Spot checking (before hyper tuning)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **F1 Score** | **Precision** | **Recall** | **Accuracy** | **ROC AUC** |
| Linear Discriminant Analysis | 0.67 | 0.67 | 0.69 | 0.69 | 0.65 |
| Random Forest Classifier | 0.66 | 0.67 | 0.70 | 0.69 | 0.69 |
| Logistic Regression | 0.64 | 0.64 | 0.67 | 0.67 | 0.65 |
| Gradient Boosting Classifier | 0.63 | 0.65 | 0.65 | 0.65 | 0.64 |
| Ada Boost Classifier | 0.62 | 0.62 | 0.63 | 0.63 | 0.59 |
| KNeighbors Classifier | 0.60 | 0.59 | 0.63 | 0.63 | 0.58 |
| Decision Tree Classifier | 0.58 | 0.62 | 0.61 | 0.60 | 0.53 |
| Gaussian NB | 0.58 | 0.62 | 0.57 | 0.57 | 0.58 |
| SVC | 0.56 | 0.58 | 0.67 | 0.67 | 0.66 |

Model performance (F1 Score) after hyper parameter tuning

|  |  |  |
| --- | --- | --- |
| **Model** | **F1 Score** | **Accuracy** |
| Random Forest | 0.816 | 0.721 |
| Gradient Boosting | 0.812 | 0.718 |
| Logistic Regression | 0.805 | 0.690 |
| Support Vector Classifier | 0.805 | 0.674 |
| AdaBoost | 0.796 | 0.697 |
| Linear Discriminant Analysis | 0.789 | 0.690 |
| K-Nearest Neighbors | 0.768 | 0.653 |
| Decision Tree | 0.748 | 0.629 |
| Gaussian Naive Bayes | 0.639 | 0.568 |

FOR PERSISTENT IMD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **F1 Weighted** | **Precision Weighted** | **Recall Weighted** | **Accuracy** | **ROC AUC** |
| Logistic Regression | 0.769 | 0.722 | 0.823 | 0.823 | 0.638 |
| Gaussian NB | 0.649 | 0.722 | 0.652 | 0.652 | 0.541 |
| KNeighbors Classifier | 0.804 | 0.786 | 0.844 | 0.844 | 0.659 |
| Decision Tree Classifier | 0.763 | 0.764 | 0.776 | 0.762 | 0.522 |
| SVC | 0.777 | 0.718 | 0.847 | 0.847 | 0.631 |
| Random Forest Classifier | 0.775 | 0.717 | 0.844 | 0.851 | 0.677 |
| Gradient Boosting Classifier | 0.787 | 0.751 | 0.830 | 0.833 | 0.677 |
| Ada Boost Classifier | 0.765 | 0.739 | 0.800 | 0.800 | 0.608 |
| Linear Discriminant Analysis | 0.759 | 0.724 | 0.799 | 0.799 | 0.646 |

Afret hyper parameter

Best Model Selected: Random Forest

|  |  |  |
| --- | --- | --- |
| **Model** | **F1 Score** | **Accuracy Score** |
| Random Forest | 0.919 | 0.847 |
| Logistic Regression | 0.917 | 0.847 |
| Support Vector Classifier | 0.917 | 0.847 |
| Gradient Boosting | 0.917 | 0.847 |
| AdaBoost | 0.917 | 0.847 |
| K-Nearest Neighbors | 0.916 | 0.847 |
| Linear Discriminant Analysis | 0.887 | 0.799 |
| Decision Tree | 0.880 | 0.799 |
| Gaussian Naive Bayes | 0.739 | 0.652 |