3-2019

REVIEWS TO RATING CONVERSION AND ANALYSIS USING MACHINE LEARNING TECHNIQUES

Charitha Chanamolu
005768174@coyote.csusb.edu

Follow this and additional works at: https://scholarworks.lib.csusb.edu/etd
Part of the Computer and Systems Architecture Commons

Recommended Citation
Chanamolu, Charitha, "REVIEWS TO RATING CONVERSION AND ANALYSIS USING MACHINE LEARNING TECHNIQUES" (2019). Electronic Theses, Projects, and Dissertations. 792.
https://scholarworks.lib.csusb.edu/etd/792

This Project is brought to you for free and open access by the Office of Graduate Studies at CSUSB ScholarWorks. It has been accepted for inclusion in Electronic Theses, Projects, and Dissertations by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
REVIEWS TO RATING CONVERSION AND ANALYSIS USING MACHINE LEARNING TECHNIQUES

A Project
Presented to the
Faculty of
California State University,
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
in
Computer Science

by
Charitha Chanamolu
March 2019
REVIEWS TO RATING CONVERSION AND ANALYSIS USING MACHINE LEARNING TECHNIQUES

A Project
Presented to the
Faculty of
California State University,
San Bernardino

by
Charitha Chanamolu
March 2019
Approved by:

Dr. Owen J. Murphy, Advisor, Computer Science and Engineering

Dr. Ernesto Gomez, Committee Member

Dr. Yunfei Hou, Committee Member
ABSTRACT

With the advent of technology in recent years, people depend more on online reviews to purchase a product. It is hard to determine whether the product is good or bad from hundreds of mixed reviews. Also, it is very time-consuming to read many reviews. So, opinion mining of reviews is necessary.

The main aim of this project is to convert the reviews of a product into a rating and to evaluate the ratings using machine learning algorithms such as Naïve Bayes and Support Vector Machine. In the process of converting the reviews to a rating, score words are created using SentiWordNet and transformed into seven categories from highly positive to highly negative.
ACKNOWLEDGEMENTS

I would like to express my special thanks and sincere gratitude to my advisor and mentor Dr. Owen J. Murphy who supported me all these days in completing my project and who guided me in my academics as well. I would also like to thank my committee members Dr. Ernesto Gomez and Dr. Yunfei Hou for their valuable suggestions and support.

I would also like to thank my parents Mr. Narasimha Swamy Chanamolu and Mrs. Latha Vallabhaneni and my brother, Suraj Chanamolu for being a part of my journey, supporting and helping me to grow physically and mentally.
# TABLE OF CONTENTS

ABSTRACT .............................................................................................................................. iii

ACKNOWLEDGEMENTS ...................................................................................................... iv

LIST OF TABLES .................................................................................................................. vii

LIST OF FIGURES .............................................................................................................. viii

CHAPTER ONE: INTRODUCTION

Background .......................................................................................................................... 1

Existing System .................................................................................................................... 1

Disadvantages of the Existing System .............................................................................. 1

CHAPTER TWO: SYSTEM ANALYSIS

Proposed System ................................................................................................................ 2

System Requirement Specifications ..................................................................................... 4

Hardware Requirements ..................................................................................................... 4

Software Requirements ..................................................................................................... 4

Programming Used .......................................................................................................... 5

Graphical User Interface ................................................................................................... 5

Swing .................................................................................................................................... 5

SVM ..................................................................................................................................... 6

Naïve Bayes ......................................................................................................................... 7

CHAPTER THREE: SYSTEM DESIGN

UML Diagrams ................................................................................................................... 8

Use Case Diagram ............................................................................................................. 8

Sequence Diagram ........................................................................................................... 9

Data Flow Diagram .......................................................................................................... 10
LIST OF TABLES

Table 1. User Acceptance Testing................................................................. 26
LIST OF FIGURES

Figure 1. Hierarchy of Java Swing API............................................................5
Figure 2. Use Case Diagram for the User...........................................................8
Figure 3. Sequence Diagram ........................................................................9
Figure 4. Data Flow Diagram.........................................................................10
Figure 5. Activity Diagram...........................................................................11
Figure 6. Component Diagram.................................................................12
Figure 7. Login Screen ..............................................................................13
Figure 8. Home Page .................................................................................14
Figure 9. Load Dictionary .........................................................................15
Figure 10. SVM Sentiment Analysis for Bromwell High Series.............16
Figure 11. SVM Sentiment Analysis for Device iPhone 8+....................17
Figure 12. SVM Sentiment Analysis for Dark Knight Movie...............17
Figure 13. Naïve Bayes Sentiment Analysis for Device iPhone 8+........18
Figure 14. Naïve Bayes Sentiment Analysis for Bromwell High Series.....19
Figure 15. Naïve Bayes Sentiment Analysis for Dark Knight Movie......19
Figure 16. Confusion Matrix for a Review in the Dataset of Bromwell High...20
Figure 17. SVM Accuracy for Bromwell High Animated Series ..........21
Figure 18. SVM Accuracy for Dark Knight Movie.....................................22
Figure 19. SVM Accuracy for iPhone 8+....................................................22
Figure 20. Naïve Bayes Accuracy for the Movie Dark Knight...............23
Figure 21. Naïve Bayes Accuracy for the Bromwell High Series.............24
Figure 22. Naïve Bayes Accuracy for the Device iPhone 8+....................24
Figure 23. Correctly Classified Chart...
CHAPTER ONE
INTRODUCTION

Background

It is often hard for an individual to come to a conclusion after reading numerous reviews of a product from numerous websites. Sometimes people found the product to be useful while some did not. An individual evaluation of the reviews is required to grade it for a final decision.

