Experiment B:

The same 5 cross validation split was used to compare the SVM classifier with 2 other classifiers – RandomForestClassifier and LogisticRegression.

RandomForestClassifier:
Parameters: **n_estimators**: Number of trees in the forest.

max_features : The number of features to consider when looking for the best split:

- If "auto", then max_features=sqrt(n_features).
- If "sqrt", then max_features=sqrt(n_features) (same as "auto").
- If "log2", then max_features=log2(n_features).

GridSearchCV was used to tune hyperparameter with the above mentioned 3 values and the best one was chosen.

Then the best value was chosen to be used to train the model on the training set and the accuracy and F1 score were provided for the testing set.

Logistic Regression:

Parameters: 'C': [0.001, 0.01, 0.1]

GridSearchCV was used to tune hyperparameter with the above mentioned 3 values and the best one was chosen.

The best parameter C = 0.001 was used to train the model and the corresponding Accuracy and F1 score were obtained.

The accuracy and F1 score for all the three classifiers were plotted.

Logistic Regression:

Best C params:

```
{'C': 0.001}
{'C': 0.01}
{'C': 0.01}
{'C': 0.01}
```

```
parameter metric
0.001     Accuracy   0.92036
          F1        0.895148
```

Random Forest Classifier:

```
{'max_features': 'auto', 'n_estimators': 200}
{'max_features': 'sqrt', 'n_estimators': 700}
{'max_features': 'log2', 'n_estimators': 200}
{'max_features': 'auto', 'n_estimators': 200}
```

```
                        Test
parameter metric
auto      Accuracy  0.911357
          F1        0.882353
```

Comparison of different classifiers:

	SVM(linear=kernel)	SVM(LinearSVC)	RFC	Logistic
Accuracy	0.923823	0.919668	0.911357	0.920360
F1	0.900437	0.895661	0.882353	0.895148