**Ricky Lin / Zhang Zhang**

1. Logistic regression (20% Testing , 80% Training) \_ TF-IDF

Accuracy: 99%

**Tahir Jamil / Chris Mclauglin**

1. Data exploration (Old & New dataset)

**A graph of blue squares

AI-generated content may be incorrect.**

**Rayile / Dustin Van:**

Training of models

1. Bag of World: 93% -> Multinomial
2. TF-IDF: 97% -> Multinomial
3. Text embedding – progressing (SentenseTransformer)

In the future:

* Find optimal hyperparameter

A screenshot of a computer screen

AI-generated content may be incorrect.

A graph with a line

AI-generated content may be incorrect.

**Next Phase:**

**K fold cross validation for finding optimal hyperparameter**

**Testing with uvic dataset (see how much email they can catch)  
  
Question:**

**Can we add some more feature (number of email, link,) beside the number of word (bag of word, TF IDF)**

**\_ E**