Problem 1

For my implementation of the Naive Bayes classifier, it had a training accuracy of 77.4 percent, and testing accuracy of 72.6 percent.

Problem 2

Naive Bayes makes the assumption that independence holds amongst predictors. That is, Naive Bayes assumes that any two given features of the data in a class are unrelated, and independent of the other. However, it is not necessarily true that features are in fact always independent. Thus, the results obtained from Naive Bayes need to be taken with a grain of salt, especially because typically, there is some amount of dependence between features. In the context of credit scores, features are most certainly not dependent. A person's credit score depends entirely on their spending hahabits, and actions they have taken. Applying Naive Bayes would basically assume that these features have no impact on their credit scores, which is not the case.

Problem 3

In order to be able to use the dataset in a Bernoulli Naive Bayes model, the **Month**, **Credit Amount**, and **Number of Credit** sections would have to be binarized, so that that attributes could appear either as 1s, or 0s. For instance, with the credit amount column, this could be done in terms of the maximum credit amount. The **Credit** section is already in terms of 1s and 2s, so each of the '2' rows could be converted into 1, and the '1' rows could be converted into 0. This would allow for the dataset to be used in the model.

Problem 4

Used in US legal theory, disparate impact is a principle referring to practices in employment, policies, and/or regulations that have an adverse effect on protected groups of individuals. If a practice has a disproportionately adverse effect on members of a protected group, then it is considered to be illegal. The term, "disproportionately adverse", is calculated using the 80% rule, which is:

$$P[\hat{Y} = 1|S! = 1]/P[\hat{Y} = 1|S = 1] \leq (t = 0.8)$$

Where $\hat{Y}$ is the output of the hypothesis, and S indicates whether the person is in a protected group.

This might be a useful measure because it prevents models from being discriminatory. For any protected attributes, such as race, then for any two given races, the probability of having a higher credit score should be the same. This measure therefore prevents the presence of a protected attribute from negatively impacting an individual's credit score.

However, conversely, some limitations to this measure are that, if one non-protected and one protected attribute where to have different credit scores for other reasons (nondiscriminatory), then one would not want the model to predict that the two probabilities of both attributes would be equal. Thus, probabilities should not always necessarily be equal as a result of the 80% rule.

Problem 5

In the context of credit ratings, the false positive rate corresponds to there being a 'good' (i.e., higher) credit score (categorized by 1) when in reality, it should be lower (0 instead). The false negative rate, on the other hand, corresponds to there being a 'bad', lower, credit score (categorized by 0), when it reality, it should be higher (1 instead). Thus, for false negatives, the score should be one, but is actually zero, and with false positives, the score should be zero, when it is actually one.

If one group's FPR is much higher than another's, that would mean that they are being more privileged by the model, compared to other attribute groups. For there to be more false positives, there would have to be many cases of individuals scored with a credit score of 1, rather than zero. Thus, this would be indicative of that group being more privileged by the model.

Contrarily, if one group's FNR is much higher than another's, then that would likely mean that they are being discriminated against by the model. In this case, there would be many instances of the model scoring these individuals with a 'bad' credit of '0', when it should instead have been '1'. Therefore, this would imply that they were being discriminated against by the model.