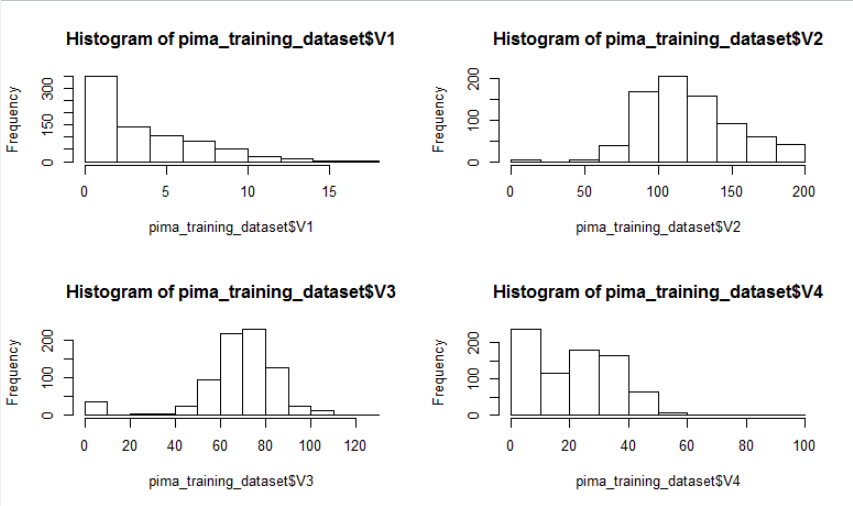
**Machine Learning Assignment -3**

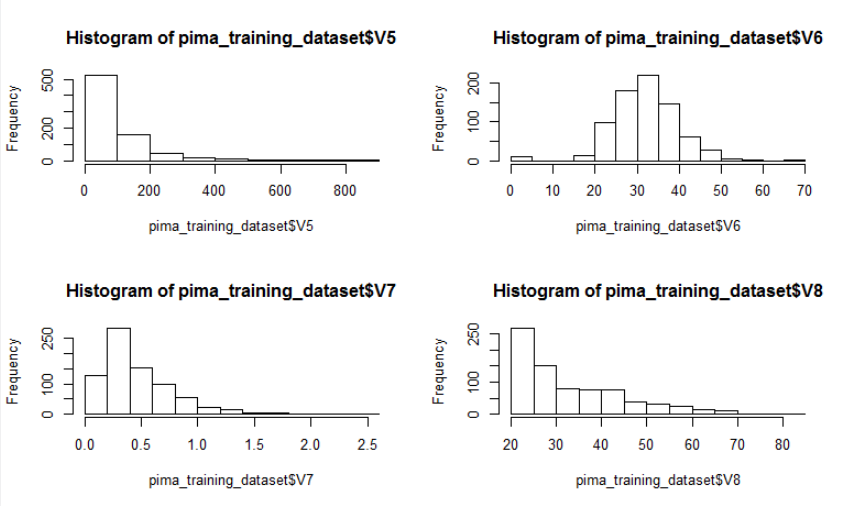
NAVANEETH VENUGOPALA RAO

[nbv140130@utdallas.edu](mailto:nbv140130@utdallas.edu)

