

# Dataset A + SVM

Hannah LeBlanc CISC873

# Recap

- Radial Basis Function kernel
- $C = 10$ ,  $\gamma = 0.01$

Accuracy: 0.803

	precision	recall	f1-score	support
0	0.84	0.90	0.87	5174
1	0.66	0.52	0.59	1869
avg / total	0.79	0.80	0.80	7043

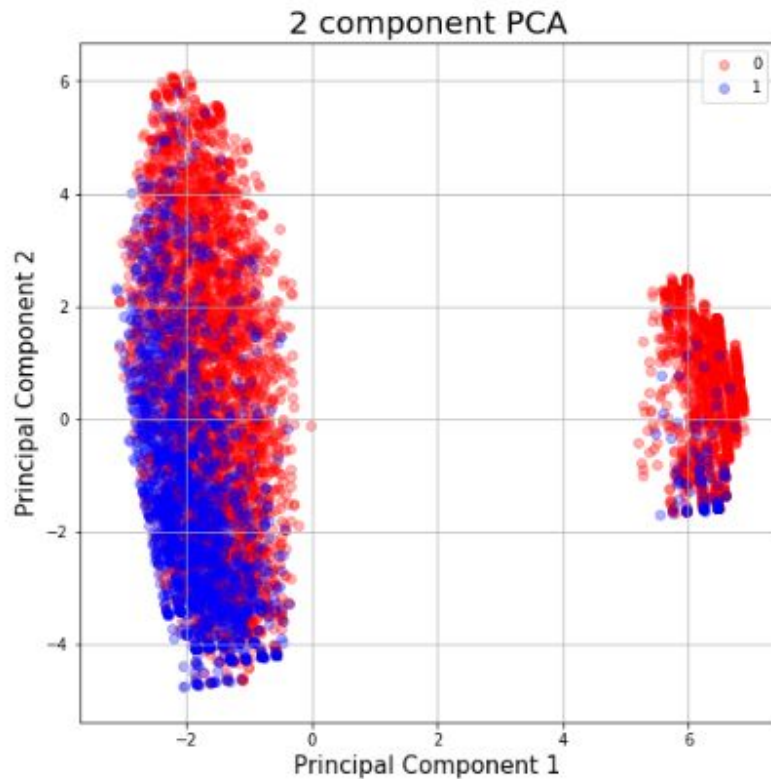
```
Data columns (total 44 columns):
customerID      7043 non-null object
gender          7043 non-null int32
SeniorCitizen  7043 non-null int64
Partner        7043 non-null int32
Dependents     7043 non-null int32
tenure         7043 non-null int64
PaperlessBilling 7043 non-null int32
MonthlyCharges 7043 non-null float64
TotalCharges   7043 non-null float64
Churn          7043 non-null int32
PhoneService_No 7043 non-null uint8
PhoneService_Yes 7043 non-null uint8
MultipleLines_No 7043 non-null uint8
MultipleLines_No phone service 7043 non-null uint8
MultipleLines_Yes 7043 non-null uint8
InternetService_DSL 7043 non-null uint8
InternetService_Fiber optic 7043 non-null uint8
InternetService_No 7043 non-null uint8
OnlineSecurity_No 7043 non-null uint8
OnlineSecurity_No internet service 7043 non-null uint8
OnlineSecurity_Yes 7043 non-null uint8
OnlineBackup_No 7043 non-null uint8
OnlineBackup_No internet service 7043 non-null uint8
OnlineBackup_Yes 7043 non-null uint8
DeviceProtection_No 7043 non-null uint8
DeviceProtection_No internet service 7043 non-null uint8
DeviceProtection_Yes 7043 non-null uint8
TechSupport_No 7043 non-null uint8
TechSupport_No internet service 7043 non-null uint8
TechSupport_Yes 7043 non-null uint8
StreamingTV_No 7043 non-null uint8
StreamingTV_No internet service 7043 non-null uint8
StreamingTV_Yes 7043 non-null uint8
StreamingMovies_No 7043 non-null uint8
StreamingMovies_No internet service 7043 non-null uint8
StreamingMovies_Yes 7043 non-null uint8
Contract_Month-to-month 7043 non-null uint8
Contract_One year 7043 non-null uint8
Contract_Two year 7043 non-null uint8
PaymentMethod_Bank transfer (automatic) 7043 non-null uint8
PaymentMethod_Credit card (automatic) 7043 non-null uint8
PaymentMethod_Electronic check 7043 non-null uint8
PaymentMethod_Mailed check 7043 non-null uint8
extra_fees      7043 non-null float64
dtypes: float64(3), int32(5), int64(2), object(1), uint8(33)
```

