# Overview of Homework 4

Abed Ahmed (ASA190005) & Dylan Kapustka (DLK190000)

10/02/2022

Homework 4

Machine Learning

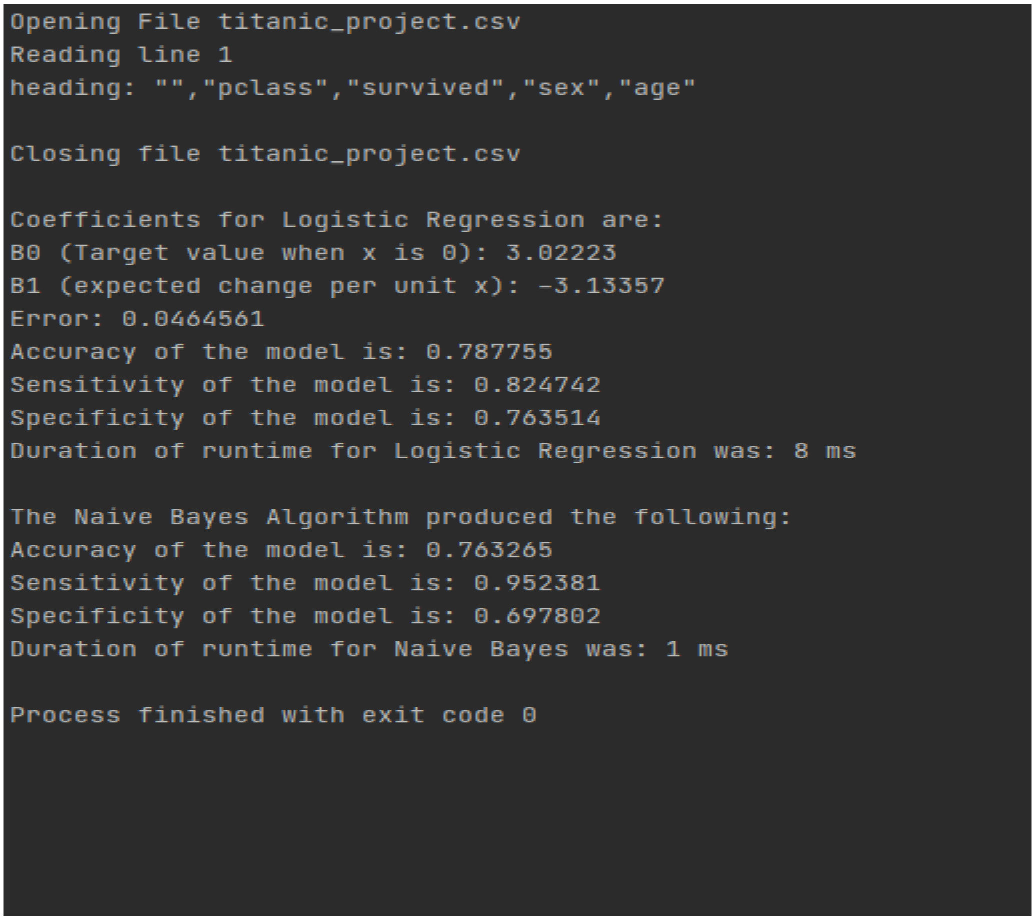
# Analysis of Algorithms

Logistic Regression and Naïve Bayes turned out to be almost identical in terms of Accuracy.

Naïve bayes was extremely sensitive, and quite a bit more sensitive than Logistic Regression.

Specificity was close to the same, with a Logistic Regression being a little bit ahead in that area.

## Output:



```
Opening File titanic_project.csv
Reading line 1
heading: "","pclass","survived","sex","age"

Closing file titanic_project.csv

Coefficients for Logistic Regression are:
B0 (Target value when x is 0): 3.02223
B1 (expected change per unit x): -3.13357
Error: 0.0464561
Accuracy of the model is: 0.787755
Sensitivity of the model is: 0.824742
Specificity of the model is: 0.763514
Duration of runtime for Logistic Regression was: 8 ms

The Naive Bayes Algorithm produced the following:
Accuracy of the model is: 0.763265
Sensitivity of the model is: 0.952381
Specificity of the model is: 0.697802
Duration of runtime for Naive Bayes was: 1 ms

Process finished with exit code 0
```

## Generative Classifiers vs Discriminative classifiers

Discriminative models learn the boundary between classes, while Generative models model the distribution of individual classes. In other words Discriminative models draw boundaries in the data space while generative models model how the data is places throughout the space. From a mathematical perspective, a discriminative ML model trains by learning the parameters of the conditional probability P(Y|X), while a generative model learns the parameters by maximizing the join probability P(X,Y).

Generative models need fewer data points to train as compared to discriminative ones. Generative models make strong assumptions and as a result are more biased. Generative models can also work with missing data while discriminative ones cannot; However, discriminative models are accurate. Both at the end of the day, are classifiers that can classify data into two or more categories.

## Reproducible research in machine learning

Reproducibility with respect to machine learning means that you can repeatedly run your algorithm on certain datasets and obtain the same (or similar) results on a particular project. This process encompasses design, reporting, data analysis and interpretation [1].

Not only does reproducibility guarantee accurate findings, it also promotes transparency and offers us confidence in our ability to comprehend exactly what was done. Reproducibility generally lowers the possibility of mistakes, boosting the dependability of an experiment. It is now well acknowledged that making research reproducible should be a regular activity and that it is a crucial component of any scientific process. [2].

Costs and financial restrictions are another area where repeatability has an impact. Adopting new algorithms might incur significant costs and significant research work, only to produce unreliable results because it lacks information about the hardware, processing capability, training data, and more complex factors like hyper-parameter tweaking [3].

Projects involving machine learning should start by considering reproducibility. It needs to be used in all areas of the project. A reproducibility-focused approach is necessary for everything, from the software and environment through the development and deployment. Making sure that documentation begins on day one is a vital step in developing projects that can be replicated. The documentation process should provide justification for decisions taken as well as a variety of crucial information required to carry out the project successfully, or what Philip Stark refers to as "reproducibility." Additionally, proposed ideas, experiments, and results ought to be tracked [1].

# Work Cited

[1] "The Importance of Reproducibility in Machine Learning Applications." *DecisivEdge*, 14 Oct. 2020, https://www.decisivedge.com/blog/the-importance-of-reproducibility-in-machine-learning-applications/#:~:text=Reproducibility%20with%20respect%20to%20machine,reporting%2C%20data%20analysis%20and%20interpretation.

[2] Ding, Zihao. "5 - Reproducibility." *Machine Learning Blog | ML@CMU | Carnegie Mellon University*, 24 Aug. 2020, https://blog.ml.cmu.edu/2020/08/31/5-reproducibility/.

[3] Hemant, Preeti. "Reproducible Machine Learning." *Medium*, Towards Data Science, 7 Apr. 2020, https://towardsdatascience.com/reproducible-machine-learning-cf1841606805.