1. **Exploratory Data Analysis**

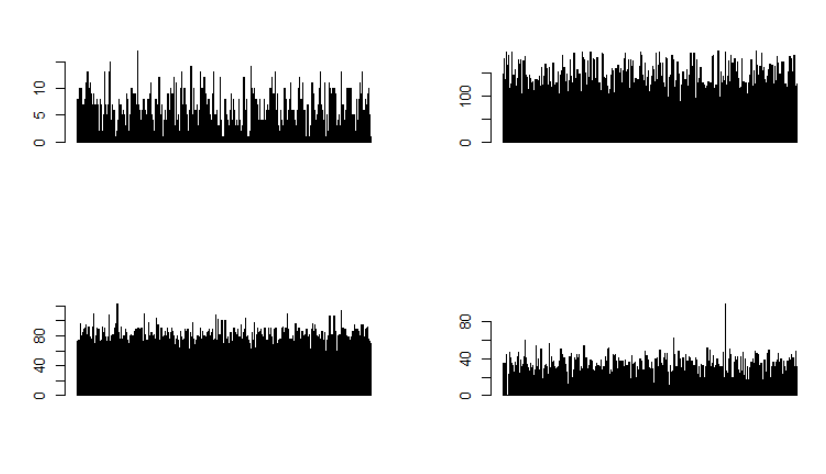
* *Histograms*

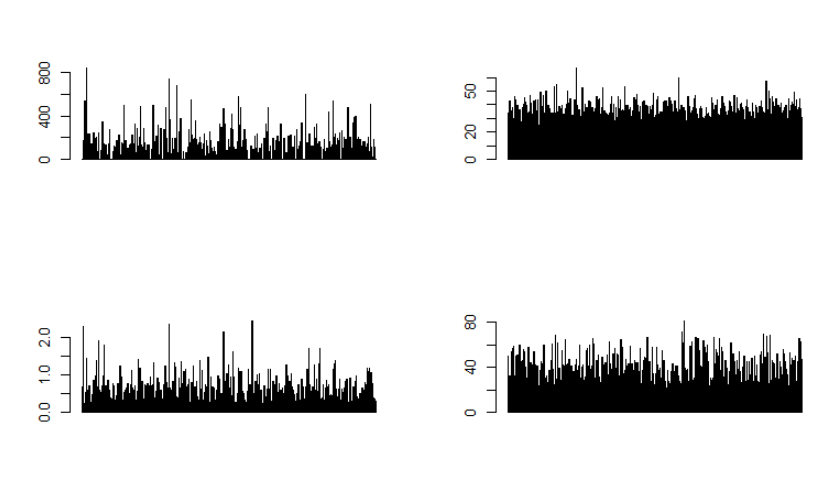




***Note :*** *From Histogram plot, it is observed that variables V2, V3 and V6 are normally distributed.*

* *BarPlots*





* *Correlation between each attributes and class variable are :*

[1] "The Correlation between : "

[1] "Attribute1 and the class variable is: 0.221898153033987"

[1] "Attribute2 and the class variable is: 0.466581398306874"

[1] "Attribute3 and the class variable is: 0.0650683595503328"

[1] "Attribute4 and the class variable is: 0.0747522319183194"

[1] "Attribute5 and the class variable is: 0.130547954884048"

[1] "Attribute6 and the class variable is: 0.292694662644446"

[1] "Attribute7 and the class variable is: 0.17384406565296"

[1] "Attribute8 and the class variable is: 0.238355983027198"

* *Two Attributes that has the highest mutual correlation:*

[1] "Maximum correlation whereabouts :"

[1] " a. Max Correlation Value : 0.544341228402339"

[1] " b. The First attribute : V1"

[1] " c. The Second atrribute : V8"

1. **Naïve Bayesian Classifier**

|  |  |
| --- | --- |
| ***Experiment*** | ***Accuracy (in %)*** |
| 1 | 76.62 |
| 2 | 75.32 |
| 3 | 66.23 |
| 4 | 79.22 |
| 5 | 70.12 |
| 6 | 80.51 |
| 7 | 74.02 |
| 8 | 75.32 |
| 9 | 80.51 |
| 10 | 74.02 |

**Overall Accuracy = 74.78%**

1. **SVM Classifier**

* *Default SVM*

|  |  |
| --- | --- |
| ***Experiment*** | ***Accuracy (in %)*** |
| 1 | 81.81 |
| 2 | 84.41 |
| 3 | 79.22 |
| 4 | 87.01 |
| 5 | 81.81 |
| 6 | 87.01 |
| 7 | 83.11 |
| 8 | 84.41 |
| 9 | 81.81 |
| 10 | 84.41 |

**Overall Accuracy = 78.31%**

* *For Different Kernels*

|  |  |
| --- | --- |
| ***Kernel*** | ***Average Accuracy of 10 Experiments (in %)*** |
| Linear | 77.79 |
| Polynomial | 76.23 |
| Radial | 76.33 |
| Sigmoid | 75.668 |

1. **kNN Classifier**

|  |  |
| --- | --- |
| ***K*** | ***Average Accuracy for 10 Experiments (in %)*** |
| 3 | 68.13 |
| 5 | 70.17 |
| 7 | 71.42 |
| 9 | 73.68 |
| 11 | 75.13 |

**Note: The default SVM is the best classifier for this Dataset. Naïve Bayesian’s accuracy is close enough but it’s value deviates a lot and hence not considered. The same holds true for kNN.**