Existing System

SentiWordNet is a lexical resource for sentiment analysis. It provides scores to the parts of text based on numbers, adverbs or adjectives. The main drawback with this existing system is that it may not always give a good result for the sentimental analysis.

Disadvantages of the Existing System

• It is difficult to find whether the outcome is positive or negative from the scores obtained from SentiWordNet.
• There is no performance evaluation in the existing system.
CHAPTER TWO
SYSTEM ANALYSIS

Proposed System

The goal of the project is to develop a GUI application where the user can view a final rating of a product, movie or series by selecting the dataset of reviews and running the application.

- This application starts with a Login screen where the user can enter his username and password. If any of the details are missing or incorrect a message box will be displayed.
- After logging in, the user will be able to find the project title at the top panel and seven buttons at the bottom panel namely Load Dictionary, SVM Sentiment Analysis, Naïve Bayes Sentiment Analysis, Confusion Matrix, SVM accuracy, Naïve Bayes accuracy, and Correctly Classified Chart.
- Load dictionary is the main button which contains stop word removal, porter stemmer algorithm, and SentiWordNet. The user should load the dictionary every time they log in.
- Support Vector Machine(SVM) sentiment analysis is used to convert the score obtained from SentiWordNet to a rating. The analysis is done by choosing the required dataset for converting reviews. A table displays on
the left panel having the dataset name, review of the product, SVM rating of positive, negative or neutral and the SentiWordNet score.

- Naïve Bayes sentiment analysis is used in converting the score obtained from SentiWordNet into a rating. Naïve Bayes Sentiment Analysis is also done by choosing the required dataset for converting reviews. A table displays on the right panel having the dataset name, review of the product, Naïve Bayes rating of positive, negative and neutral and the SentiWordNet score.

- Confusion matrix contains a table comparing SVM and Naïve Bayes ratings. Both SVM and Naïve Bayes divides the score obtained from SentiWordNet into seven categories from highly positive to highly negative.

- SVM accuracy is the ratio of total positive ratings obtained from the SVM sentiment analysis to the overall scores from the dataset.

- Naïve Bayes is the ratio of the total positive ratings obtained from the Naïve Bayes sentiment analysis to the total ratings from the dataset.

- The Correctly classified chart contains the graphical representation of SVM and Naïve Bayes accuracy.

- Logout terminates the operation from the user.
System Requirement Specifications

The end-user needs to possess the following hardware and software to run the application.

Hardware Requirements

- Laptop having Windows and Eclipse IDE
- RAM-8GB

Software Requirements

- Operating System- Windows 10
- Programming Language- Java
- Toolkit- Java Swing
Programming Used

**Graphical User Interface:**

In the Graphical User Interface, the user can interact graphically with the screen rather than with the text commands. AWT and Swing are the two sets of Java API for the graphical programming.

**Swing**

Swing is a part of Java Foundation Classes (JFC) software that implements a set of GUI components. The JFC components are lightweight, and the look and feel of JFC are the same on all platforms. List controls, labels, tree controls, buttons, and table controls are the components included in the swing tool kit.

![Hierarchy of Java Swing API](image)

Figure 1. Hierarchy of Java Swing API
Support Vector Machine (SVM) is used to classify the texts into either positive, negative or neutral. SVM has the distinct advantage of handling large texts. The individual scores obtained for each score word from SentiWordNet are categorized into seven different categories (strong-positive, strong-negative, positive, weak-positive, negative, weak-negative and neutral) using SVM.

- strong-positive(sp)=$score>0.4$
- positive(p)=$0.3<score<0.4$
- weak-positive(wp)=$0.2<score<0.3$
- strong-negative(sn)=$0.1<score<0.2$
- negative(n)=$0<score<0.1$
- weak-negative(wn)=$score<0$
- neutral(n)=0

Total positive score for a review $(t.p)=svm.sp+svm.p+svm.wp$

Total negative score for a review $(t.n)=svm.sn+svm.n+svm.wn$

SVM accuracy=number of positive reviews/(total number of reviews).
Naïve Bayes

Naïve Bayes Sentiment analysis is used to classify the texts into either positive, negative or neutral. Naïve Bayes has the distinct advantage of handling small texts. The individual scores obtained for each score word from SentiWordNet are categorized into seven different categories (strong-positive, strong-negative, positive, weak-positive, negative, weak-negative and neutral) using Naïve Bayes Sentiment analysis.

