Project Report-1:

Project Milestone1:
Tirumal Reddy Ramidi
U1414933
u1414933@umail.utah.edu

- So far I have downloaded the data sets provided on Kaggle.
- For the first milestone, I have used the Roberta features set among the three feature sets provided.
- Got the best results when I used this feature set for training my model.
- Also, the Average perceptron would give the maximum accuracy, hence I have used this algorithm.
- I have developed a linear classifier using the perceptron variation using average weights.
- For this algorithm to find the only hyperparameter used in this algorithm, I have passed all three values to the algorithm to get the best learning rate that gives the highest accuracy.
- All Accuracies:
    - Test Accuracy of Averaged Perceptron with learning rate: 1 is: 78.93350888742594
    - Test Accuracy of Averaged Perceptron with learning rate: 0.1 is: 79.98683344305464
    - Test Accuracy of Averaged Perceptron with learning rate: 0.01 is: 76.30019749835418
- The hyperparameter value that is the learning rate with the highest accuracy of 79.98683344305464 is 0.1
- Using this best learning rate, I've trained a linear classifier where the epoch is 20. I retained the best weights with good accuracy on the test dataset.
- With this hyperparameter, I have passed the average weights learned from the training dataset to the prediction function to predict accuracy.
- Then read the eval dataset values into a pandas data frame and passed the eval_x, and averaged weights learned earlier to the eval_predict function. This predicts the labels on the evaluation dataset.
-
- The number of Positive label examples in train data is 3559
- The number of Negative label examples in train data is 3529

Value counts on the eval dataset after prediction is:
- The number of 0's is: 797
- The number of 1's is: 723

- Evaluation accuracy using the Roberta feature set is 80.921

- For the next milestone, I plan on using the Naive Bayes Classifiers, Decision Trees, and SVM.
- Looking forward to starting working on building linear classifiers using some of these exciting machine learning models.
- I plan on using the other two feature sets for the next milestone, and I am eager to work on them and improving the accuracy of the models for these feature sets. This enables me to analyze what type of classifier works best on certain feature sets.