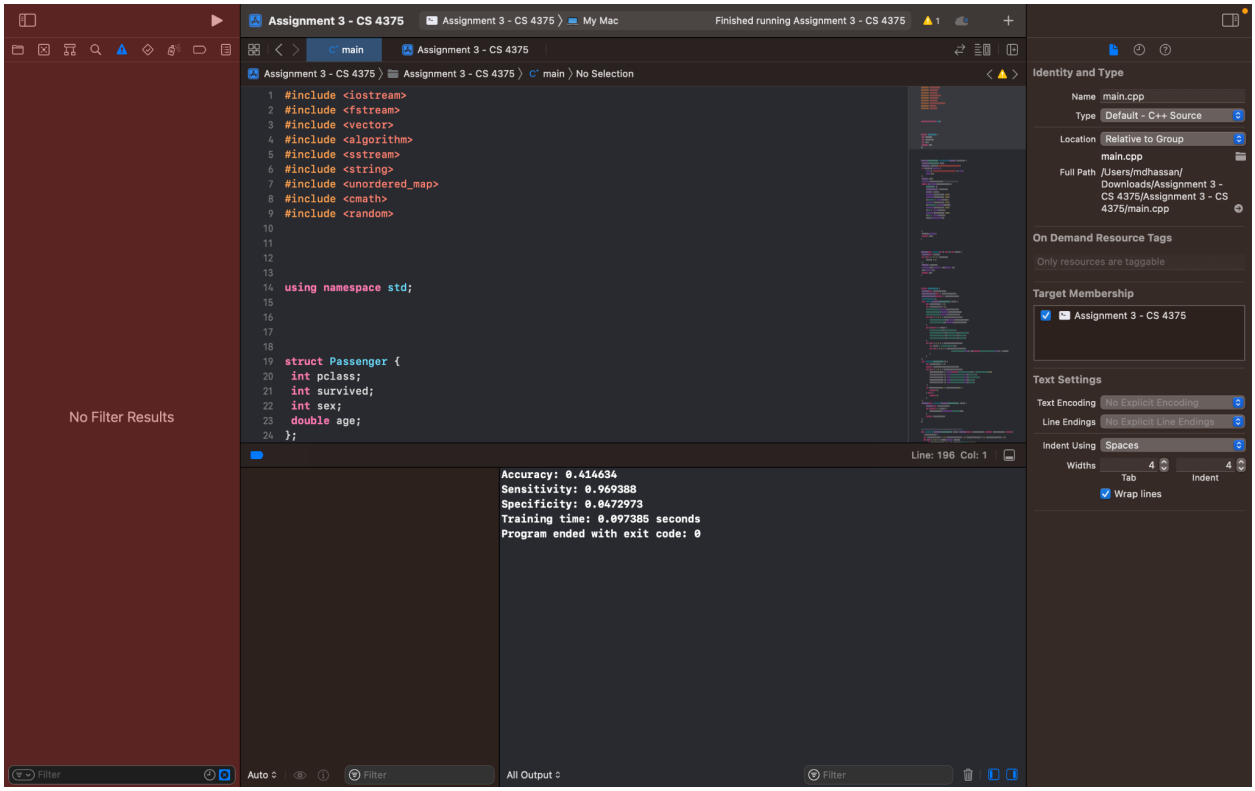
Md Hassan

Jovanni Ochoa

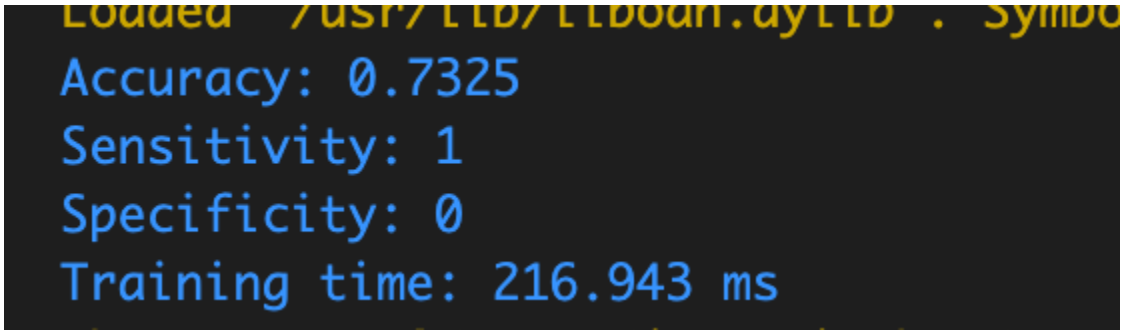
CS 4375

March 5 2022

1. Naive Bayes Model:



Logistic Regression Model:



2.

a. In the Logistic Regression Model we can see that the accuracy is 73% which means that the model is predicting the correct outcome 73% of the time. A sensitivity of 1 represents that the model correctly identifies all the positive outcomes and no false negative results. A specificity of 0 is the complete opposite of sensitivity. In our case it is zero which means that there are no false negative results. Lastly, it took 216.943 ms to train the data that was provided in the csv file.
b. In the Naive Bayes model, our accuracy is 42% which means that the model is correct. Our sensitivity is 97% which is close to one and it identifies positive outcomes and barely any false negative results. Specificity is at 5%. The training time was 0.097s.

3. A Generative model learns the joint probability distribution, while a discriminative model learns the conditional probability distribution. Generative classifiers are used in Naive Bayes, Bayesian networks, Markov random fields and Hidden Markov Models. Discriminative classifiers can be found in Logistic regression, Scalar Vector Model, traditional neural networks, nearest neighbor and conditional random fields.
   Joint Probability distribution refers to probability of specific values of variables occurring at the same time. For example in this project joint probability distribution would refer to the passenger with the same class, sex, age and survival status. Conditional probability distribution refers to the likelihood of an event occurring given that another event has already occurred. For example, we can find the number of passenger that survived based on gender.
4. Reproducible research in machine learning refers to the practice of making the program transparent so that users can repeatedly run the program on different databases and obtain accurate results. It is very important because it adds continuous integration or

