Marketing Campaigns - by Portuguese banking institution

Objective

Find prospective customers who will subscribe for deposit

We do not want to miss any customer so **Recall** is considered as more appropriate performance metrics

The models used for comparison are:

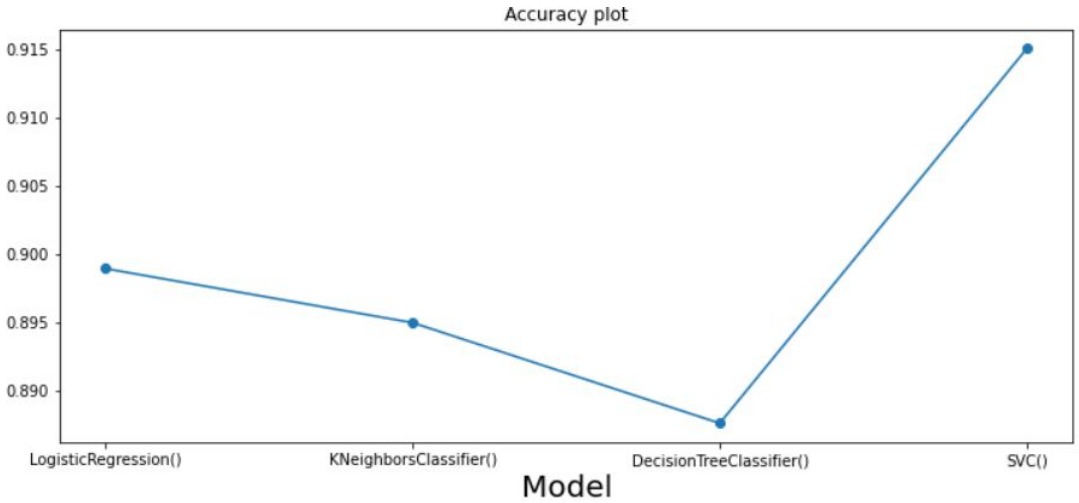
- KNN
- Decision tree
- SVC
- Logistic Regression

After the data analysis the features are used for model calculation are in the next slide

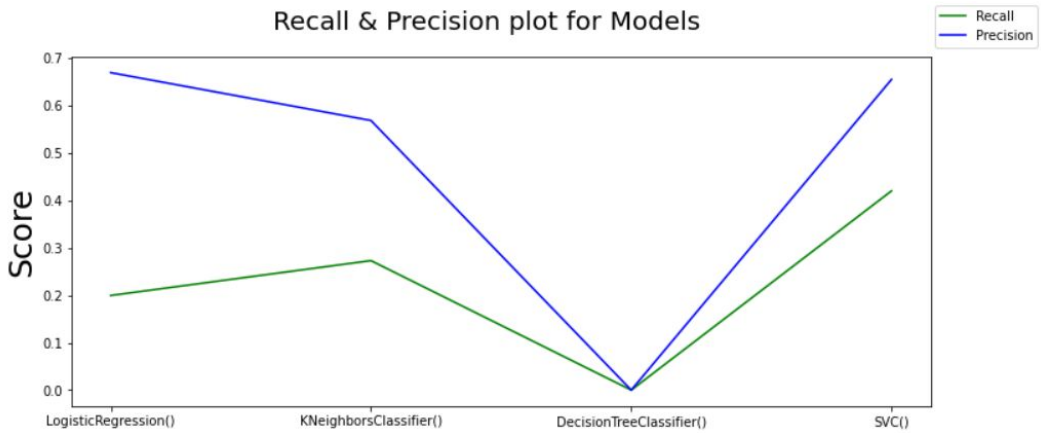
Feature	Status	Reason for dropping
age	Dropped	From the feature selection
job	Used	
marital	Used	
education	Used	
default	Used	
housing	Used	
loan	Dropped	From the feature selection
contact	Dropped	From the feature selection
month	Used	
day_of_week	Dropped	Subscription is same for all 5 days
duration	Dropped	Known only after the call
campaign	Used	
pdays	Used	
previous	Used	
poutcome	Used	
emp.var.rate	Dropped	it is highly corellated to euribor3m & cons.conf.idx, the model will unstable when we have multicollinearity
cons.price.idx	Used	
cons.conf.idx	Used	
euribor3m	Used	
nr.employed	Used	

KNN model suites our requirements, it has better recall score.

Below plots show the accuracy of different models.



Below plots show the recall & precision of different models.



SVC model is executed for 10 % of the data.
When considering the time, recall, and accuracy KNN will suit the dataset.

	Model	Execution Time	Accuracy	Recall	Precision	Best Paramater
0	LogisticRegression()	24.651042	0.898924	0.199424	0.669082	{'model__C': 2.7825594022071245, 'model__penal...
1	KNeighborsClassifier()	198.567001	0.894958	0.272858	0.568216	{'model__leaf_size': 15, 'model__n_neighbors':...
2	DecisionTreeClassifier()	4.425010	0.887594	0.000000	0.000000	{'model__max_depth': 2, 'model__min_impurity_d...
3	SVC()	26.647332	0.915049	0.419847	0.654762	{'model__gamma': 0.1, 'model__kernel': 'linear'}

Recommendations from the analysis

Proposed model : KNN

Proposed parameters:

```
{'model__leaf_size': 15,
 'model__n_neighbors': 9,
 'model__weights': 'uniform'}
```