**Q1. Human Ratings Task:**

**a) Get 3 classmates (opinion holders) to write three different opinions about their phone.**

**b) Get 3 different people (raters) to rate these comments as positive, negative, neutral or can’t-say**

**c) Take this 3 x 3 matrix and find the inter-rater reliability between your 3 raters using Kappa**

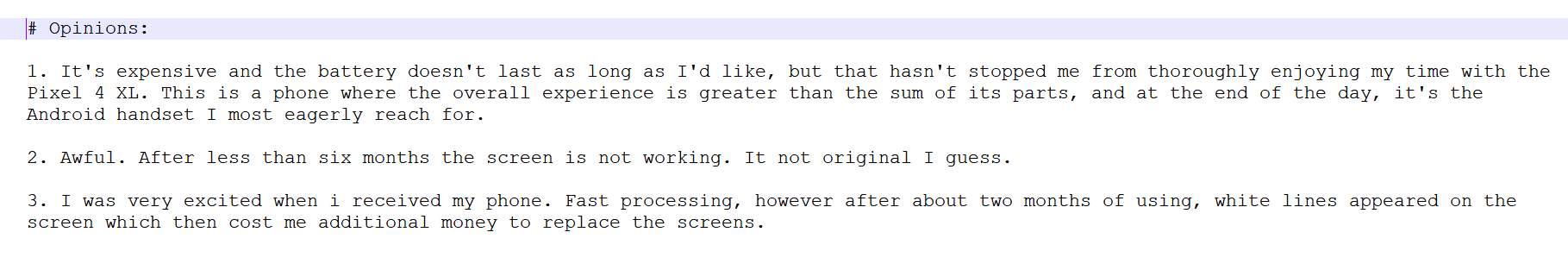
**d) If you wanted to get the correlation between raters (using Pearson’s rho) what would you do? Then do this.**

**Ans1a.**

**Opinions:** Opinion is defined as a judgement which is not conclusive. Opinion holder is one who expresses their opinions on a subject.

3 opinion holders with 3 different opinions about their phone are considered.

Opinions are as follows:

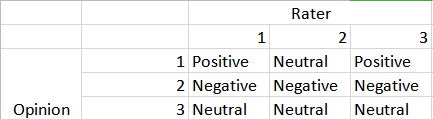


**Ans1b.**

**Rater:** One who determines a rating for a particular subject is called a Rater.

Possible rating values can be (Positive, Negative , Neutral).

Rating for the 3 different opinions by 3 different raters are considered. Ratings are as follows:

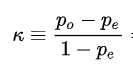


**Ans1c.**

**Inter-rater Reliability:** It is defined as the extent of agreement among the raters. It is the score of how much consensus exists between the ratings given by different raters. High value of Inter- reliability infers to high degree of agreement among the raters. It can be evaluated by different statistics such as Kappa, Pearson, Percentage agreement etc.

**Cohen’s Kappa Coefficient (k):** It is a statistic used to calculate the inter-rater reliability for categorical values.

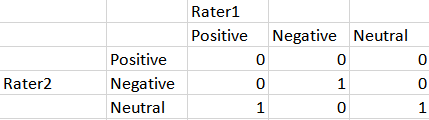
The coefficient k considers the agreement occurring by choice. K value always less than or equal to 1. A value 1 depicts perfect agreement.

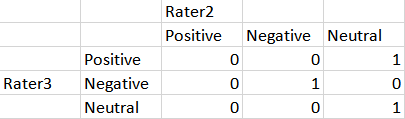
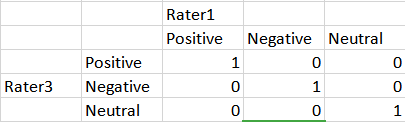


