector Machine Algorithm which is had a variety of task including social network filtering. Rapid miner is used for creating twitter sentiment classify model. The method used is template-based frameworks. It provides text mining and machine learning. The model used in the rapid miner is polynomial modelling which is a non-linear statistical data modelling and also as a decision tools.

1.6 EXPECTED RESULT:

This project focuses on the classifying sentiment in social media by integrating text based affective method with Support Vector Machine Algorithm. This study was expected to classify text in social media (twitter) by using the focused method.
CHAPTER 2:

LITERATURE REVIEW:

2.1 INTRODUCTION:

Literature review is a process to read, analyse, evaluate, and summarize the techniques and methods that should be taken to the implementation of this project. The information has been gathered to get a better understanding for this project.

Classifying sentiment in social media is a thorough research of how opinions of peoples can be related to their emotions and attitude that shows in natural language. It has been a widely used area over the years and still, it leaves a lot to be researched. From this research, it used many instruction to classify the sentiment in social media especially on twitter. Support Vector Machine (SVM) Algorithm had been used to classify the sentiment on twitter. There are some pros and cons to this algorithm. SVM algorithm entails more processing power and is mainly employed in graphics processing units. However, SVM and its family algorithms give better accuracy in many sentiment applications, providing that they have enough data to train them.

2.2 SUPPORT VECTOR MACHINE (SVM):

Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for classification or regression problems. It use a technique called the kernel trick to transform the data. Based on these transformations, it finds an optimal boundary between the possible outputs. However, this project will focus on using SVM for classification. In particular it will be focusing on non-linear SVM, or SVM using a non-linear kernel. Non-linear SVM means that the boundary that the algorithm calculates doesn't have to be a straight line. The benefit is that it can capture much more complex relationships between the data points without having to perform difficult transformations
on the own. The downside is that the training time is much longer as it’s much more computationally intensive.

2.3 SOCIAL NETWORK:

Social network sites provides hundreds of thousands of users with a platform to interact, cooperate, create, and share information. This project focus on Twitter social media application. On twitter, users can acquire information about others by viewing their profile, status, and photo entries to enabling them for develop friendships. For this project, twitter is being use to gain information and dataset to analyse the sentiment.

2.4 TABLE OF LITERATURE REVIEW:

<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>TITLE</th>
<th>DESCRIPTION</th>
<th>METHODOLOGY</th>
</tr>
</thead>
</table>
2. Analysis of user reviews can provide substantive information for business processing.  
3. Provide a CNN-based sentiment classification approach that can be used in Android applications | 1. Natural language processing (NLP).  
2. Convolutional Neural Network (CNN) |
| 2. V. Subramaniyaswamy        |                                                                        |                                                                                                                                                  |                                                                           |
| 3. V. Vijayakumar             |                                                                        |                                                                                                                                                  |                                                                           |
| 4. Sangaiah Arun Kumar        |                                                                        |                                                                                                                                                  |                                                                           |
| 5. R. Logesh                  |                                                                        |                                                                                                                                                  |                                                                           |
| 6. A. Umamakeswari            |                                                                        |                                                                                                                                                  |                                                                           |
2. Implications for management, practice and research.  
3. Perform sentiment classification on movie reviews | 1. Lexicon-based methods  
2. Machine learning (traditional machine learning)  
3. Neural networks |
| 2. Suzana Ilicb               |                                                                        |                                                                                                                                                  |                                                                           |
| 3. Mathias Krausa             |                                                                        |                                                                                                                                                  |                                                                           |
| 4. Stefan Feuerriegel         |                                                                        |                                                                                                                                                  |                                                                           |
| 5. Helmut Prendingerb         |                                                                        |                                                                                                                                                  |                                                                           |
| 1. Van Gerven  | Editorial: Artificial Neural Networks as Models of Neural Information Processing. (2017) | 1. Explore the use of ANNs in the context of computational neuroscience from various perspectives. 2. ANNs have also been embraced as a new tool for understanding neural information processing in the brain. | 1. Artificial Neural network |
| 2. Marcel      |                                           |                                                                         |                             |
| 3. Sander Bohte|                                           |                                                                         |                             |

