

## 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: Linear):

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	
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):

param_C	param_class_weight	param_gamma	param_probability	param_shrinking	mean_test_score	std_test_score	mean_train_score	std_train_score	
1.151395	None	scale	True	False	0.918274	0.017669	0.929838	0.005522	
52.641797	balanced	auto	True	False	0.918274	0.019653	0.949591	0.002306	
52.641797	balanced	auto	False	False	0.918274	0.019653	0.949591	0.002306	
52.641797	balanced	auto	True	True	0.918274	0.019653	0.949591	0.002306	
04.811313	balanced	auto	False	False	0.918274	0.020573	0.945845	0.003473	
04.811313	balanced	auto	False	True	0.918274	0.020573	0.945845	0.003473	
04.811313	balanced	auto	True	False	0.918274	0.020573	0.945845	0.003473	
04.811313	balanced	auto	True	True	0.918274	0.020573	0.945845	0.003473	
1.151395	None	scale	True	True	0.918274	0.017669	0.929838	0.005522	
52.641797	balanced	auto	False	True	0.918274	0.019653	0.949591	0.002306	

Validación casos test ajenos al modelamiento:

```


-----
Error de test
-----
Número de clasificaciones erróneas de un total de 184 clasificaciones: 16
% de error: 8.695652173913043

-----
Matriz de confusión
-----

```

	Minsa_Peru	DanielUrresti1
Minsa_Peru	101	3
DanielUrresti1	13	67

c) Modelo 1 SVC (kernel: polinómico):

param_C	param_class_weight	param_degree	param_gamma	param_probability	param_shrinking	mean_test_score	std_test_score	mean_train_score	std_train_score	
0.790604	balanced	1	scale	False	False	0.919635	0.021199	0.948911	0.003038	
0.790604	balanced	1	scale	True	True	0.919635	0.021199	0.948911	0.003038	
0.790604	balanced	1	scale	True	False	0.919635	0.021199	0.948911	0.003038	
0.790604	balanced	1	scale	False	True	0.919635	0.021199	0.948911	0.003038	
152.641797	balanced	1	auto	True	True	0.918274	0.019653	0.949591	0.002306	
104.811313	balanced	1	auto	True	True	0.918274	0.020573	0.945845	0.003473	
104.811313	balanced	1	auto	False	False	0.918274	0.020573	0.945845	0.003473	
152.641797	balanced	1	auto	False	False	0.918274	0.019653	0.949591	0.002306	
152.641797	balanced	1	auto	False	True	0.918274	0.019653	0.949591	0.002306	
152.641797	balanced	1	auto	True	False	0.918274	0.019653	0.949591	0.002306	

Mismo resultados que el lineal , el aumento de grados de polinomio no aporó mejores resultados

```

-----
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

GANADOR :  
SVC kernel SIGMOID

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

```
-----  
Error de test  
-----  
Número de clasificaciones erróneas de un total de 184 clasificaciones: 16  
% de error: 8.695652173913043  
  
-----  
Matriz de confusión  
-----
```

	Minsa_Peru	DanielUrresti1
Minsa_Peru	101	3
DanielUrresti1	13	67