delivery cycle. It also promotes transparency which can help in eliminating errors in the program.

In order to achieve transparency, programmers have to keep track and record changes in the algorithm during the experimentation stage. It is very important to keep track of every change made. There are tools like DVC, Neptune command and MLflow tracking that can help programmers achieve that goal. It is also very important to store metadata in a repository as it describes the dataset, computational environment and the model. Lastly, model versioning is also crucial when performing reproducible research. Model Versioning is the organization of controls, changes and implementing policies for the model.

5. Link to Md Hassan: https://github.com/mdyahassan/CS-475-ML
6. Link to Jovanni Ochoa: https://github.com/jovanniochoa/MachineLearning/tree/main

Sources:

- Joshi, P. M. (2018, September 1). *Generative vs discriminative models*. https://medium.com/@mlengineer/generative-and-discriminative-models-af5637a66a3
- Libretexts. (2020, July 27). *5.3: Conditional probability distributions*. Statistics LibreTexts. https://stats.libretexts.org/Courses/Saint_Mary's_College_Notre_Dame/MATH_345__-_Probability_(Kuter)/5%3A_Probability_Distributions_for_Combinations_of_Random_Variables/5.3%3A_Conditional_Probability_Distributions#:~:text=Informally%2C%20we%20can%20think%20of,a%20given%20characteristic%20of%20interest.
- decisivedge. (2022, December 7). *The importance of reproducibility in machine learning applications*. DecisivEdge. Retrieved March 5, 2023, from 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.

- Onose, E. (2023, January 26). *How to solve reproducibility in ML*. neptune.ai. https://neptune.ai/blog/how-to-solve-reproducibility-in-ml#:~:text=Code%3A%20To%20achieve%20reproducibility%2C%20you,the%20outcome%20of%20a%20model.