where

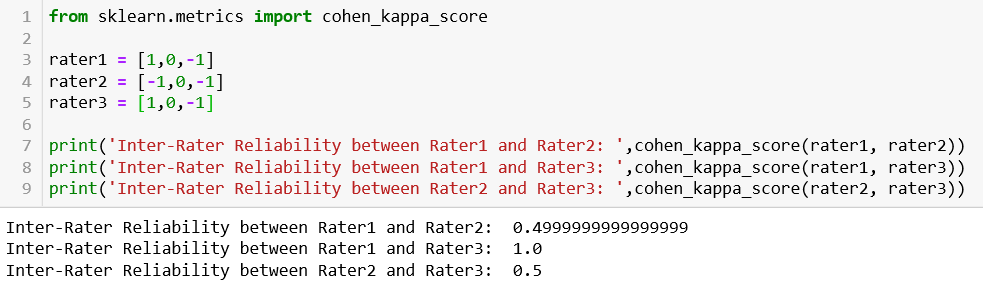
po: Relative agreement among raters.

pe: Probability of chance agreement.



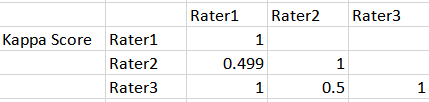
Using Cohen’s Kappa, the inter-rater reliability among 3 raters is shown below:



Cohen’s Kappa Coefficient,k is calculated for Rater 1,2 and 3.

The inter-rater reliability among Rater 1 and Rater2 comes out to be 0.4999 which infers that Rater 1 and rater2 are slightly agreeing on the ratings. This is due to their common Negative and Neutral rating for opinion number 2 and 3.

Similar case with Rater 2 and 3. Also, the inter-rater reliability among Rater 1 and Rater 3 comes out to be 1.0 which infers that both the raters are having perfect agreeing on the ratings. This is due to their similar rating for all three opinions.



**Ans1d.**

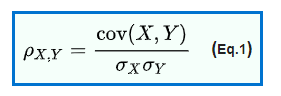
**Pearson’s rho:** It is defined as the measure of linear correlation among 2 variables. It’s value lies in between -1 and +1.

+1 infers total positive linear correlation among raters in this case.

-1 infers total negative linear correlation among raters in this case.

0 infers zero linear correlation among raters in this case.

It can be calculated as:



where,

cov(X,Y): Co variance between X and Y.

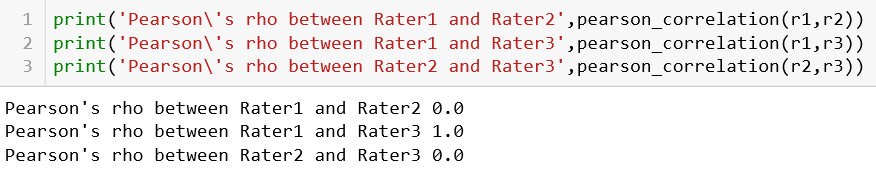
σX : Standard Deviation of X

σY : Standard Deviation of Y

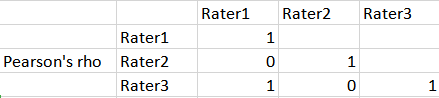
The correlation among raters is calculated using the Pearson’s:

1. Mean of ratings from both raters is calculated.
2. From every rating, mean is subtracted.
3. Summation of product of every element is divided by root of product of summation of square of entity 1 and entity 2.

This gives the Pearson’s rho among two raters.



Pearson’s rho among all the raters is shown below:



The correlation among Rater 1 and Rater2 comes out to be 0 which infers that Rater 1 and rater2 are not correlated to each other in terms of ratings. This is due to their different rating for all the opinions.

Similar case with Rater 2 and 3. But, correlation among Rater 1 and Rater 3 comes out to be 1.0 which infers that both the raters are highly correlated to each other on the basis of ratings. This is due to their similar rating for all three opinions.