strong-positive(sp)=score>0.25
positive(p)=0.2<score<0.25
weak-positive(wp)=0.15<score<0.2
strong-negative(sn)=0.1<score<0.15
negative(n)=0<score<0.1
weak-negative(wn)=score<0
neutral(n)=0

Total positive score for a review(t.p)=nb.sp+nb.p+nb.wp
Total negative score for a review(t.n)= nb.sn+nb.n+nb.wn
Naïve Bayes accuracy=number of positive reviews/(total number of reviews)
CHAPTER THREE
SYSTEM DESIGN

UML Diagrams

Use Case Diagram

Use case diagram mainly shows the interactions between user and the system. It gives the clear overview of what steps are going in the system.

Figure 2. Use Case Diagram for the User
Sequence Diagram

A sequence diagram shows how processes operate with one another and in what order.

Figure 3. Sequence Diagram
Data Flow Diagram

A Data Flow Diagram gives the exact information on how the data flows in the system.

Figure 4. Data Flow Diagram
Activity Diagram

Activity Diagram is a flow chart to represent flow from one activity to another.

Figure 5. Activity Diagram
**Component Diagram**

The Component diagram is a special kind of UML diagram that describes the components that are used to form the system. It doesn’t give any information about the working of the system.

![Component Diagram]

Figure 6. Component Diagram
CHAPTER FOUR
SYSTEM SCREENSHOTS

System Working

Login Screen

At the Login Screen, the user can either login to the application or reset the username and password. A Message box displays if there is any incorrect username or password.

Figure 7. Login Screen
Home Page

This Homepage contains project title on the top panel. Eight buttons in the bottom panel are useful in converting the reviews to a rating.

![Opinion Mining](image)

Figure 8. Home Page
Load Dictionary

Load Dictionary loads stop word removal, the porter stemmer algorithm, and SentiWordNet. It displays a message box when the dictionary is loaded.

Figure 9. Load Dictionary
SVM Sentiment Analysis

The dataset for which the ratings should be calculated is chosen from the list of the datasets available. Here the dataset containing the reviews of “Bromwell High” which is an animated entertainment series is chosen for SVM Sentiment Analysis.

Figure 10. SVM Sentiment Analysis for Bromwell High Series
Figure 1. SVM Sentiment Analysis for Device iPhone 8+

Figure 2. SVM Sentiment Analysis for Dark Knight Movie
Naïve Bayes Sentiment Analysis

The dataset for which the ratings should be calculated is chosen from the list of datasets available. Here the dataset containing the reviews of iPhone 8+ is chosen for Naïve Bayes Sentiment Analysis.

Figure 13. Naïve Bayes Sentiment Analysis for Device iPhone 8+
Figure 14. Naïve Bayes Sentiment Analysis for Bromwell High Series

<table>
<thead>
<tr>
<th>File Name</th>
<th>Review</th>
<th>Opinion</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0_9.txt</td>
<td>bromwell high cartoon</td>
<td>Negative</td>
<td>0.2</td>
</tr>
<tr>
<td>10.9.txt</td>
<td>male women movies</td>
<td>Negative</td>
<td>1.4</td>
</tr>
<tr>
<td>11.9.txt</td>
<td>stanley irish acting</td>
<td>Positive</td>
<td>1.4</td>
</tr>
<tr>
<td>12.9.txt</td>
<td>stanley irish acting</td>
<td>Positive</td>
<td>1.4</td>
</tr>
<tr>
<td>13.7.txt</td>
<td>production quality</td>
<td>Negative</td>
<td>0.9</td>
</tr>
<tr>
<td>14.10.txt</td>
<td>film special place</td>
<td>Positive</td>
<td>2.4</td>
</tr>
<tr>
<td>15.7.txt</td>
<td>guess film magic</td>
<td>Negative</td>
<td>0.4</td>
</tr>
<tr>
<td>16.7.txt</td>
<td>found romance drama</td>
<td>Negative</td>
<td>3.2</td>
</tr>
<tr>
<td>17.9.txt</td>
<td>complex film explores</td>
<td>Negative</td>
<td>-2</td>
</tr>
<tr>
<td>18.7.txt</td>
<td>stanley irish heart</td>
<td>Positive</td>
<td>1.3</td>
</tr>
<tr>
<td>19.10.txt</td>
<td>read comments toremey</td>
<td>Negative</td>
<td>2.1</td>
</tr>
<tr>
<td>1.7.txt</td>
<td>adult comedy cartoons</td>
<td>Positive</td>
<td>0.5</td>
</tr>
<tr>
<td>20.9.txt</td>
<td>stanley irish show tri</td>
<td>Positive</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Figure 15. Naïve Bayes Sentiment Analysis for Dark Knight Movie.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Review</th>
<th>Opinion</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0_1.txt</td>
<td>pretty spot time</td>
<td>Positive</td>
<td>2.3187</td>
</tr>
<tr>
<td>0_10.txt</td>
<td>batman</td>
<td>Positive</td>
<td>0.0</td>
</tr>
<tr>
<td>0_11.txt</td>
<td>perfection movie</td>
<td>Positive</td>
<td>0.0</td>
</tr>
<tr>
<td>0_12.txt</td>
<td>brilliant haunting as</td>
<td>Positive</td>
<td>0.2015</td>
</tr>
<tr>
<td>0_13.txt</td>
<td>stupendous effort nec</td>
<td>Neutral</td>
<td>0.0</td>
</tr>
<tr>
<td>0_2.txt</td>
<td>is gonna start love n</td>
<td>Neutral</td>
<td>0.6767</td>
</tr>
<tr>
<td>0_3.txt</td>
<td>couldn't dark knight li</td>
<td>Negative</td>
<td>1.5013</td>
</tr>
<tr>
<td>0_4.txt</td>
<td>dark complex ambition</td>
<td>Neutral</td>
<td>1.8771</td>
</tr>
<tr>
<td>0_5.txt</td>
<td>dark knight years ago</td>
<td>Negative</td>
<td>-0.372</td>
</tr>
<tr>
<td>0_6.txt</td>
<td>leave theatre highly</td>
<td>Positive</td>
<td>5.6453</td>
</tr>
<tr>
<td>0_7.txt</td>
<td>honor watching tdk sc</td>
<td>Positive</td>
<td>4.7907</td>
</tr>
<tr>
<td>0_8.txt</td>
<td>comic movie</td>
<td>Positive</td>
<td>0.3333</td>
</tr>
<tr>
<td>0_9.txt</td>
<td>dark knight good easy</td>
<td>Positive</td>
<td>0.7856</td>
</tr>
</tbody>
</table>
Confusion Matrix

