Q1: Human Ratings Task:

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

Classmate A's comments on "LG G8": The phone itself, feels great in the hands and is overall a wonderful device. The sound from the display is good, but not as clear as a regular speaker when turned up to higher volume, setting the phone down on a flat surface does help to boost the volume a little. The camera does lag slightly and the quality isn't as good as other high-end phones, but it is not bad by any means. The volume and quality of the auxiliary are actually better than I expected to come from a previous Galaxy S/N series owner. Call quality, is pretty good, but not top-notch (this may vary with carrier and coverage). The display, is very good. I enjoy the phone a lot, and the haptic feedback is really good on this device, better than any phone I have ever used. The battery life, is absolutely better than my old phone (Galaxy Note 8). I had to keep my Samsung phone on charge or battery saver to keep it from dying halfway through the day or a little after half, but this phone barely gets to 70 with moderate to heavy use. With light to moderate use, I can still have upwards of 80 percent battery percentage left.

Classmate B's comments on "Sony XZ2": I used to use iPhone X for a year. When I used my new Sont XZ2 phone to take photos compared to the old iPhone X, I found that it looks great on the phone but the pictures on a desktop and big screen the Apple X is better. Plus when I text the image stays it stays much clearer iPhone to iPhone vs. Sony to iPhone. Simply not worth the cost when new iPhones on the horizon.

Classmate C's comments on "Samsung Galaxy Note9": I love all the features and whatnot. I love my ability to do basically anything with this phone. But I cannot get past the size. This phone is monstrous. I can't hold it. I'm worried about when it starts to get warmer and I want to go hiking or jogging... this thing is a whole tablet! Although I'm already a little disappointed about the battery life. It's not necessarily bad, but it's not as good as they advertised. It lasts about the same as my coworker's iPhone 8 plus.

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

Rater-1: Classmate A: positive
Classmate B: neutral
Classmate C: negative
Rater-2: Classmate A: positive
Classmate B: neutral
Classmate C: positive
Rater-3: Classmate A: neutral
Classmate B: neutral
Classmate C: positive

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

Cohen's kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories. The definition of this is:

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

where po is the relative observed agreement among raters (identical to accuracy), and pe is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly seeing each category.

		Rater-1	Rater-2	Rater-3
Classmate A	positive	1	1	0
Classmate B	positive	0	0	0
Classmate C	positive	0	1	1
Classmate A	neutral	0	0	1
Classmate B	neutral	1	1	1
Classmate C	neutral	0	0	0
Classmate A	negative	0	0	0
Classmate B	negative	0	0	0
Classmate C	negative	1	0	0

		Rater-1	Rater-2	Rater-3
Classmate A		positive	positive	neutral
Classmate B		neutral	neutral	neutral
Classmate C		negative	positive	positive

Rater-1 X Rater-2:

			Rater-2		
		positive	neutral	negative	total
Rater-1	positive	1	0	0	1
	neutral	0	1	0	1
	negative	1	0	0	1
	total	2	1	1	3
		positive	neutral	negative	total
	Agreements	1	1	0	2
	Chance	0.5	0.5	0	1
	Kappa	(2-1)/(3-1)			
	Kappa	0.5			

Rater-1 X Rater-3:

			Rater-3		
		positive	neutral	negative	total
Rater-1	positive	0	1	0	1
	neutral	0	1	0	1
	negative	1	0	0	1
	total	1	2	0	3
		positive	neutral	negative	total
	Agreements	0	1	0	1
	Chance	0	1	0	1
	Kappa	(1-1)/(3-1)			
	Kappa	0			

Rater-2 X Rater-3:

			Rater-3		
		positive	neutral	negative	total
Rater-2	positive	1	1	0	2
	neutral	0	1	0	1
	negative	0	0	0	0
	total	1	2	0	3
		positive	neutral	negative	total
	Agreements	1	1	0	2
	Chance	0.5	0.5	0	1
	Kappa	(2-1)/(3-1)			
	Kappa	0.5			

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

In terms of the Pearson correlation, the rating should give the numbers, so I ask the raters to give a specific sentiment score to these comments, the results is as follows:

		Rater-1	Rater-2	Rater-3
Classmate A	positive	80	80	0
Classmate B	positive	0	0	0
Classmate C	positive	0	70	60
Classmate A	neutral	0	10	60
Classmate B	neutral	70	90	70
Classmate C	neutral	20	0	20
Classmate A	negative	0	0	0
Classmate B	negative	0	0	0
Classmate C	negative	80	0	20

Lastly, I use the PEARSON() function in excel to get the correlation between raters:

	positive	neutral	negative
R1XR2	0.596039561	0.92770835	#DIV/0!
R1XR3	-0.5	0.419313935	1
R2XR3	0.397359707	0.727903458	#DIV/0!

Because that the correlation coefficient ranges from -1 to 1. A value of 1 implies that a linear equation describes the relationship between X and Y perfectly, with all data points lying on a line for which Y increases as X increases. A value of -1 implies that all data points lie on a line for which Y decreases as X increases. A value of 0 implies that there is no linear correlation between the variables.

We can have a conclusion from the table above, that the neutral has relatively the highest correlation, this mean most people can judge the neutral comments precisely.

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).

1. Liu and Hu opinion lexicon contains around 6800 positive and negative opinion words or sentiment words for English language. This list was composed over many years.
2. SentiWordNet is a lexical resource in which each WordNet synset is associated to three numerical scores Obj(s), Pos(s) and Neg(s), describing how objective, positive, and negative the terms contained in the synset are.
3. The MPQA (Multi-Perspective Question Answering) Subjectivity Lexicon is a list of subjectivity clues that is part of OpinionFinder. It helps to determine text polarity. The list on this page contains more than 6000 positive words and phrases making it one of the longest and best list of positive words.

