Task 1:

Q1:

```
KNN from scratch results:
Accuracy: 0.83  with k = 7
Precision: 0.75
Recall: 0.67
F1 score: 0.7077464788732396
```

```
Naïve Bayes results:
Accuracy: 0.83
Precision: 1.0
Recall: 0.78
F1 score: 0.8764044943820225
```

For code, please visit https://github.com/lumalav/CAP5610/blob/master/HW3/HW3.ipynb

Q2:

KNN from scratch

	ID	Label
0	25	Win
1	26	Win
2	27	Win
3	28	Win
4	29	Lose
5	30	Lose
6	31	Win
7	32	Lose
8	33	Win
9	34	Win
10	35	Win
11	36	Lose

Naïve Bayes

	ID	Label
0	25	Win
1	26	Lose
2	27	Win
3	28	Win
4	29	Win
5	30	Lose
6	31	Lose
7	32	Win
8	33	Win
9	34	Lose
10	35	Win
11	36	Lose

For code, please visit https://github.com/lumalav/CAP5610/blob/master/HW3/HW3.ipynb

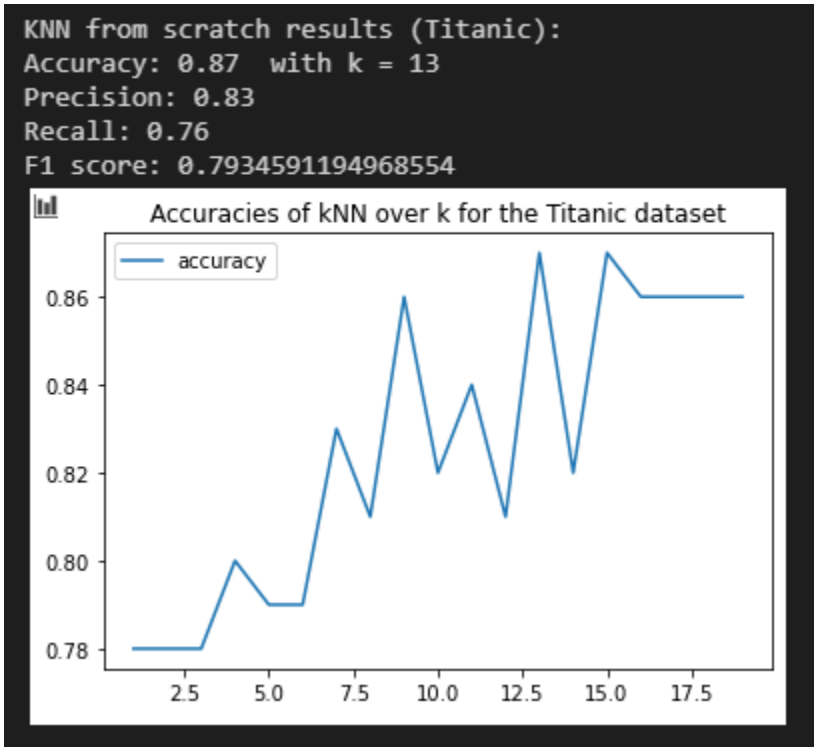
Task 2:

Q1:

```
Naïve Bayes results (Titanic):
Accuracy: 0.88
Precision: 0.76
Recall: 0.95
F1 score: 0.8444444444444444
```

It seems that Naïve Bayes is more effective on bigger datasets. Even though we had bigger recall than for the Football dataset, we had better accuracy.

Q2:



Ask K increases, accuracy tends to increase as well. However, there is a point where no more increasing occurs and tends to decrease.

Q3: I think they are suited for different problems. Bayesian is clear the winner for big data problems like the Titanic dataset. Because of the nature of the algorithm, KNN can be slow with a big dataset. A clear disadvantage of Bayes is that it can suffer from the zero-probability problem where attributes conditional probability equal to zero. When this occurs, Bayes won't be able to predict. But for the titanic dataset is clearly the winner. Even over decision trees where I've got 82% of accuracy on last week's homework.