Rafael Melendez

B.) Performing logistic regression on the titanic data set yielded an accuracy measure of about 76% for survivability solely based on the sex of the passengers. As for the other metrics, its sensitivity were about 81% and 69% respectively. On the other hand, the Naive Bayes program yielded an accuracy of about 79% and therefore marginally higher than that of the linear regression model.

C.)

Classifiers do exactly what their name suggests. They are algorithms that take sets of data from some source and categorize them into classes. Generative Classifiers are defined as algorithms that work to model joint probability distributions of target variables and inputs. On the other hand, Discriminative Classifiers are created to model conditional probabilities of a target given some input variable. Both classifiers can be used to create distinct modes of analysis.

Some factors differentiating these classifiers include how they react to the presence of certain quantities of data. Logistic regression is a type of discriminative classifier that, on average, performs well on binary classification tasks of varying sizes. Naive Bayes, a kind of generative classifier, works exceptionally better with small data sets in the same scenario but starts to dwindle as data grows larger. In general, discriminative algorithms will perform better than generative algorithms because of its simplicity relative to generative classifiers, though there are cases where generative models can come in handy.

D.)

Reproducible research as is defined in a machine learning context is the replication of accurate data from some original source or study. As its name suggests, this kind of computational "research" is conducted in order to replicate data in the absence of sufficient resources/means. The caveat in this case comes with trying to recreate the data as closely as possible given the available resources, including the source study, code, etc. (Alston). In turn, researchers can use this data to cross-examine and confirm already existing findings (Making Your Research Reproducible). Within machine learning, reproducibility is handled through the collection of datasets that have been reformatted or normalized for streamlined usage (Moving Towards Reproducible Machine Learning).

The implications of this method of research include the variability of new data compared to its source. If reproducible research is conducted haphazardly without the attention to detail it requires, it can yield data uncharacteristic to its origin. Not all research studies can be reproduced either, as the lack of detail or depth can leave replication with much to be desired.

Works Cited

Alston, Jesse M, and Jessica A Rick. "A Beginner's Guide to Conducting Reproducible
Research." *Esajournals*,
https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/bes2.1801.

"Making Your Research Reproducible." *NC State University Libraries*,
https://www.lib.ncsu.edu/do/data-management/reproducible-research.

"Moving towards Reproducible Machine Learning." *Nature News*, Nature Publishing
Group, 12 Oct. 2021, https://www.nature.com/articles/s43588-021-00152-6.