Practica Nro 3 – parte 2 Modelamiento

Víctor Mariano Villacorta Plasencia

Configuración inicial:

Variables de experimento para la vecorización por tdf-idf:

```
min_df = 0.01,
max_df = 0.9,
```

Métrica de seguimiento:

```
scoring = 'accuracy',
```

a) Modelo 1 SVC (kernel: Lineal):

Mejores parametros de entrenamiento:

param_C	param_class_weight	param_probability	param_shrinking	mean_test_score	std_test_score	mean_train_score	std_train_score	
0.790604	balanced	True	True	0.922365	0.020387	0.948570	0.003468	I
0.790604	balanced	True	False	0.922365	0.020387	0.948570	0.003468	
0.790604	balanced	False	True	0.922365	0.020387	0.948570	0.003468	
0.790604	balanced	False	False	0.922365	0.020387	0.948570	0.003468	
0.372759	balanced	True	False	0.918265	0.008526	0.941415	0.005137	
0.372759	balanced	False	False	0.918265	0.008526	0.941415	0.005137	
0.372759	balanced	False	True	0.918265	0.008526	0.941415	0.005137	
0.372759	balanced	True	True	0.918265	0.008526	0.941415	0.005137	
1.151395	balanced	False	False	0.916914	0.018877	0.951635	0.003828	
1.151395	balanced	False	True	0.916914	0.018877	0.951635	0.003828	

Validación casos test ajenos al modelamiento:

```
Error de test

Número de clasificaciones erróneas de un total de 184 clasificaciones: 17
% de error: 9.239130434782608

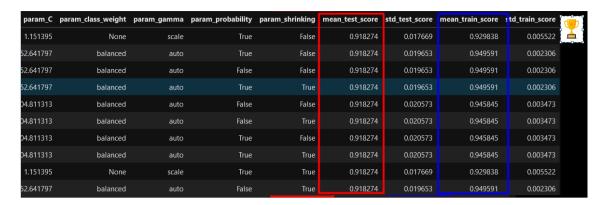
Matriz de confusión

Minsa_Peru DanielUrresti1

Minsa_Peru 100 4

DanielUrresti1 13 67
```

b) Modelo 1 SVC (kernel: sigmoid):



Validación casos test ajenos al modelamiento:



c) Modelo 1 SVC (kernel: polinómico):



Mismo resultados que el lineal , el auemnto de grados de polinomio no aporto mejores resultados



GANADOR:

SVC kernel SIGMOID

Su matriz de confusión con data test nueva es mejor (menor error)