The positive words I choose are:
admire easy envy glory helpful luck bravo clear well-known worth
The negative words I choose are:
anger alarm error cold crowded cry lose lying mediocre mistake

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.

Admire: it means that If one holds someone in high esteem or look up to someone, then he or she will admire him or her. So admire is really positive. While in some case, it could be interpreted with the opposite valence, for example:
It is dangerous to admire a person who is not righteous.
Easy: the adjective easy can describe anything that comes without too much effort, so it is a positive word, but when using it to describe a person, then it is not a positive word. For example:
She is a easy girl.
Envy: Wanting what someone else has and resenting them for having it is envy. I think this word often explains a negative sentiment of the user, so I don't think it is really a positive word.
Glory: glory means a state of high honor. So it is really a positive word. But it may be interpreted with the opposite valence in some cases, for example:
This little child actually thinks that engaging in dangerous activities is a glory.
Helpful: One's action can be called helpful if he or she is inclined to assist others in any situation. This word is really positive. However it can be interpreted with the opposite valence in some cases, for example:
He is never a helpful person.
Luck: luck means an unknown and unpredictable phenomenon that leads to a favorable outcome. It is often used in a context to explain one's joyful sentiment, so it is really positive, while it can be interpreted with the opposite valence in some cases:
He has no real talents other than luck.
Bravo: Bravo means cheer but could also mean a murder, so I don't think it is really a positive word.
Clear: Clear means clean and uncluttered, without distractions or confusion. It is really positive but when someone is impatient with another person, he maybe say that: "Am I clear?" with a angry voice. In this case it will be negative.
Well-known: means widely or fully known. But considering when it is followed by a negative word, so it is not really a positive word.
Worth: We usually see it as a positive word, but sometimes we may see it not worth it which has the negative sentiment.

Anger: It describes a bad emotion so it is really a negative word, but it can be interpreted with the opposite valence in some cases:

Effective release of anger is good for health.
Alarm: It is often used in a dangerous situation so it is a negative word, but it can be interpreted with the opposite valence in some cases:
The city's alarm system is perfect
Error: it means a mistake in a procedure which is really a negative word,but it can be interpreted with the opposite valence in some cases:
The code is perfect, there is no error in it.
Cold: it describe a environment which has low temperature, which is really a negative word, but it can be interpreted with the opposite valence in some cases:
He really love those cold places and snowy weather.
Crowded: It describes a place which has a lot of people which is a negative word, but it can be interpreted with the opposite valence in some cases:
Only crowded streets have a festive atmosphere.
Cry: It means a sad feeling, so it is a negative word, but it can be interpreted with the opposite valence in some cases:
That baby is so pretty, even though she is crying.
Lose: It means to be defeated in some competition, but it can be interpreted with the opposite valence in some cases:
He is so brave and not afraid to lose.
Lying: It means someone lie to others, but it also means to lie on the bed, so it is not really a negative word.
Mediocre: It means ordinary, which is really a negative word but it can be interpreted with the opposite valence in some cases:
He yearns for a mediocre life.
Mistake: It means doing something wrong in a process which is really a negative word, but it can be interpreted with the opposite valence in some cases:
There is no mistake in this essay.

Q3: Human Ratings Task:

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.

a) Now consider ways to improve the training. Eg if you removed stop words from the inputs what do you think might happen?

I think the training can be improved by removing the stop words, before removing the stop words:

```
using all words as features
train on 7998 instances, test on 2666 instances
accuracy: 0.77344336084021
pos precision: 0.7881422924901186
pos recall: 0.7479369842460615
neg precision: 0.7601713062098501
neg recall: 0.7989497374343586
Most Informative Features
              engrossing = True              pos : neg    =     17.0 : 1.0
                   quiet = True              pos : neg    =     15.7 : 1.0
                mediocre = True              neg : pos    =     13.7 : 1.0
               absorbing = True              pos : neg    =     13.0 : 1.0
                portrait = True              pos : neg    =     12.4 : 1.0
               inventive = True              pos : neg    =     12.3 : 1.0
                   flaws = True              pos : neg    =     12.3 : 1.0
              refreshing = True              pos : neg    =     12.3 : 1.0
            refreshingly = True              pos : neg    =     11.7 : 1.0
                triumph = True               pos : neg    =     11.7 : 1.0
```

After removing the stop words:
accuracy: 0.7618154538634658
pos precision: 0.7612275449101796
pos recall: 0.7629407351837959
neg precision: 0.762406015037594
neg recall: 0.7606901725431358

b) Implement this or another solution in the program and report what happens to the precision and recall of the classifier.

In this question, I change the ratio of training set and test set ,I change the k to the value of 4/5 And we can find that the all the accuracy and precision recall has been increased and the training process has been improved.

```
using all words as features
train on 8530 instances, test on 2134 instances
accuracy: 0.7788191190253045
pos precision: 0.7874396135265701
pos recall: 0.7638238050609185
neg precision: 0.7707006369426752
neg recall: 0.7938144329896907
Most Informative Features
              engrossing = True              pos : neg    =     18.3 : 1.0
                mediocre = True              neg : pos    =     13.7 : 1.0
                   flaws = True              pos : neg    =     13.7 : 1.0
                 generic = True              neg : pos    =     13.0 : 1.0
               absorbing = True              pos : neg    =     13.0 : 1.0
                  boring = True              neg : pos    =     12.4 : 1.0
               inventive = True              pos : neg    =     12.3 : 1.0
              refreshing = True              pos : neg    =     12.3 : 1.0
                    flat = True              neg : pos    =     11.8 : 1.0
            refreshingly = True              pos : neg    =     11.7 : 1.0
```