SVM and Naïve Bayes Sentiment Analysis divides the given review into seven categories from strong positive to strong negative. The Confusion matrix here is the comparison of ratings between SVM and Naïve Bayes for any dataset of reviews.

![Confusion Matrix for a Review in the Dataset of Bromwell High](image)

Figure 16. Confusion Matrix for a Review in the Dataset of Bromwell High
SVM Accuracy

SVM accuracy is the ratio of total positive ratings in the dataset of reviews to the total reviews. The below figure shows that there is only 10.8 percent of positive reviews for the Bromwell High animated series.

Figure 17. SVM Accuracy for Bromwell High Animated Series
Figure 18. SVM Accuracy for Dark Knight Movie

<table>
<thead>
<tr>
<th>File Name</th>
<th>Review</th>
<th>Opinion</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0_3_1.txt</td>
<td>pretty spot time rem... Positive</td>
<td>2.3187</td>
<td></td>
</tr>
<tr>
<td>0_3_10.txt</td>
<td>batman</td>
<td>Positive</td>
<td>0.0</td>
</tr>
<tr>
<td>0_3_11.txt</td>
<td>perfection movie</td>
<td>Positive</td>
<td>0.0</td>
</tr>
<tr>
<td>0_3_12.txt</td>
<td>brilliant haunting x</td>
<td>Positive</td>
<td>0.2815</td>
</tr>
<tr>
<td>0_3_13.txt</td>
<td>stupendous effort new... Neutral</td>
<td>Neutral</td>
<td>0.0</td>
</tr>
<tr>
<td>0_3_2.txt</td>
<td>im gonna start love m... Neutral</td>
<td>Neutral</td>
<td>0.6767</td>
</tr>
<tr>
<td>0_3_3.txt</td>
<td>couldn't dark knight l! Negativ</td>
<td>Negative</td>
<td>1.5013</td>
</tr>
<tr>
<td>0_3_4.txt</td>
<td>dark complex ambition... Neutral</td>
<td>Neutral</td>
<td>1.8771</td>
</tr>
<tr>
<td>0_3_5.txt</td>
<td>dark knight years ago... Negative</td>
<td>Negative</td>
<td>0.3742</td>
</tr>
<tr>
<td>0_3_6.txt</td>
<td>leave theatre highly... Positive</td>
<td>Positive</td>
<td>5.6453</td>
</tr>
<tr>
<td>0_3_7.txt</td>
<td>honor watching tdk so... Positive</td>
<td>Positive</td>
<td>4.7907</td>
</tr>
<tr>
<td>0_3_8.txt</td>
<td>comic movie</td>
<td>Positive</td>
<td>0.3333</td>
</tr>
<tr>
<td>0_3_9.txt</td>
<td>dark knight good eax</td>
<td>Positive</td>
<td>0.7856</td>
</tr>
</tbody>
</table>