# PCA and SVM

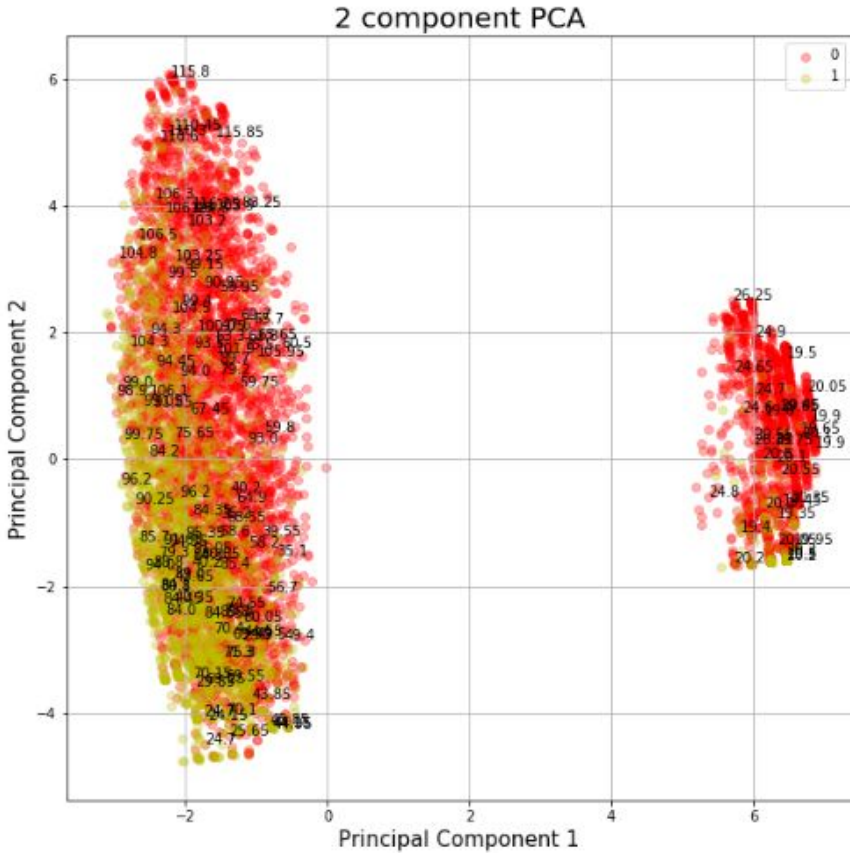
Accuracy: 0.7765156893369303

```
[[4744 1144]
 [ 430  725]]
```

