**MMAI 891**

**Natural Language Processing**

**Dr. Stephen W. Thomas**

**Individual Assignment 1**

**June 28, 2020**

**He Li**

# Sentiment Analysis via Lexicon-Based Approach

## Tasks

1. [Text] Given the accuracy and F1-score of your model, are you satisfied with the results? Explain.
   * **Model Accuracy: 79.85%**
   * **Model F1: 80.77%**
   * **I’m satisfied with the results above because it is much better than guessing every sentence in the test set as positive polarity (54.40%).**
2. [Text] Show five example instances in which your model was incorrect. Describe why the model was wrong.

|  |  |  |
| --- | --- | --- |
| **Examples of Incorrect Instances** | **Mistake Type** | **Reasoning** |
| It's a feel-good film and that's how I felt when I came out of the cinema! | FN | * Failed to extract “good” from “feel-good” * all neutral sentiments got hard-coded to polarity of 0 |
| The directing is sloppy at best. | FP | * VADER categorized “best” as positive sentiment |
| Not too screamy not to masculine but just right. | FN | * 2 “not”s and 1 “right” * all neutral sentiments got hard-coded to polarity of 0 |
| The least said about the acting the better. | FP | * VADER categorized “better” as positive sentiment |
| Still, it was the SETS that got a big "10" on my "oy-vey" scale. | FN | * VADER didn’t understand “10 on my … scale” as positive sentiment * all neutral sentiments got hard-coded to polarity of 0 |

# Sentiment Analysis via ML-Based Approach

## Tasks

* [Text] Given the accuracy and F1-score of your model, are you satisfied with the results? Explain.
* **Model Accuracy: 80.04%**
* **Model F1: 80.71%**
* **Because my ML-Based Approach includes VADER sentiments as one of the features in addition to TF-IDF n-gram BOW, I was expecting better performance than Question 1. Since the performance did not get better, I was not satisfied.**
* [Text] Show five example instances in which your model was incorrect. Describe why the model was wrong.

|  |  |  |
| --- | --- | --- |
| **Examples of Incorrect Instances** | **Mistake Type** | **Reasoning** |
| I'll give this film 10 out of 10! | FN | * VADER do not understand “10 out of 10” as positive * Not enough training data to make TF-IDF n-gram BOW useful |
| What this film lacks is a convincing script. | FP | * VADER miss-categorization * Not enough training data to make TF-IDF n-gram BOW useful |
| See it. | FN | * Not enough training data to make TF-IDF n-gram BOW useful |
| i wouldnt see this movie again for free. | FP | * VADER miss-categorization * Not enough training data to make TF-IDF n-gram BOW useful |
| I advise you to look out for it. | FN | * Not enough training data to make TF-IDF n-gram BOW useful |

* [Text] Compare and contrast the performance of the lexicon-based approach from Q1 with the ML-based approach here.
  + **When comparing the performances between Q1 and Q2. I found them to be extremely similar. This mean VADER did most of the heavy lifting and TF-IDF n-gram BOW didn’t help much likely because the training data was too small for TF-IDF n-gram BOW to provide any meaningful value to the RandomForestClassifier.**

# (Optional) Sentiment Analysis via Deep ML-Based Approach

As this is an optional bonus question, a maximum of 10% will be added to your mark.

## Tasks

1. [Text] Given the accuracy and F1-score of your model, are you satisfied with the results? Explain.
   * **Model Accuracy: 91.21%**
   * **Model F1: 91.58%**
   * **I’m very satisfied with the results, since I didn’t do any manual pre-processing and after only 1 epoch. The model outperformed Q1 and Q2**
2. [Text] Show five example instances in which your model was incorrect. Describe why the model was wrong.

|  |  |  |
| --- | --- | --- |
| **Examples of Incorrect Instances** | **Mistake Type** | **Reasoning** |
| The last 15 minutes of movie are also not bad as well. | FN | * “not bad as well” not understood as good * Not enough epochs trained (1 so far) |
| Lame would be the best way to describe it. | FP | * sarcasm * Not enough epochs trained (1 so far) |
| The soundtrack wasn't terrible, either. | FN | * Double negative * Not enough epochs trained (1 so far) |
| You can find better movies at youtube. | FP | * sarcasm * Not enough epochs trained (1 so far) |
| I advise you to look out for it. | FN | * Confusing sentiment even from human eye * Not enough epochs trained (1 so far) |

1. [Text] Compare and contrast the performance of this approach(Q3) with the 2 other approaches above (Q1 and Q2)
   * **This transformer model out-performed Q1 and Q2 and managed to reach an amazing F1 and Accuracy or 91%+. It requires no manual preprocessing, and only trained for 1 epoch with GPU being used. Very impressive!**