Figure 19. SVM Accuracy for iPhone 8+

<table>
<thead>
<tr>
<th>File Name</th>
<th>Review</th>
<th>Opinion</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0_2_1.txt</td>
<td>a good phone package ... Positive</td>
<td>1.0631</td>
<td></td>
</tr>
<tr>
<td>0_2_10.txt</td>
<td>phone features good p... Positive</td>
<td>Positive</td>
<td>0.6337</td>
</tr>
<tr>
<td>0_2_2.txt</td>
<td>common belief modern ... Neutral</td>
<td>Neutral</td>
<td>-0.081</td>
</tr>
<tr>
<td>0_2_3.txt</td>
<td>dont buy phone iphone... Negative</td>
<td>Negative</td>
<td>-0.613</td>
</tr>
<tr>
<td>0_2_4.txt</td>
<td>phone android device ... Negative</td>
<td>Negative</td>
<td>-0.613</td>
</tr>
<tr>
<td>0_2_5.txt</td>
<td>had</td>
<td>Negative</td>
<td>-0.570</td>
</tr>
<tr>
<td>0_2_6.txt</td>
<td>dont buy phone iphone... Negative</td>
<td>Negative</td>
<td>-0.613</td>
</tr>
<tr>
<td>0_2_7.txt</td>
<td>phone price is way high... Negative</td>
<td>Negative</td>
<td>-0.0454</td>
</tr>
<tr>
<td>0_2_8.txt</td>
<td>phone hangs quick sud... Neutral</td>
<td>Neutral</td>
<td>0.0319</td>
</tr>
<tr>
<td>0_2_9.txt</td>
<td>phone features good t... Neutral</td>
<td>Neutral</td>
<td>0.7237</td>
</tr>
</tbody>
</table>
Naïve Bayes Accuracy

Naïve Bayes Accuracy is the ratio of total positive ratings in the dataset of reviews to the total reviews. The below figure shows that there are 84.61 percent positive reviews for the movie Dark Knight.

Figure 20. Naïve Bayes Accuracy for the Movie Dark Knight
Figure 21. Naïve Bayes Accuracy for the Bromwell High Series

Figure 22. Naïve Bayes Accuracy for the Device iPhone 8+
Correctly Classified Chart

The correctly classified chart contains graphical representation of SVM and Naïve Bayes accuracy. The below figure shows the Correctly Classified Chart for Bromwell high animated series.

Figure 23. Correctly Classified Chart
Testing has become a fundamental part of product development and equally important to application development. Testing is done to verify the quality of the product. It also decreases the maintenance cost of the product. A User acceptance test is conducted with nine test cases to ensure proper working of the application.

User Acceptance Testing

In this testing, all the components are equally verified. A total of nine test cases are verified at each level of the application starting from login page to the logout.

Test Cases

Table 1. User Acceptance Testing

<table>
<thead>
<tr>
<th>Test Case Id</th>
<th>Test Case Name</th>
<th>Test Case Disc.</th>
<th>Test Case Status</th>
<th>Test Steps</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Login</td>
<td>Verify either user data is uploaded or not</td>
<td>We cannot get further operations</td>
<td>Logging successfully</td>
<td>High</td>
</tr>
<tr>
<td>02</td>
<td>Load Dictionary</td>
<td>Verify the dictionary is loaded or not</td>
<td>We cannot get the parts of speech from words</td>
<td>Dictionary loaded</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>03</td>
<td>SVM Sentiment Analysis</td>
<td>Verify the dataset is loaded or not</td>
<td>If it's not loaded</td>
<td>We cannot discover the opinions</td>
<td>Displaying sentiment results for each review with scores</td>
</tr>
<tr>
<td>04</td>
<td>Naïve Bayes Sentiment Analysis</td>
<td>Verify the dataset is loaded or not</td>
<td>If it's not loaded</td>
<td>We cannot discover the opinions</td>
<td>Displaying sentiment results from Naïve Bayes algorithm</td>
</tr>
<tr>
<td>05</td>
<td>View Confusion Matrix</td>
<td>Verify any one row is selected or not</td>
<td>If it's not selected</td>
<td>We cannot apply the confusion matrix</td>
<td>Get the count of 7 types</td>
</tr>
<tr>
<td>06</td>
<td>SVM Accuracy</td>
<td>Verify any one row is selected or not</td>
<td>If it's not selected</td>
<td>We cannot apply the confusion matrix</td>
<td>Displays the SVM accuracy</td>
</tr>
<tr>
<td>07</td>
<td>Naïve Bayes Accuracy</td>
<td>Verify any one row is selected or not</td>
<td>If it's not selected</td>
<td>We cannot apply the confusion matrix</td>
<td>Displays the Naïve Bayes accuracy</td>
</tr>
<tr>
<td>08</td>
<td>Correctly Classified Chart</td>
<td>Verify the SVM &amp; Naïve Bayes datasets exist or not</td>
<td>If it's not existed</td>
<td>We cannot get the chart</td>
<td>Display the chart for no of positive reviews discovered from each algorithm</td>
</tr>
<tr>
<td>09</td>
<td>Logout</td>
<td>Verify all the operations are completed or not</td>
<td>If it's not complete</td>
<td>We cannot exit</td>
<td>Logout from the screen</td>
</tr>
</tbody>
</table>
CHAPTER SIX

CONCLUSION

This project will be useful for people in making decisions based on online reviews that are incorporated in the review section of websites.

