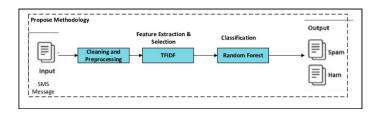
Spam Detection Group #7

General Flowchart:



Motivation:

Spam messages flood email inboxes, clutter text message threads, and invade personal space. Sorting through countless spam messages wastes valuable time and energy.

Process & Models:

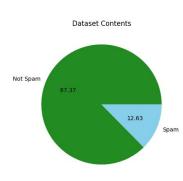
- Preprocessing (stemming, stopwords, Tfidfvectorizer. etc)
- We chose three models:
 Logistic Regression, Naive
 Bayes, and Random Forest
- Split the data 90:10
- Classifier Report for results

What to Improve:

- Our dataset was very unbalanced so we would either oversample or find a more balanced dataset
- More hyperparameter tuning
- Use lemmatization instead of stemming

Results:





WordCloud for most frequent words found in spam messages

