Python programming language has been used to carry out the experiment work. For evaluating the efficacy of the various machine learning models, confusion matrix and ROC curve calculations have been made. A 2X2 matrix called the confusion matrix contains numerical values for the variables TP, FP, TN, and FN. At the matrix, TP and TN are kept in the diagonal places. Measures of the expected and predicted results using TP and TN agree. Type 1 error, or FP, is another name for it in the classification. It reflects the numbers that were genuinely negative but were mistakenly categorized as positive. Type 2 error is another name for FN in the categorization. It depicts the values that were mistakenly categorized as negative but were actually good. The dataset is taken from JSPL power plant, which consist of 10 features about the major operational factors which are measured during power generation. Given dataset consist of 1000 measured data elements namely, Date/Time, LOAD, VACUUM-1, VACUUM-2, LP1EXH\n HOOD TMP, LP2 EXH\n HOOD TMP, COND-A CW\n l/L TMP, COND-A CW\n O/L TMP, COND-B CW\n l/L TMP, COND-B CW\n O/L TMP. LOAD is our dependent variable for experimental prediction and VACUUM-1 is independent variable. VACUUM-1 elements have been converted to binary numbers using threshold technique for better prediction, where 80% data is used in training and 20% is used for testing part.
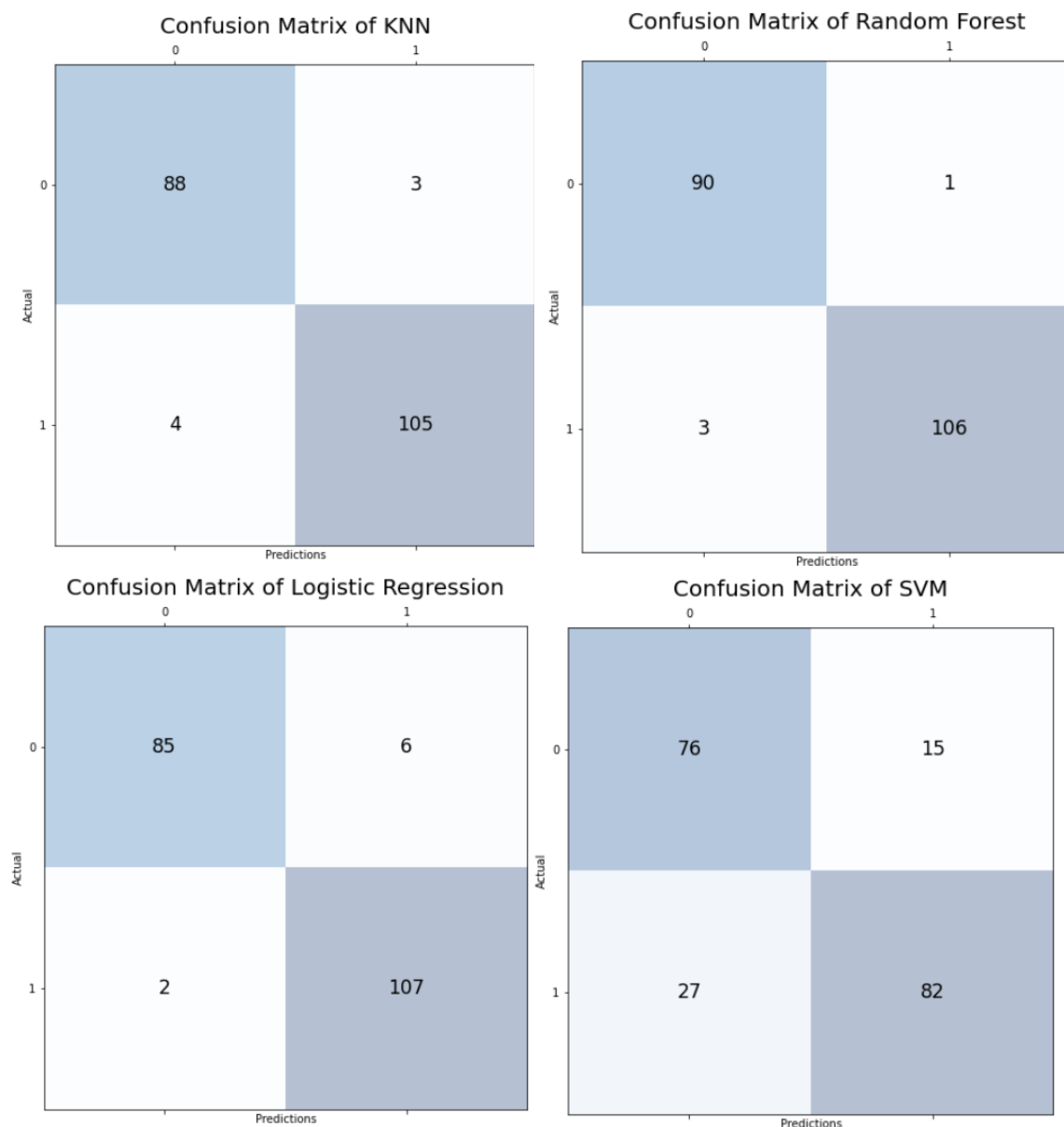

**Figure 1. Confusion Matrix of Different ML Models**

By using figure 1 suggested models have undergone quantitative evaluation using Precision, Recall, and F1-score. Table 1 presents quantitative information gathered using several machine learning techniques. According to table 1, Random Forest offers greater accuracy when compared to models for logistic regression, support vector machines, and KNN. Table 1 also shows that Random Forest provides superior Precision and Recall ratios. Using F1-score as well, it was discovered that Random Forest provides a greater performance ratio. As a result, it has been found through quantitative evaluation that Random Forest is better competent than other models to analyse the forecast.