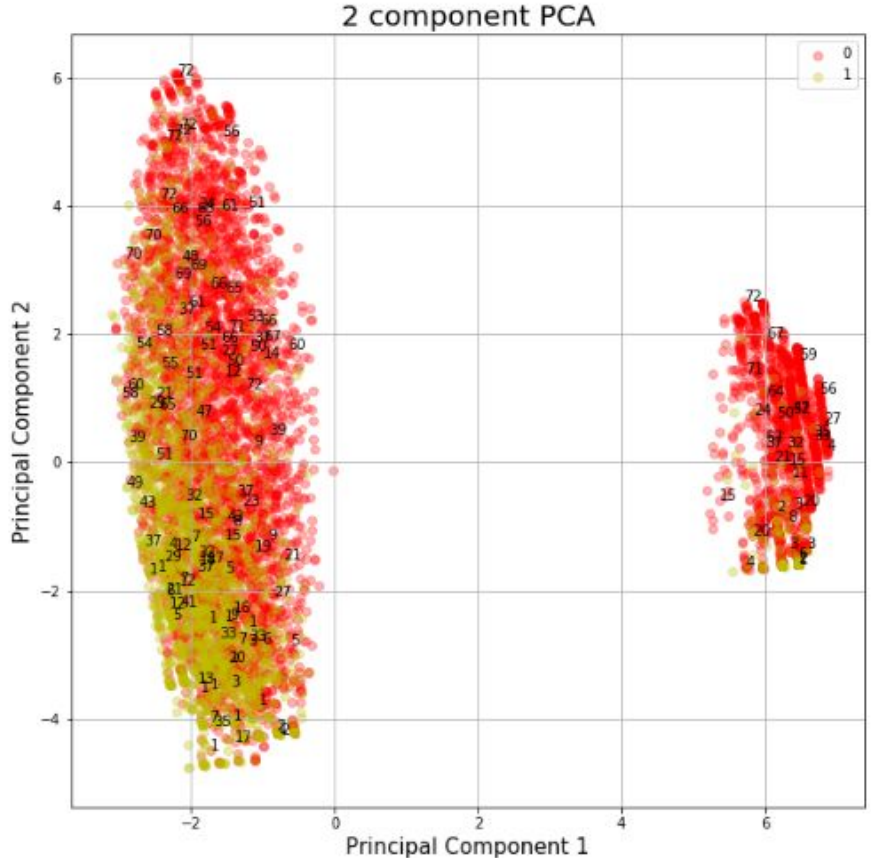
	precision	recall	f1-score	support
0	0.81	0.92	0.86	5174
1	0.63	0.39	0.48	1869
avg / total	0.76	0.78	0.76	7043