==================================================================================================

**Q2. Do, some searches and find 3 sentiment lists that are commonly used in previous research. For 2 of these lists, select 10 positive and 10 negative words (randomly). Evaluate each word, discussing whether it is really positive/negative; for each one try to find a sentential context in which it might be interpreted with the opposite valence.**

**Ans2.**

**Sentiment:** It is defined as the emotion towards something.  
Following 3 Sentiment lists are considered:

1. SentiWordNet ***[1]***
2. MultiPurpose Question Answering MPQ Sentiment List ***[2]***
3. Opinion-Lexicon Sentiment List ***[3]***

20 random positive and negative words from MPQ list are considered and shown below:

P -> Positive and N->Negative

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr. No | Entry | P | N | Opposite Valence |
| 1 | Abnormal (exceptional) | ✓ |  | **N:** The people around him got disturbed by his abnormal(undesirable) behavior.  **P:** The company made abnormal profits this time. |
| 2 | Alarm (alert) | ✓ |  | **N:** She hesitated, not wanting to alarm(feeling of fear)her friend with the details.  **P:** The loud fire alarm alarmed the police. |
| 3 | Bait (lure) | ✓ |  | **N:** The boys revelled in baiting(insult) him about his love of stories.  **P:** Best stuff placed closest to sidewalk,served as a bait to the customers. |
| 4 | Bent (determined) | ✓ |  | **N:** This is as bad as a bent(corrupt) cop forging evidence to put a real criminal away.  **P:** An army man bent on saving souls. |
| 5 | Bicker (flash) | ✓ |  | **N:** Individual directors may disagree, bicker(argue over trivial mattters) the chairman.  **P:** The bicker of the headlights saved them from robbery at night. |
| 6 | Blah (ordinary) | ✓ |  | **N:** Having fun is the best way to banish the blahs(depression) and reap physical benefits  **P:** I am not sure if I would rather have an amazing year, or a blah year like 2018. |
| 7 | Cancer (constellation) | ✓ |  | **N:** Some of those volunteers will develop cancer(disease), heart disease or mental illness.  **P:** A cancer in the sky is said to represent crab and it’s crazy. |
| 8 | Bomb (looks hot) | ✓ |  | **N:** The channel has produced a string of critical and commercial bombs (event that fails badly).  **P:** He is simply ‘Bomb’ where ladies are concerned. |
| 9 | fall (season) | ✓ |  | **N:** She fall into a trap.  **P:** In October, comes the beautiful fall of snow. |
| 10 | Bunk (bed built over another) | ✓ |  | **N:** Anyone with a brain cell would never believe such bunk(non sense).  **P:** Bunks they used to take during their school time made memories. |
| 11 | Swagger(arrogant manner) |  | ✓ | **P:** I will take you somewhere swagger(smart and stylish).  **N:** He swaggered along the corridor. |
| 12 | Wild (angry) |  | ✓ | **P:** I'm not wild (enthusiastic) about the music.  **N:** She was wild. She just flipped. It was as if she had voices in her head. |
| 13 | Overshadow (dominate) |  | ✓ | **P:** He always overshadowed(more impressive by comparison) his brilliant elder brother.  **N:** The tragedy overshadowed the couple's happiness. |
| 14 | Wicked (unpleasant) |  | ✓ | **P:** Sophie makes wicked(wonderful) cakes.  **N:** Wicked weather passing by is responsible for scattered power outages. |
| 15 | Abrasive (harsh) |  | ✓ | **P:** The wood should be rubbed down with fine abrasive(capable of polishing hard surface) paper.  **N:** Her abrasive nature won her no friends. |
| 16 | Awe (feeling of fear) |  | ✓ | **P:** The famous professor awed(inspired) the undergraduates.  **N:** He stared over the edge with a feeling of awe. |
| 17 | Brisk (sharp) |  | ✓ | **P:** She adopted a brisk(quick), businesslike tone that helped her to survive in IT industry.  **N:** His brisk walk made him to fall. |
| 18 | Dote (foolish due to old age) |  | ✓ | **P:** Grandmother dotes(shower love) on her the twins.  **N:** She is now old and dotes. |
| 19 | Free (completely lacking) |  | ✓ | **P:** The researchers set the birds free(no longer confined).  **N:** The free cash backs made the companies to lose heavily |
| 20 | Keen (sharp) |  | ✓ | **P:** I have a keen(highly developed) eyesight.  **N:** A keen blade went through his arm and injured him. |

There are few words which were really negative in context but given in positive list like Abnormal. Similar case is with positive words but they were present in Negative lists. For example, awe which means to admire, inspire.