| 2. Katrin Ihler     |                                           |                                                                         |                             |
| 3. Julian Kniephoff |                                           |                                                                         |                             |
| 4. Oliver Vornberger|                                           |                                                                         |                             |

| 1. Feifei Kou  | Social Network Search Based on Semantic Analysis and Learning. (2016) | 1. Analyzes the characteristics of social network data. 2. Gives the main purpose of social network search. 3. Points out the significance of social network cross-modal search. 4. Overcome sparsity issue of short texts. | 1. Vector space model. 2. Latent Dirichlet Allocation (LDA) model. 3. deep neural network (AlexNet and VGG-Net) |
| 2. Junping Du   |                                           |                                                                         |                             |
| 3. Yijiang He   |                                           |                                                                         |                             |
| 4. Lingfei Ye   |                                           |                                                                         |                             |

<p>| 1. Preslav Nakov | Sentiment Analysis in Twitter. (2016) | 1. Estimating the prevalence of positive and negative tweets about a given topic. 2. Detecting whether a textual item (e.g., a product review, a blog post, an editorial, etc.) | 1. Data mining (iterative quantification method) 2. SentiWordNet |
| 2. Alan Ritter  |                                           |                                                                         |                             |
| 3. Sara Rosenthal |                                          |                                                                         |                             |
| 4. Fabrizio Sebastiani |                                        |                                                                         |                             |
| 5. Veselin Stoyanov |                                         |                                                                         |                             |</p>
<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
<th>Methods</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saif M. Mohammad, Xiaodan Zhu, Svetlana Kiritchenko, Joel Martin</td>
<td>Sentiment, emotion, purpose, and style in electoral tweets. (2015)</td>
<td>Twitter Search API, Amazon’s Mechanical Turk, CrowdFlow (a.k.a Figure Eight)</td>
<td>1. Predict emotion and purpose labels, respectively, in new unseen tweets. 2. Determining how public sentiment is shaped. 3. Detecting sentiment or emotion.</td>
</tr>
<tr>
<td>Kumar Ravi, Vadlamani Ravi</td>
<td>A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. (2015)</td>
<td>Natural Language Processing (information retrieval and information extraction), Data Mining</td>
<td>1. Share views across the world. 2. Classified existing studies on the basis of opinion mining tasks 3. Extract and analyze public mood and views.</td>
</tr>
</tbody>
</table>
CHAPTER 3:
METHODODOLOGY:

3.1 INTRODUCTION:

Methodology is an approach, process and step that being used during the implementation of this project. The selection for the most suitable methodology for the development of the project is very important as the side effect of choosing wrong methodology is chaotic enough because the project might will not be completed on the right schedule or fail.

3.2 FRAMEWORK:

1. Install Twitter Application
2. Get the Twitter API
3. Install Rapid Miner

- Windows 10
- Twitter Application
- Twitter API
- Rapid Miner
- Put the dataset from twitter API into Rapid Miner.
- Get dataset from twitter through Twitter API
3.2.1 Install Twitter Application:

- Install Twitter Application and register Twitter application to get our own credentials. Registering an application with Twitter is critical, as it is the only way to get authentication credentials. As soon as getting the credentials, start to write the code.

3.2.2 Get the Twitter API.

- To get the twitter API, twitter developer must be accessed and create the apps.

3.2.3 Install Rapid Miner.

- RapidMiner is a tools to collect and classify sentiment in social networks. For example, tweets in twitter.

3.2.4 Get dataset from twitter through Twitter API.

- Twitter API enable to manage Tweets, publish and curate Tweets, filter and search for Tweet topics or trends, and much more.

3.2.5 Put the dataset from twitter API into Rapid Miner.

- Dataset from text-based will put into Rapid Miner to classify sentiments. There are some steps in using Rapid Miner:
  2. Classify their Sentiment using the Classify Sentiment Operator.
  3. Assign the tweets to different categories using the Categorize Operator.
  4. Visualize our results and make them more consumable and understandable.
REFERENCES:


