# Assignment 5 – ML – Empirical study of different training models

## Dharmam Buch – DSB150030

1. We would be using the following data sets :-
   1. Wine
   2. Car
   3. Haberman
   4. Abalone
   5. Iris
2. Summary Of Observation :-

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Number Of Instance** | **Number Of Attributes** | **Bagging** | **Boosting** | **Gradient Boosting** | **Random Forest** | **KNN** |
| Wine | 135 | 13 | 89.01 | 100 | 49.6 | 99.5 | 68.8 |
| Car | 1588 | 6 | 90.5 | 97.6 | 53.4 | 99.4 | 89.5 |
| Haberman | 248 | 4 | 80.2667 | 83.4 | 64.9 | 97.1 | 75.2 |
| Abalone | 3978 | 9 | 26.4 | 25.7 | 24.9 | 99.91 | 78.2 |
| Iris | 117 | 5 | 95.68 | 100 | 56.5 | 99.6 | 94.3 |
| **Total** |  |  | 76.37134 | 81.34 | 49.86 | 99.102 | 81.2 |

Note : **Abalone was a skewed data set,** for example, Rings = 1,2,3 had 15/30 training data each and rings like 28,29 had just 1 training example. Still Random Forest Performed the best.

1. **Hence, the experiment was fairly successful as all the concepts taught in class were numerically visualized and validated.**
2. **Random Forest** was the best classifier followed by bagging and boosting.