# Monthly Charges



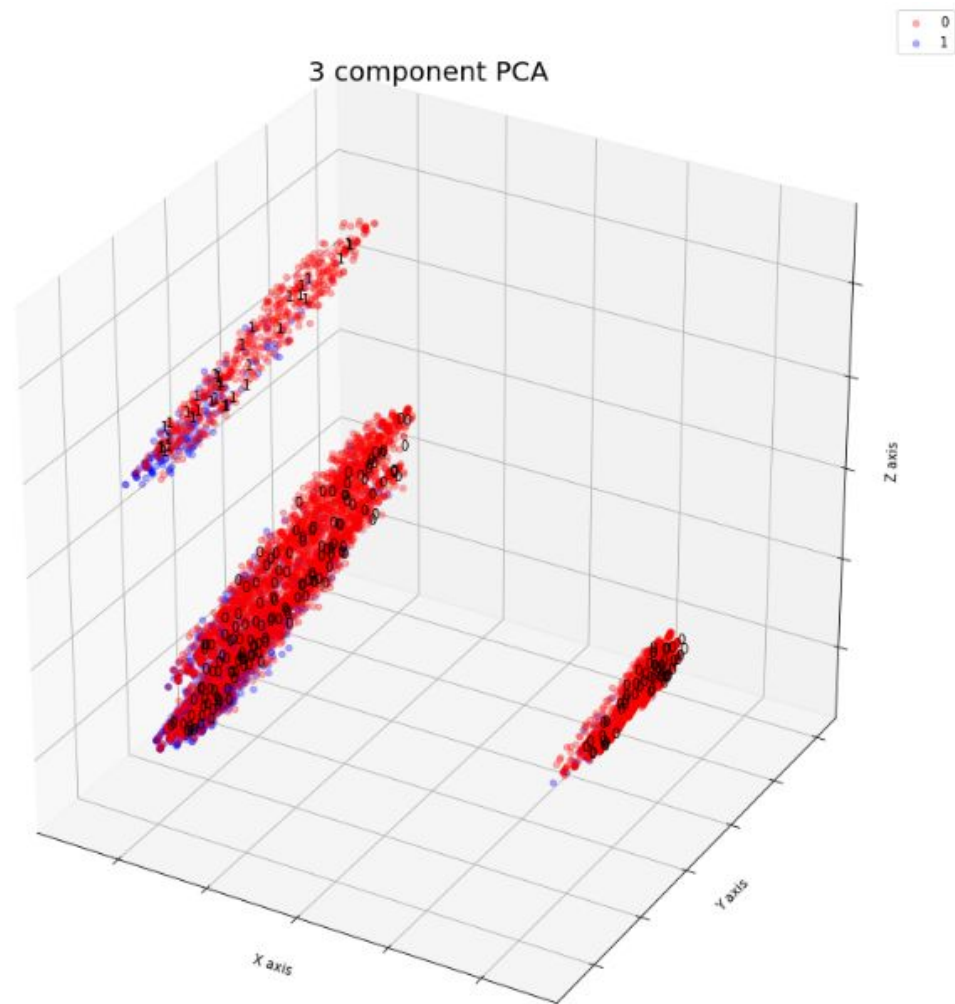
# Tenure



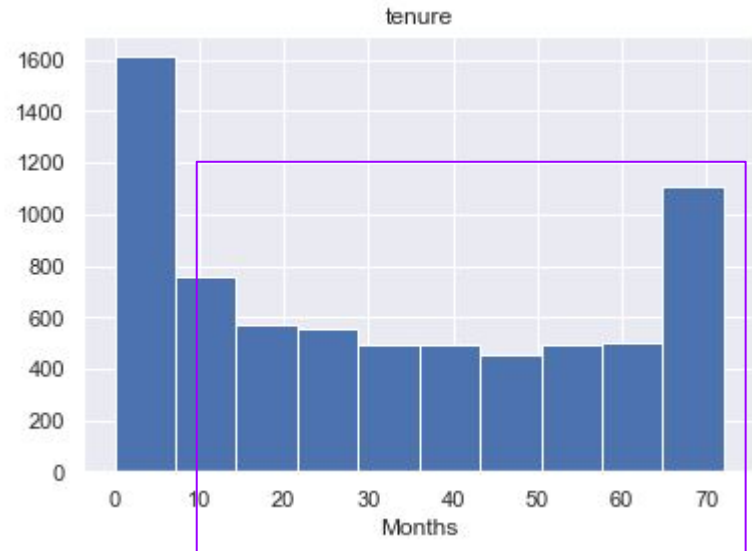
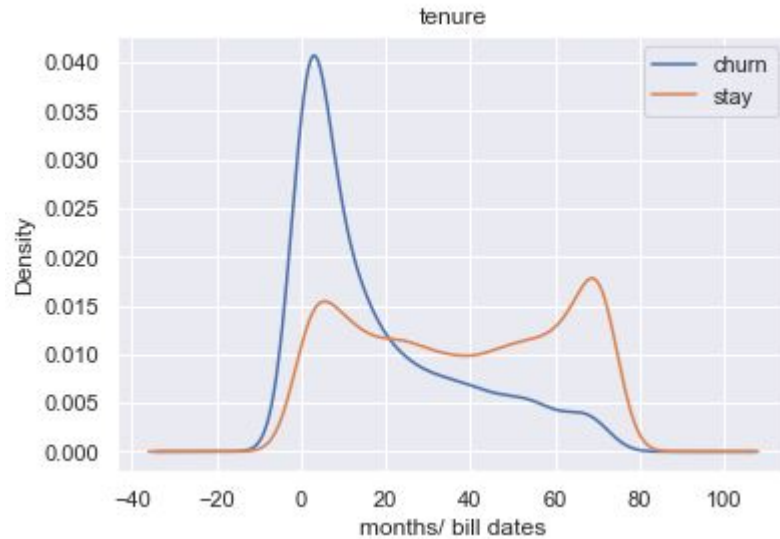
Accuracy: 0.7939798381371574

```
[[4690 967]  
 [ 484 902]]
```

	precision	recall	f1-score	support
0	0.83	0.91	0.87	5174
1	0.65	0.48	0.55	1869
avg / total	0.78	0.79	0.78	7043



# Customers for > 12 months



```
Churned/Total all customers 0.27
Churned/Total customers w. tenure > 12 0.17
Churned/Total customers w. tenure <= 12 0.47
```

## Using class\_weight='balanced'

Accuracy: 0.8439365863701873

```
[[3904 637]
 [ 121 195]]
```

	precision	recall	f1-score	support
0	0.86	0.97	0.91	4025
1	0.62	0.23	0.34	832
avg / total	0.82	0.84	0.81	4857

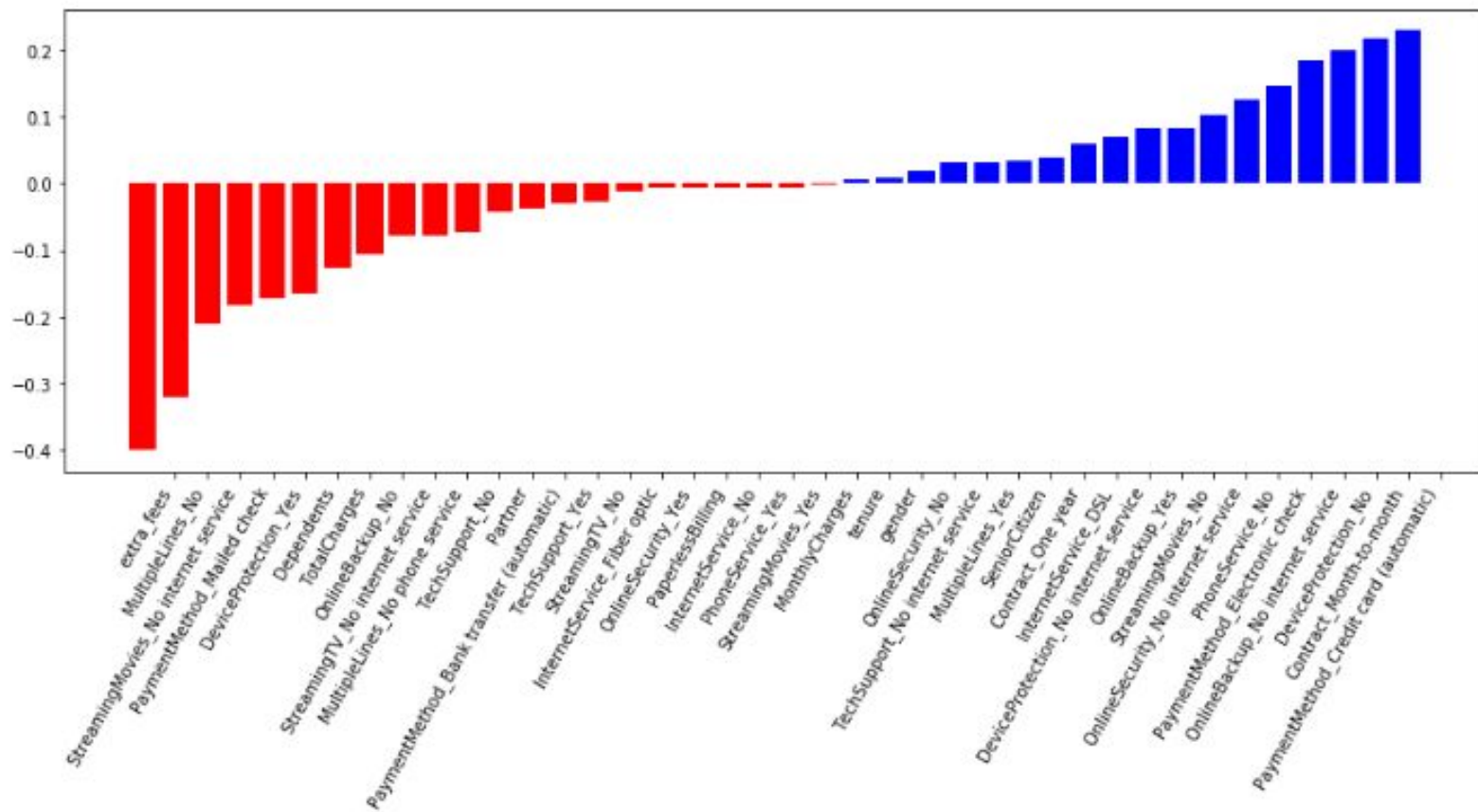
Accuracy: 0.7358451719168211

```
[[2927 185]
 [1098 647]]
```

