

## What is Naive bayes

Naive Bayes is a classifier that works by multiplying, for each class:

- the likelihood that the item is from that class, calculated from the amount of items in that class divided by the total number of items
- for each feature(in this case word) in the item multiply by the number of times that word appears in messages that are in that class.

## Bernoulli Naive Bayes

Bernoulli Naive Bayes is a Naive Bayes classifier that works with boolean features, in this case the probability of spam/ham for a specific word will only be multiplied if it is in the message.

## Results

Each set listed below was trained on 70% of the data in the set(randomly chosen) and tested on the remaining 30%, to avoid overfitting.

Dataset	Lowest accuracy	Highest accuracy
All	84	91
Psy	91	97
Katy Perry	81	92
LMFAO	91	96
Eminem	80	85
Shakira	60	79

Conclusion:

The Bernoulli Naive Bayes Classifier is passable, it could likely be improved by better cleaning of the input data