The proposed approach has been tested successfully on the reviews for an animated entertainment series which has the least rating, a device which has a neutral rating and from a movie which has high ratings.

Future Enhancements

Context-Based Sentiment Analysis

SentiWordNet cannot identify the context of the sentence. SentiWordNet can misjudge a positive sentence as a negative or vice versa without considering the context. The accuracy of the current rating system can be improved by considering the context of the sentence.
APPENDIX A

APPLICATION CODE
package com;
import javax.swing.JFrame;
import javax.swing.JLabel;
import javax.swing.JTextField;
import javax.swing.JButton;
import javax.swing.JPanel;
import java.awt.event.ActionEvent;
import java.awt.event.ActionListener;
import javax.swing.UIManager;
import java.awt.BorderLayout;
import java.awt.Dimension;
import java.awt.Color;
import java.awt.Font;
import javax.swing.JPasswordField;
import javax.swing.JOptionPane;
import com.jd.swing.custom.component.panel.HeadingPanel;
import com.jd.swing.util.PanelType;
import com.jd.swing.util.Theme;
public class Login extends JFrame
{
    CUSTOMPANEL P1;
JLabel l1,l2;
    JTextField tf1,tf2;
    JButton b1,b2;
    Font f1;
    
    public Login(){
        super("Login ");
        p1 = new CustomPanel("LoginScreen");
        p1.setTitle("Login Screen");
        p1.setLayout(null);
        
        JPanel main = new
        HeadingPanel("","Theme.GLOSSY_METALIC_BLUE_THEME");
        main.setLayout(new BorderLayout());
        
        f1 = new Font("Microsoft Sanserif",Font.BOLD,13);
        JPanel pan1 = new
        HeadingPanel("","Theme.GLOSSY_METALIC_BLUE_THEME");
        l1 = new JLabel("Username");
        l1.setForeground(Color.white);
        l1.setFont(f1);
        pan1.add(l1);
tf1 = new JTextField(15);
tf1.setFont(f1);
pan1.add(tf1);

JPanel pan2 = new HeadingPanel("", Theme.GLOSSY_METALIC_BLUE_THEME);
l2 = new JLabel("Password");
l2.setForeground(Color.white);
l2.setFont(f1);
pan2.add(l2);

tf2 = new JPasswordField(15);
tf2.setFont(f1);
pan2.add(tf2);

JPanel pan3 = new HeadingPanel("", Theme.GLOSSY_METALIC_BLUE_THEME);
b1 = new JButton("Login");
b1.setFont(f1);
pan3.add(b1);

b1.addActionListener(new ActionListener(){
    public void actionPerformed(ActionEvent ae){
        login();
    }
});
b2 = new JButton("Reset");
b2.setFont(f1);
pan3.add(b2);
b2.addActionListener(new ActionListener(){
    public void actionPerformed(ActionEvent ae){
        tf1.setText(""");
        tf2.setText(""");
    }
});
main.add(pan1, BorderLayout.NORTH);
main.add(pan2, BorderLayout.CENTER);
main.add(pan3, BorderLayout.SOUTH);
main.setBounds(50,80,300,100);
p1.add(main);
getContentPane().add(p1, BorderLayout.CENTER);
}
public static void main(String a[]) throws Exception{
    UIManager.setLookAndFeel(UIManager.getSystemLookAndFeelClassName());
}
Login login = new Login();
login.setVisible(true);
login.setSize(400,300);
login.setLocationRelativeTo(null);
login.setResizable(false);
}
public void clear(){
    tf1.setText(""");
    tf2.setText(""");
}
public void login(){
    String user = tf1.getText();
    String pass = tf2.getText();
    if(user == null || user.trim().length() <= 0){
        JOptionPane.showMessageDialog(this,"Username must be entered");
        tf1.requestFocus();
        return;
    }
    if(pass == null || pass.trim().length() <= 0)
    {

JOptionPane.showMessageDialog(this, "Password must be entered");
    tf2.requestFocus();
    return;
}

try{
    if (user.equals("xxxx") && pass.equals("*****")){
        setVisible(false);
        UploadDocument ud = new UploadDocument(this);
        ud.setVisible(true);
        ud.setExtendedState(JFrame.MAXIMIZED_BOTH);
    }else{
        JOptionPane.showMessageDialog(this, "invalid user");
    }
}catch(Exception e){
    e.printStackTrace();
}
package com;
import javax.swing.JFrame;
import javax.swing.JPanel;
import javax.swing.JLabel;
import java.awt.BorderLayout;
import java.awt.Font;
import java.awt.Color;
import javax.swing.JButton;
import java.awt.event.ActionEvent;
import java.awt.event.ActionListener;
import java.io.BufferedReader;
import java.io.FileReader;
import java.io.File;
import java.util.ArrayList;
import javax.swing.JOptionPane;
import javax.swing.JScrollPane;
import javax.swing.JTable;
import javax.swing.table.DefaultTableModel;
import javax.swing.table.TableColumn;
import javax.swing.JFileChooser;
import java.awt.Cursor;
import com.jd.swing.custom.component.panel.HeadingPanel;
import com.jd.swing.util.PanelType;
import com.jd.swing.util.Theme;
import java.awt.Dimension;
import java.io.BufferedReader;
import java.awt.Cursor;
public class UploadDocument extends JFrame{
37