TABLE 1

MODELS	ACCURACY	PRECISION	RECALL	F1-SCORE
Logistic Regression	0.9600	0.9341	0.9816	0.9551
Random Forest	0.9800	0.9890	0.9724	0.9783
KNN	0.9650	0.9670	0.9633	0.9617
SVM	0.7900	0.8352	0.7522	0.7835

ROC Curve

ROC curve of different machine learning models has been presented in figure 2. For better understanding of model's performance. ROC is a graph that displays how well a classification model performs across all thresholds. True Positive Rate and False Positive Rate, two parameters, are plotted on this graph. TPR vs. FPR are plotted on a ROC curve at various categorization criteria. More items are classified as positive when the classification threshold is lowered, which raises the number of both False Positives and True Positives. A typical ROC curves of Random Forest, Logistic Regression, SVM and KNN models are shown in figure 2.
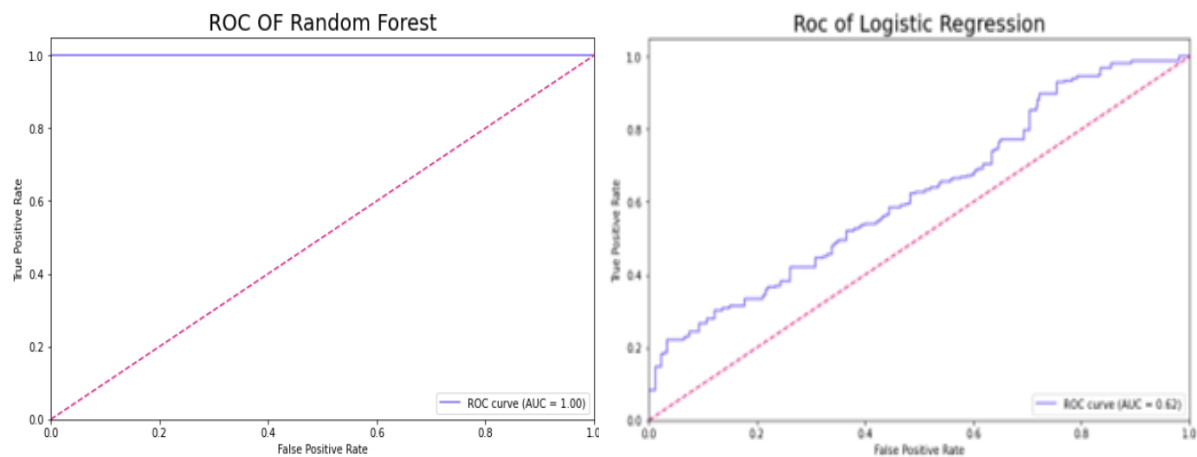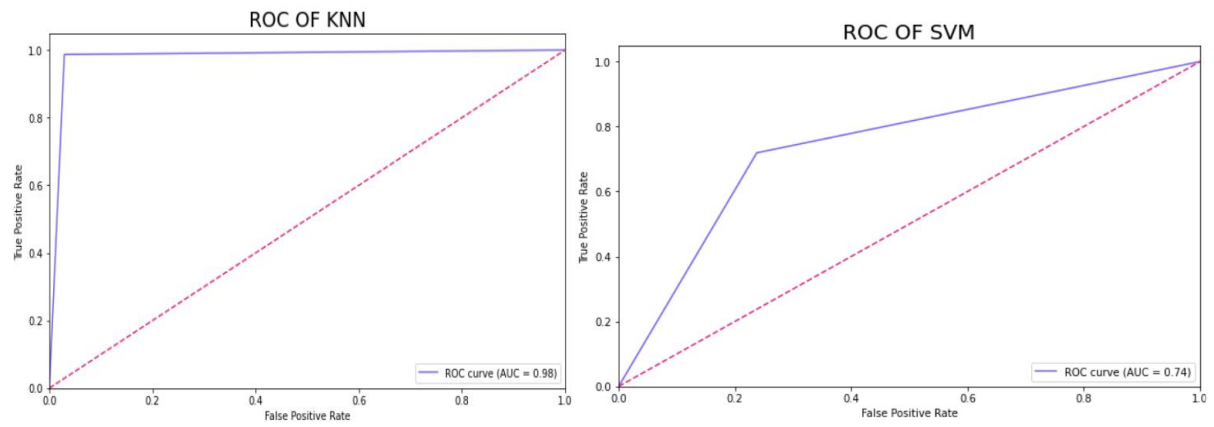
Figure 2. ROC Curve of Different ML Models

By above analysis it has been found that Random Forest is best fitted for our experimental work and more accurate to predict the LOAD required during power generation by considering the VACCUM measures.