20 random positive and negative words from Opinion-Lexicon Sentiment list are considered and shown below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr. No | Entry | P | N | Opposite Valence |
| 1 | Fiery (bright color of fire) | ✓ |  | **N:** He gave a fiery(aggressive) speech.  **P:** The fiery red color of the car made it looked amazing. |
| 2 | Obsession (passion for something) | ✓ |  | **N:** He was in the grip of an obsession(unhealthy preoccupation) he was powerless to resist.  **P:** Her obsession for her career made her to succeed. |
| 3 | Prefer (like better) | ✓ |  | **N:** The police will prefer charges.  **P:** I prefer Paris over Dublin. |
| 4 | Enough (as much as required) | ✓ |  | **N:** I've had enough(unwilling to tolerate) of this arguing.  **P:** I have taken enough food for me. |
| 5 | Fine (high quality) | ✓ |  | **N:** A fine(thin,narrow) needle can cause pain.  **P:** This was a fine piece of art. |
| 6 | Hale (strong and healthy) | ✓ |  | **N:** He haled(dragged) an old man out of the audience.  **P:** He's in his sixties,bur very hale and hearty. |
| 7 | Sweet (pleasant in taste) | ✓ |  | **N:** It was far too sweet(less pleasant) and had a bitter aftertaste.  **P:** She came with a bunch of sweet-scented flowers. |
| 8 | Appropriate(apt) | ✓ |  | **N:** The accused had appropriated(take for one’s use without owner’s permission) the  property.  **P:** The title of the story is appropriate according to its content. |
| 9 | Daring (courage) | ✓ |  | **N:** The daring(over adventurous) of the players ruined the game.  **P:** The daring of the players brings fortune. |
| 10 | Easy (relaxed) | ✓ |  | **N:** As a taxi driver he was an easy(vulnerable) target.  **P:** It was an easy victory over the opponent. |
| 11 | Giddy (Disorienting) |  | ✓ | **P:** He was a charmingly giddy(excitable) young man.  **N:** We will wait to see whether he ever rises to those giddy heights. |
| 12 | Throttle (Attack by choking) |  | ✓ | **P:** Liquid fuel engines can be throttled(control) up and down during a flight.  **N:** She was sorely tempted to throttle him. |
| 13 | Taut (tense) |  | ✓ | **P:** A taut text(concise) of only a hundred pages helped to predict the judgement.  **N:** Her nerves were taut as the strings of a bow. |
| 14 | Tank (jail, cell) |  | ✓ | **P:** A huge tank(storage chamber) is used to store either liquid or gas.  **N:** He drove me straight to the tank. |
| 15 | Sick (Disappointment) |  | ✓ | **P:** The company organized sick funds for its ill workers so as to help them in illness.  **N:** He had a sick fear of returning. |
| 16 | Ferocious (Violent) |  | ✓ | **P:** A ferocious(great) headache helped him to know about his tumor.  **N:** The animal is a most ferocious looking one. |
| 17 | Suffer (appear worse) |  | ✓ | **P:** Faiz will no longer suffer the trauma.  **N:** Her daughter suffered from a prolonged illness. |
| 18 | Suspect (Doubtful) |  | ✓ | **P:** A suspect package on the platform increased the awareness among the crowd.  **N:** He suspected the gas leak, hence saved the house from fire. |
| 19 | Steal (Dishonesty) |  | ✓ | **P:** He stole the ball from Kay to run on and score his seventh League goal.  **N:** She was found guilty of stealing from the shop. |
| 20 | Crazy (Mad) |  | ✓ | **P:** She is crazily gorgeous.  **N:** It was crazy to hope that good might come out of this mess. |

There are few words which were really negative but given in positive list like Obsession. Similar case with positive words but they are present in Negative lists. For example, taut which means tense.