JLabel l1;
JPanel p1,p2,p3,p4;
Font f1;
JScrollPane jsp,jsp1;
Login;
JButton b1,b2,b3,b4,b5,b6,b7,b8;
FileChooser chooser;
DefaultTableModel dtm,dtm1;
JTable table,table1;
POSMModel model;
boolean loaded = false;
SentiWordNetDemoCode sentiwordnet;
SVMAnalysis svm;
NaiveBayes nb;
double svm_accuracy;
double nb_accuracy;
Porter stemmer = new Porter();
ArrayList<Count> svmcount = new ArrayList<Count>();
ArrayList<Count> nbcount = new ArrayList<Count>();

public UploadDocument(Login log){
    super("Opinion Mining");
    login = log;
    p1 = new HeadingPanel("Project Title",Theme.GLOSSY_METALIC_BLUE_THEME);
    p1.setPreferredSize(new Dimension(600,50));
    l1 = new JLabel("<html><body><center>A Novel approach for SentiWordNet for Reviews to Ratings Conversion<br/>using Web Data</center></body></html">".toUpperCase());
    l1.setFont(new Font("Times New Roman",Font.BOLD,18));
I1.setForeground(Color.white);
p1.add(I1);
getContentPane().add(p1, BorderLayout.NORTH);

f1 = new Font("Courier New", Font.BOLD, 14);

p2 = new JPanel();
p2.setLayout(new BorderLayout());
dtm = new DefaultTableModel()
{
    public boolean isCellEditable(int r, int c)
    {
        return false;
    }
}

table = new JTable(dtm);
table.setRowHeight(30);
table.setAutoResizeMode(JTable.AUTO_RESIZE_OFF);
table.setFont(f1);
table.getTableHeader().setFont(f1);
dtm.addColumn("File Name");
dtm.addColumn("Review");
dtm.addColumn("Opinion");
dtm.addColumn("Score");
table.getColumnModel().getColumn(0).setPreferredWidth(100);
table.getColumnModel().getColumn(1).setPreferredWidth(200);
table.getColumnModel().getColumn(2).setPreferredWidth(100);
table.getColumnModel().getColumn(3).setPreferredWidth(100);
jsp = new JScrollPane(table);
p2.add(jsp, BorderLayout.CENTER);
p4 = new JPanel();
    p4.setLayout(new BorderLayout());
    dtm1 = new DefaultTableModel()
    {
        public boolean isCellEditable(int r, int c){
            return false;
        }
    }
    table1 = new JTable(dtm1);
    table1.setRowHeight(30);
    table1.setAutoResizeMode(JTable.AUTO_RESIZE_OFF);
    table1.setFont(f1);
    table1.getTableHeader().setFont(f1);
    dtm1.addColumn("File Name");
    dtm1.addColumn("Review");
    dtm1.addColumn("Opinion");
    dtm1.addColumn("Score");
    table1.getColumnModel().getColumn(0).setPreferredWidth(100);
    table1.getColumnModel().getColumn(1).setPreferredWidth(200);
    table1.getColumnModel().getColumn(2).setPreferredWidth(100);
    table1.getColumnModel().getColumn(3).setPreferredWidth(100);
    jsp1 = new JScrollPane(table1);
    p4.add(jsp1, BorderLayout.CENTER);

    p3 = new HeadingPanel("", Theme.GLOSSY_METALIC_BLUE_THEME);
    p3.setPreferredSize(new Dimension(150,80));
    chooser = new JFileChooser(new File(".")));
chooser.setFileSelectionMode(JFileChooser.DIRECTORIES_ONLY);

b1 = new JButton("Load Dictionary");
b1.setFont(f1);
p3.add(b1);

b1.addActionListener(new ActionListener(){
    public void actionPerformed(ActionEvent ae){
        try{
            if(!loaded){
                Cursor hourglassCursor = new Cursor(Cursor.WAIT_CURSOR);
                setCursor(hourglassCursor);
                loadDictionary();
                svm = new SVMAnalysis();
                nb = new NaiveBayes();
                sentiwordnet = new SentiWordNetDemoCode("SentiWordNet_3.0.0_20130122.txt");
                Stopwords.readWords(stemmer);
                Cursor normalCursor = new Cursor(Cursor.DEFAULT_CURSOR);
                setCursor(normalCursor);
            }
        }catch(Exception e){
            e.printStackTrace();
        }
    }
}

