**CS 470 Final Project Implementation and Performance**

Email Spam Classification using Naïve Bayes, Linear Regression, and K Nearest Neighbor

GitHub: <https://github.com/Devin-Jay/EmailSpamClassification>

**Project Details**

1. Import Dataset.

A screen shot of a computer code

Description automatically generated

1. Get training set and evaluation set.

A computer screen shot of text

Description automatically generated

1. Divide the training set into K fold, which in this case is 5.

I did this by looping through the training set and manipulating the indices to get the fold



1. For each fold:
   1. Train model where applicable.





* 1. Evaluate performance on fold.

A screen shot of a computer

Description automatically generated

* 1. Use model on evaluation set.



A blue and white text on a black background

Description automatically generated

A blue and white text

Description automatically generated

1. Display results.

A screenshot of a computer program

Description automatically generated

**Project Classes**

Naïve Bayes Model Class

1. Get training set
2. Train model by calculating attribute likelihoods and spam and non-spam probabilities
3. Use model on eval set

A screen shot of a computer screen

Description automatically generated

A computer screen shot of text

Description automatically generated

A screen shot of a computer code

Description automatically generated

A computer screen shot of a program code

Description automatically generated

Logistic Regression Class

1. Get training set
2. Train model by fitting a line to training set
3. Use model on eval set

A screen shot of a computer code

Description automatically generated

A computer screen shot of a program

Description automatically generated

A computer code with colorful text

Description automatically generated

A computer screen shot of a program code

Description automatically generated

K Nearest Neighbor Class

1. Get training set without target classifier
2. Get eval set
3. Calculate the Euclidian distances between each email in eval set and training set
4. Classify each email in eval set using class of email with least Euclidian distance

A computer screen shot of text

Description automatically generated