	precision	recall	f1-score	support
0	0.94	0.73	0.82	4025
1	0.37	0.78	0.50	832
avg / total	0.84	0.74	0.77	4857

# Feature Selection

C=100

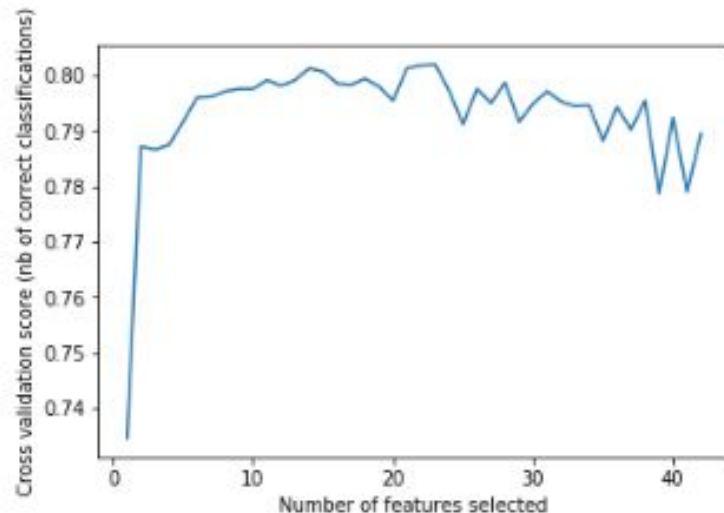




# Recursive Feature Elimination

```
# Create the RFE object and compute a cross-validated score.  
svc = LinearSVC(C=10)  
rfecv = RFECV(estimator=svc, step=1, cv=StratifiedKFold(3),  
              scoring='accuracy')  
rfecv.fit(X, y)
```

Optimal number of features : 23



## Features after RFE

```
Index(['Contract_Month-to-month', 'Contract_Two year', 'DeviceProtection_Yes',  
      'InternetService_DSL', 'InternetService_No', 'MultipleLines_No',  
      'MultipleLines_Yes', 'OnlineBackup_No internet service',  
      'OnlineSecurity_Yes', 'PaperlessBilling',  
      'PaymentMethod_Electronic check', 'PhoneService_No', 'PhoneService_Yes',  
      'StreamingMovies_No', 'StreamingMovies_No internet service',  
      'StreamingMovies_Yes', 'StreamingTV_No', 'StreamingTV_Yes',  
      'TechSupport_No internet service', 'TechSupport_Yes', 'TotalCharges',  
      'extra_fees', 'gender'],  
      dtype='object')
```

Accuracy: 0.7963935822802783

Linear SVM

```
[[4598  858]  
 [ 576 1011]]
```

	precision	recall	f1-score	support
0	0.84	0.89	0.87	5174
1	0.64	0.54	0.59	1869
avg / total	0.79	0.80	0.79	7043

# Final Results

{'C': 8, 'gamma': 0.001953125}

Accuracy: 0.8043447394576175

```
[[4718  922]
 [ 456  947]]
```

	precision	recall	f1-score	support
0	0.84	0.91	0.87	5174
1	0.67	0.51	0.58	1869
avg / total	0.79	0.80	0.79	7043

