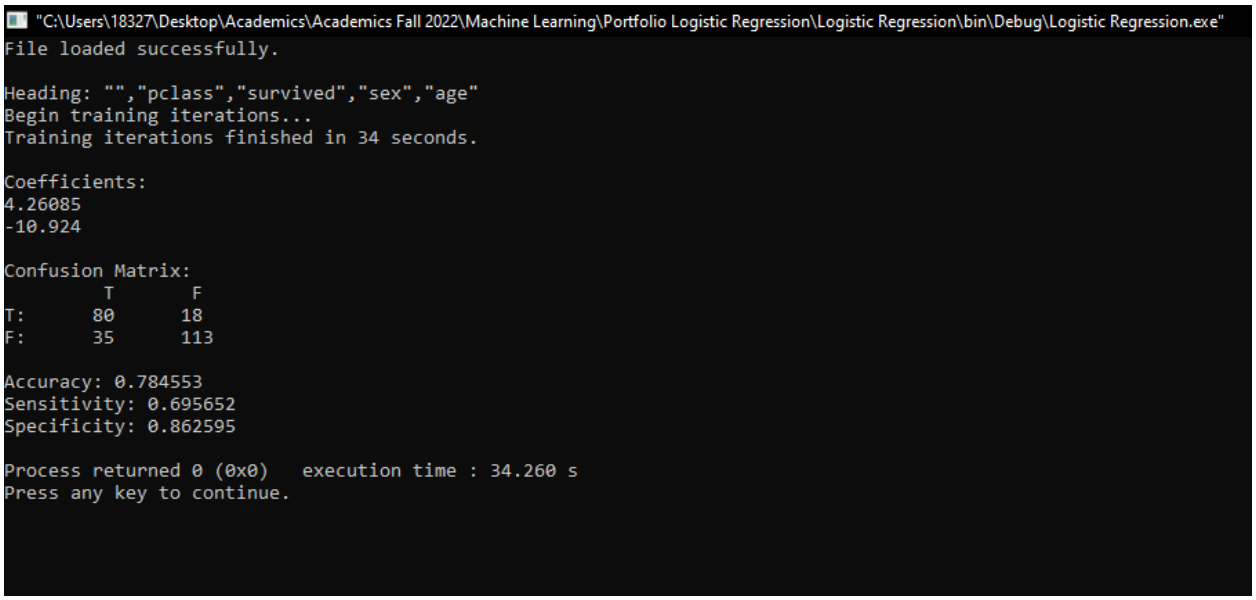
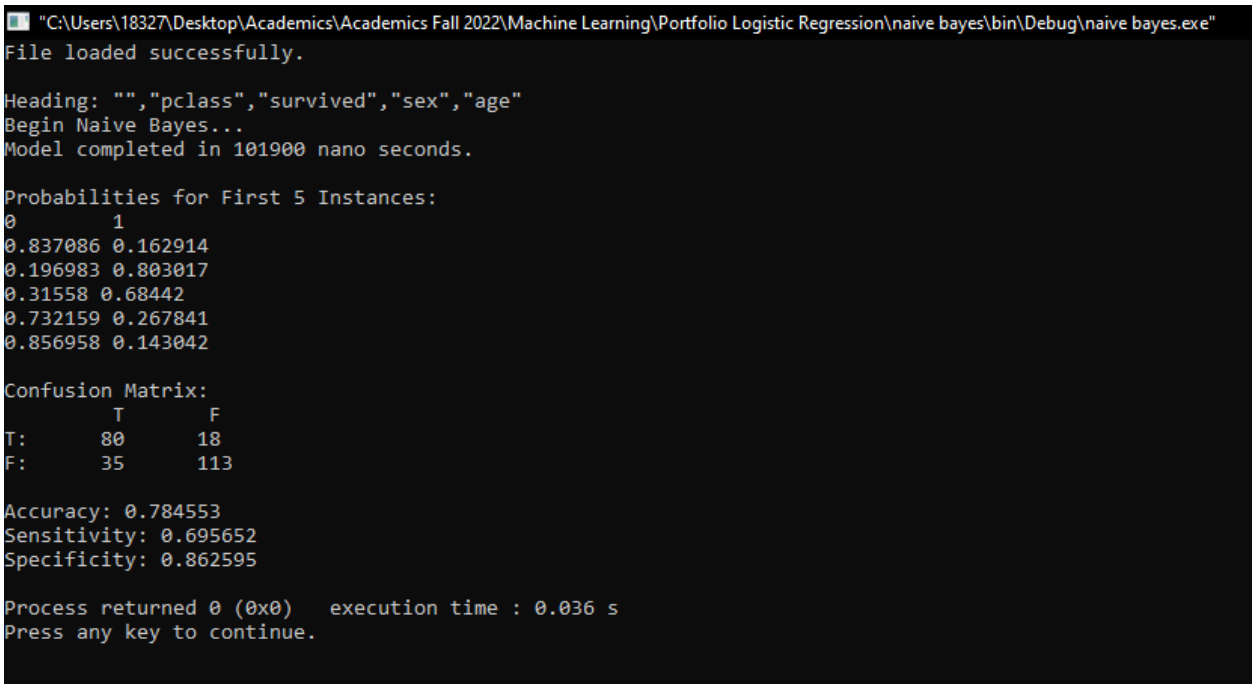
Logistic Regression Output:



Naive Bayes Output:



Discussing results of both algorithms:

Both algorithms had the same base data set which was divided into train and test subsets identically for both algorithms. Bizarrely, both models had equal accuracy, sensitivity, and specificity. This is likely due to the small size of the dataset and in turn the instances that can be used to test each model. With a larger data set, it is almost certain that the accuracy between the two models would differ. With this in mind, however, the only remaining metric to judge the models on is run time. In this regard, the clear standout of the two is the Naive Bayes model as it ran in a mere fraction of the time compared to the logistic regression model. This is likely due to the fact that the Bayes model is essentially just a series of equations the data is run through with little iterations. By comparison, the Logistic Regression model uses about half a million iterations to adjust the coefficients of the model, which takes up most of its thirty second run time. While the Bayes model may be faster and equally accurate to the regression model, it must be stated that the Bayes model is not intrinsically better. Both models have their flaws as well as their strengths.

On Generative Classifiers vs Discriminative Classifiers:

Generative classifiers and discriminative classifiers are two different approaches to classification models. A generative model uses the likelihoods of each of the predictors to generate probabilities for an instance based on its corresponding predictors. In contrast, a discriminative model uses the predictors to create a function that acts as a solid boundary between the classes. A discriminative model's capacity to draw a hard line between classes leaves it vulnerable to misclassify points that fall on the wrong side of the line, comparatively, generative models are more susceptible to outliers in the data [1].

Of the two models created for this assignment, linear regression is a discriminative classifier while the Bayes model is a generative classifier. The Bayes model generates the probabilities of each factor in a class of a given instance, as seen in the output. Meanwhile, the logistic regression model creates a function that represents the line it uses to divide the data, as can be seen by the coefficients in the output.

On Reproducible Research in Machine Learning:

Computational reproducibility is the ability to reproduce the circumstances and results of a computation given sufficient resources [2]. More specifically, in the field of machine learning this means the ability to recreate a machine learning model that can reach the same conclusions as the original work [3]. As machine learning models grow more complex, reproducibility becomes increasingly important. Reproducibility is the key to understanding how an algorithm works and keeping reproducibility in mind will help prevent an algorithm from generating a random or nonsensical output.

There are a number of ways we can and have implemented reproducibility into our work in this course. Most prominently, the set seed command in R. By setting the seed in R we ensure that the random values generated for each execution are the same, thereby ensuring the train and test data for each execution are the same. Another way we can strive for reproducibility is to have clear documentation of programs and throughout our programming process. By documenting our work, anyone who wants to reproduce it will be able to keep and mind any factors that will affect the outcome and may even detect errors in our procedures. BY striving for reproducibility we help anyone who wants to build or learn off of our work in the future.

Works Cited:

[1] C. Goyal, “Deep Understanding of Discriminative and Generative Models.” *Analytics Vidhya*, 19 July 2021, https://www.analyticsvidhya.com/blog/2021/07/deep-understanding-of-discriminative-and-generative-models-in-machine-learning/.

[2] J. Shenouda, W. U. Bajwa, “A Guide to Computational Reproducibility in Signal Processing and Machine Learning”, 15 February 2022, https://arxiv.org/abs/2108.12383

[3] P. Hemant, “Reproducible Machine Learning”, *Towards Data Science*, 17 February 2020, https://towardsdatascience.com/reproducible-machine-learning-cf1841606805