==================================================================================================

**Q3. Bromberg’s Sentiment Program:**

**Have a look at the simple program that does sentiment analysis. So, take a look at the program and see what is happening in the different variables, but adding print statements on its variables.**

1. **Now consider ways to improve the training. Eg if you removed stopwords from the inputs what do you think might happen?**
2. **Implement this or another solution in the program and report what happens to the precision and recall of the classifier.**

**Ans3.** Bromberg’s Sentiment Analysis program: ***[4]***

* It starts with reading the two input files ‘rt-polarity-pos.txt’ and ‘rt-polarity-neg.txt’.
* The block evaluate\_features () splits up the reviews by line and then builds a posFeatures variable that contains the output of feature selection mechanism with tags ‘pos’ or ‘neg’ appended to it, depending on whether the review it is drawing from is positive or negative.
* Further the code separates the data into train and test data for a Naive Bayes classifier that uses 3/4 th of the data as training data and 1/4th data as test data.
* It then trains the model using Naive Bayes Classifier.
* Lastly, function checks how well the classifier does when it classifies the test data.
* referenceSets contains the actual values for the testing data and testSets contains the predicted output.
* For each one of the testFeatures, loop through three things: an arbitrary ‘i’, as an identifier, and then the words in the review, and the actual label (‘pos’ or ‘neg’), Then ‘i’ (the unique identifier) is added to the correct bin in referenceSets.
* At last,prediction of the label based on the features using the trained classifier and put the unique identifier in the predicted bin in testSets is performed.
* Accuracy,Precision for both positive,negative words is displayed in the end.

**Ans3a.**

The different ways to improve the training includes:

1. Pre processing (Removal of stopwords)
2. Trying with different classifier.
3. Consider whole sentence at once instead of words.
4. Adding different feature selection mechanism.

The accuracy of this program after removing the stopwords gets decreased slightly compared to that without removal of stopwords because in semantic analysis, the context is the prime priority.So if we remove the stopwords, context changes and hence the accuracy decreases. The precision for positive decreases whereas recall gets increased.

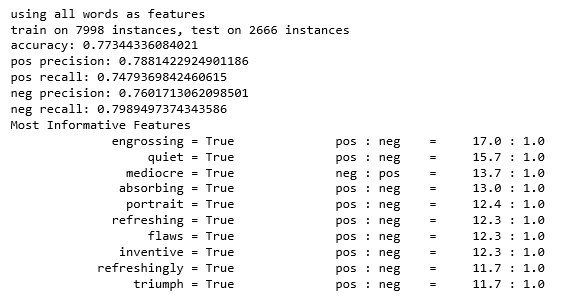
Similarly for the negative, precision increases with minute difference but the recall gets decreased.

If the classifier is changed, then also the accuracy gets decreased.

**Ans3b.**

**Implementation using Stopword removal, Changing the classifier to Support Vector Machine, different feature selection mechanism and Changing the training set.**

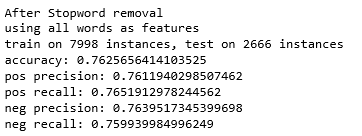
Output of Bromberg’s Sentiment Analysis program:



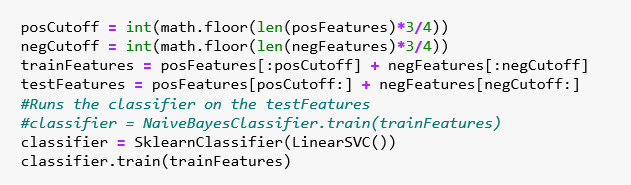
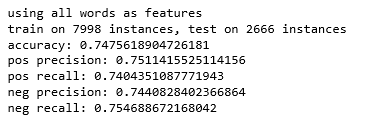
1. The accuracy of this program after removing the stopwords gets decreased slightly compared to that without removal of stopwords because in semantic analysis, the context is the prime priority.So if we remove the stopwords, context changes and hence the accuracy decreases. The precision for positive decreases whereas recall gets increased.

