# Naïve Bayes Algorithm Analysis on Universal Bank Dataset

## Write a Script for Naïve Bayes Classification.

Here I wrote the Naïve Bayes Classification Algorithm

## Binning the Variables to Get Categorical Results

I binned the variables as each bin would have the same number of points in it and tried number of bins from 1 to 7 and ran the algorithm for 5 times for each binning with different validation and training datasets each run to get an average of the performances.. The algorithm was performing best when the bin number was 4. Here are the average errors of each bin number:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Bin Number | 1 | 2 | 3 | 4 | 5 |
| Error Rate | 0.0962 | 0.0950 | 0.0960 | 0.0946 | 0.940 |

## Comparison of Error Rates with Naïve Rule Error Rate.

For each run and I mentioned in Part I – b), I also recorded Naïve Rule performance and here are the averages of error rates of Naïve Bayes Algorithm and Naïve rule.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Bin Number | 1 | 2 | 3 | 4 | 5 |
| Naïve Bayes | 0.0962 | 0.0950 | 0.0960 | 0.0946 | 0.940 |
| Naïve Rule | 0.0962 | 0.0950 | 0.0960 | 0.0946 | 0.940 |

Interestingly, although my code predicts the validation data differently than Naïve Rule, they nearly all the time give the same error rate. Thus, I cannot say that my code is good enough. Probably, reason behind this is the way data is designed. The data has very little “1” labeled data point and is not very nice to work on.