Challenging Project.

1. "watch anime now".

a. pre filter keeps all of the words.

we know the probability of watch and anime.

P("watch" | spam) = .095 P("anime" | spam) = .095

P("watch" | not spam) = .053. P("anime" | not spam) = .053.

we want to know the probability of "now"

By approach we take to find the likelihood.

$$P(\text{"now"} \mid \text{spam}) = \frac{1+1}{8+13} = \frac{2}{21} = 0.09523$$

$$P(\text{"now"} \mid \text{not spam}) = \frac{0+1}{6+13} = \frac{1}{19} = 0.052631 \approx 0.053$$

P("watch anime now" | spam) = $0.095 \times 0.095 \times 0.095 = 8.57375_E-4$.

P("watch anime now" | not spam) = $0.053 \times 0.053 \times 0.053 = 1.48877_E-4$.

this message is more likely to be a spam.

2. "takeout and anime at my house"

we pre process this string and are left with

"takeout anime my house"

we know the Probability for "anime", "my", "house"

P("anime" | spam) = .095 P("my" | spam) = .0476 P("house" | spam) = .095

P("anime" | not spam) = .053 P("my" | not spam) = .053 P("house" | not spam) = .105

we want to know the probability of "takeout" using the example message contents given:

$$P(\text{"takeout"} \mid \text{spam}) = \frac{0+1}{8+13} = \frac{1}{21} \approx 0.0476$$

$$P(\text{"takecat"} \mid \text{not spam}) = \frac{1+1}{6+13} = \frac{2}{19} \approx 0.105$$

Therefore

$$P(\text{"takeoat and anime at my house"} \mid \text{spam}) = .095 \times .0476 \times .095 \times .0476$$
$$= 2.0448484e-5$$

$$P(\text{"takeout and anime at my house"} \mid \text{not spam}) = .053 \times .053 \times .105 \times .105$$
$$= 3.0969225e-5$$

The message is more likely not to be spam

3. "sell me your anime collection"

every word is kept after pre processing

and we know "anime" 's probability

$P(\text{"anime"} \mid \text{spam}) = .095$

$P(\text{"anime"} \mid \text{not spam}) = .053$

We will calculate "probability of "sell", "me", "your", "collection"

$$P(\text{"sell"} \mid \text{spam}) = \frac{1+1}{8+13} = \frac{2}{21} \approx 0.095$$

$$P(\text{"sell"} \mid \text{not spam}) = \frac{0+1}{6+13} = \frac{1}{19} \approx 0.053$$

$$P(\text{"me"} \mid \text{spam}) = \frac{0+1}{8+13} = \frac{1}{21} \approx 0.0476$$

$$P(\text{"me"} \mid \text{not spam}) = \frac{0+1}{6+13} = \frac{1}{19} \approx 0.053$$

$P(\text{"your"} \mid \text{spam}) = \frac{1+1}{8+13} = \frac{2}{21} \approx 0.095$

$P(\text{"your"} \mid \text{not spam}) = \frac{0+1}{6+13} = \frac{1}{19} \approx 0.053$

$P(\text{"collection"} \mid \text{spam}) = \frac{0+1}{8+13} = \frac{1}{21} \approx 0.0476.$

$P(\text{"collection"} \mid \text{not spam}) = \frac{0+1}{6+13} = \frac{1}{19} \approx 0.053$

So the probability of this message being spam or not spam is

$P(\text{"sell me your anime collection"} \mid \text{spam}) = .095 \times .0476 \times .095 \times .095 \times .0476$
$= 1.94260598E-6$

$P(\text{"sell me your anime collection"} \mid \text{not spam}) = 0.053 \times .053 \times .053 \times .053 \times .053$
$= 4.18195493E-7.$

Therefore this message is likely to be a spam.