Question - 3.b

- Implementing Emotion Intensity based on following features
  - N-gram (1, 2) in multivariate fashion for each sentence
  - Overall emoticon score (1)
    - This score is calculated before any normalization, because normalization leads to removal of punctuation
  - Normalization
    - Removal of numbers, as numbers have very little to no impact on emotions
    - Removal of punctuation, as it leads to overall increase in unigrams and bigrams and similar to numbers have little impact on emotion.
    - lower case of the characters, in order to reduce number of unigrams and bigrams.
    - lemmatization (Wordnet Lemmatizer), it is performed both on training data, and the lexicons. The aim is to reduce the overall size of feature size.
  - Expanded Emotion Score (10)
    - A vector of size 10, stores aggregated emotion score for each sentence.
  - Emotion Count (10)
    - A vector of size 10, it aggregates the count of emotion. It provides coarse information about the emotion, while emotion score provides much finer details.
  - Polarity Hashtag (1)
    - Aggregates polarity of hashtags, into a single number.
  - Polarity Sentiment (2)
    - It aggregates the polarity of sentiments, into positive and negative.
  - MPQA (4)
    - It aggregates based on lexicon sentiment into positive, negative, neutral or both.
  - Bing (1)
    - It aggregates the positive sentiment in that sentence
- Evaluation
  - Setting
    - Aggregation function is sum
    - No normalization is performed
    - Dataset : Anger
  - SVM Regression

Metric	Train	Test
Pearson Correlation	0.162	0.0757
Spearman Correlation	0.146	0.0932

- Decision Tree

Metric	Train	Test
Pearson Correlation	0.999	0.342
Spearman Correlation	0.999	0.32

- Multi Layer Perceptron

Metric	Train	Test
Pearson Correlation	0.9468	0.06723
Spearman Correlation	0.9475	0.0663

- Setting
  - Dataset : Joy
- SVM Regression

Metric	Train	Test
Pearson Correlation	0.155	0.105
Spearman Correlation	0.159	0.104

- Decision Tree

Metric	Train	Test
Pearson Correlation	0.999	0.338
Spearman Correlation	0.999	0.328

- Multi Layer Perceptron

Metric	Train	Test
Pearson Correlation	0.291	0.076
Spearman Correlation	0.307	0.026