JOptionPane.showMessageDialog(UploadDocument.this,"Dictionary Loaded");

b2 = new JButton("SVM Sentiment Analysis");
b2.setFont(f1);
p3.add(b2);

    b2.addActionListener(new ActionListener(){
      public void actionPerformed(ActionEvent ae){
        clearTable();
        int option = chooser.showOpenDialog(UploadDocument.this);
        if(option == chooser.APPROVE_OPTION){
          File = chooser.getSelectedFile();
          Cursor hourglassCursor = new Cursor(Cursor.WAIT_CURSOR);
          setCursor(hourglassCursor);
          svm(file);
          Cursor normalCursor = new Cursor(Cursor.DEFAULT_CURSOR);
          setCursor(normalCursor);
        }
      }
    });

b3 = new JButton("Naive Bayes Sentiment Analysis");
b3.setFont(f1);
p3.add(b3);
b3.addActionListener(new ActionListener(){
      public void actionPerformed(ActionEvent ae){
        clearTable1();
        int option = chooser.showOpenDialog(UploadDocument.this);
        if(option == chooser.APPROVE_OPTION){
          File = chooser.getSelectedFile();
          Cursor hourglassCursor = new Cursor(Cursor.WAIT_CURSOR);
          setCursor(hourglassCursor);
        }
      }
    });
b4 = new JButton("SVM Accuracy");
b4.setFont(f1);
p3.add(b4);
b4.addActionListener(new ActionListener(){
    public void actionPerformed(ActionEvent ae){
        double positive = 0;
        double negative = 0, neutral = 0;
        for (int i = 0; i < dtm.getRowCount(); i++) {
            String value = dtm.getValueAt(i, 2).toString().trim();
            if (value.equals("Positive"))
                positive = positive + 1;
            if (value.equals("Neutral"))
                neutral = neutral + 1;
        }
        svm_accuracy = (positive + neutral) / (double) dtm.getRowCount();
        JOptionPane.showMessageDialog(UploadDocument.this, "SVM Accuracy " + svm_accuracy);
    }
});

b5 = new JButton("Naive Bayes Accuracy");
b5.setFont(f1);
p3.add(b5);
b5.addActionListener(new ActionListener(){
    public void actionPerformed(ActionEvent ae){
        double positive = 0;
        double negative = 0, neutral = 0;
        for (int i = 0; i < dtm.getRowCount(); i++) {
            String value = dtm.getValueAt(i, 2).toString().trim();
            if (value.equals("Positive"))
                positive = positive + 1;
            if (value.equals("Neutral"))
                neutral = neutral + 1;
        }
        svm_accuracy = (positive + neutral) / (double) dtm.getRowCount();
        JOptionPane.showMessageDialog(UploadDocument.this, "SVM Accuracy " + svm_accuracy);
    }
});
public void clearTable1()
{
    for(int i=table1.getRowCount()-1;i>=0;i--){
        dtm1.removeRow(i);
    }
}
public void svm(File file){
    try{
        svmcount.clear();
        String result = "none";
        File file_list[] = file.listFiles();
        for(int i=0;i<file_list.length;i++){
            File fname = file_list[i];
            String review =
                svm.analyze(fname.getPath(),model,sentiwordnet.stemmer);
            Count c = new
            Count(review,svm.neutral,svm.wp,svm.p,svm.sp,svm.n,svm.wn,svm.sn,svm.score);
            svmcount.add(c);
            int positive = svm.wp+svm.p+svm.sp;
            int negative = svm.n+svm.wn+svm.sn;
            int neutral = svm.neutral;
            if(positive > negative && positive > neutral)
                result = "Positive";
            else if(negative > positive && negative > neutral)
                result = "Negative";
            else if(neutral > positive && neutral > negative)
                result = "Neutral";
            Object row[] = {fname.getName(),review,result,svm.score};
            dtm.addRow(row);
        }
    }
}catch(Exception e){
    e.printStackTrace();
}

public void naiveBayes(File file){
    try{
        nbcount.clear();
        String result = "none";
        File file_list[] = file.listFiles();
        for(int i=0;i<file_list.length;i++){
            File fname = file_list[i];
            String review = nb.analyze(fname.getPath(),model,sentiwordnet,stemmer);
            Count c = new Count(review,nb.neutral,nb.wp,nb.p,nb.sp,nb.n,nb.wn,nb.sn,nb.score);
            nbcount.add(c);
            int positive = nb.wp+nb.p+nb.sp;
            int negative = nb.n+nb.wn+nb.sn;
            int neutral = nb.neutral;
            if(positive > negative && positive > neutral)
                result = "Positive";
            else if(negative > positive && negative > neutral)
                result = "Negative";
            else if(neutral > positive && neutral > negative)
                result = "Neutral";
            Object row[] = {fname.getName(),review,result,nb.score};
            dtm1.addRow(row);
        }
    }
catch(Exception e){
    e.printStackTrace();
}

public void loadDictionary(){
    Runnable r = new Runnable(){
        public void run(){
            try{
                model = new POSModelLoader().load(new File("en-pos-maxent.bin"));
            }catch(Exception e){
                e.printStackTrace();
            }
        }
    }
    new Thread(r).start();
}
REFERENCES