Similarly for the, precision increases with minute difference but the recall gets decreased.

Output after removing the stopwords:

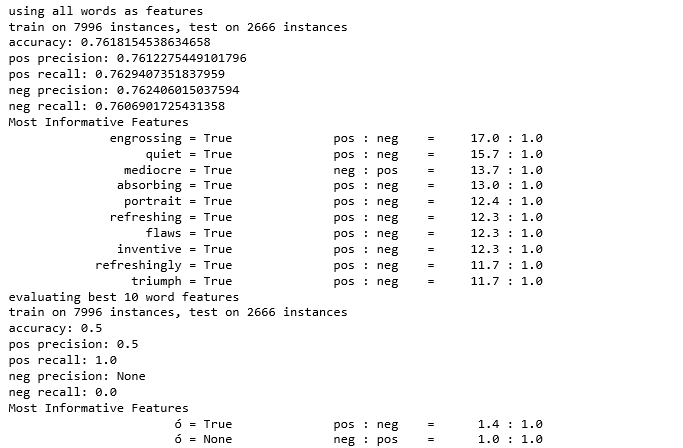
1. The Sentiment Analysis program is then tried using Support Vector Machine but still the accuracy of the model is low as compared to that of Naive Bayes Classifier.The precision,recall for positive gets decreases.Also precision and recall for the negative gets decreased.

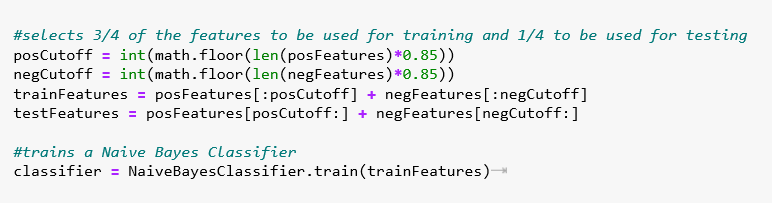
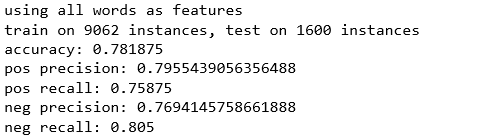
1. The Sentiment Analysis program is then tried using the different feature selection(10,100 and so on):

Using 10 features, Precision for positive becomes half of the original and recall becomes 1.0 whereas the precision and recall for negative goes down to 0.

==================================================================================================



1. If we increase the training set from 75% to 85%, accuracy gets increased to 78.12% with increase in precision and decrease in recall in case of positive. For the negative, precision slightly increases but recall goes to 80%.

**REFERENCES**

1. A. Eusli, "aesuli/SentiWordNet", GitHub, 2019. [Online]. Available: <https://github.com/aesuli/sentiwordnet> [Accessed: 17- Nov- 2019]
2. "Subjectivity Lexicon | MPQA", Mpqa.cs.pitt.edu, 2019. [Online]. Available: <http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/> [Accessed: 17- Nov- 2019]
3. B. Liu, "Opinion Mining, Sentiment Analysis, Opinion Extraction", Cs.uic.edu, 2019. [Online]. Available: <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon> [Accessed: 17- Nov- 2019]

**[4]** A. Bromberg, "Second Try: Sentiment Analysis in Python : Andy Bromberg", Andybromberg.com, 2019. [Online]. Available: <https://andybromberg.com/sentiment-analysis-python/> [Accessed: 17- Nov- 2019]