**Survival Prediction of Titanic Voyagers**

* Initially the dataset “[Titanic - Machine Learning from Disaster](https://www.kaggle.com/competitions/titanic/discussion/6240)” was downloaded from Kaggle.
* The dataset contained the following features:

'PassengerId', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked’ and the target 'Survived'.

* The column Name was dropped as it seemed less relevant and the class Ticket was also dropped since Cabin and Fare was already present.
* The missing Age and Fare values were filled based on mean. While, Embarked, and Sex were label encoded. For Cabin, we marked the null values as 0 because it meant that the corresponding id’s did not avail any Cabin and as 1 to represent that Cabin was availed by the passenger.
* The train, test and actual data was stored in separate files. Hence, no splitting was required.
* So, the training data was fed to the six classification models(Logistic Regression, Support Vector, Naive Bayes, K-NN, Decision Tree, Random Forest) and the result was analyzed using the test data. The summary may be noted below:

| Model | Accuracy | F1 | Precision | Recall | AUC |
| --- | --- | --- | --- | --- | --- |
| LR | 92.1% | 89.45% | **86.95%** | **92.10%** | **0.98** |
| SVM | **92.82%** | **90.19%** | 89.61% | 90.78% | 0.97 |
| NB | 79.66% | 73.18% | 70.30% | 76.31% | 0.85 |
| RF | 79.42% | 71.71% | 71.71% | 71.71% | 0.89 |
| DT | 74.64% | 65.80% | 64.55% | 67.10% | 0.72 |
| KNN | **63.87%** | **49.49%** | **50.30%** | **48.68%** | **0.66** |