

## Hoja de trabajo 5

### Naive Bayes

Variable objetivo, Grade, que representa la calificación de las notas de los estudiantes

- 'AA': Excelente
- 'BA': Bueno
- 'BB': Regular
- 'CB': Suficiente
- 'CC': Necesita mejorar
- 'DC': Deficiente
- 'DD': Muy deficiente
- 'Fail': Reprobado

Un total de 12 variables categóricas

	Sex	High_School_Type	Scholarship	Additional_Work	Sports_activity	Transportation	Attendance	Reading	Notes	Listening_in_Class	Project_work	Grade
Index												
0	Male	Other	50%	Yes	No	Private	Always	Yes	Yes	No	No	AA
1	Male	Other	50%	Yes	No	Private	Always	Yes	No	Yes	Yes	AA
2	Male	State	50%	No	No	Private	Never	No	No	No	Yes	AA
3	Female	Private	50%	Yes	No	Bus	Always	No	Yes	No	No	AA
4	Male	Private	50%	No	No	Bus	Always	Yes	No	Yes	Yes	AA

Un total de 3 variables numéricas

	Id	Student_Age	Weekly_Study_Hours
Index			
0	5001	21	0
1	5002	20	0
2	5003	21	2
3	5004	18	2
4	5005	22	12

Eficiencia del algoritmo

## Ver precision

```
115]: from sklearn.metrics import accuracy_score
```

```
acc = accuracy_score(y_test, y_pred)
```

```
print("La precision del modelo es de: ", acc)
```

```
La precision del modelo es de: 0.13636363636363635
```

```
116]: y_pred_train = gnb.predict(X_train)
```

```
y_pred_train
```

```
116]: array(['CC', 'DD', 'CB', 'CB', 'CC', 'DD', 'CB', 'CB', 'DD', 'Fail', 'DC',  
          'CB', 'CC', 'Fail', 'DD', 'CC', 'Fail', 'DD', 'DC', 'DC', 'CB',  
          'CC', 'Fail', 'Fail', 'Fail', 'CC', 'CC', 'Fail', 'CC', 'Fail',  
          'DD', 'CC', 'DD', 'CC', 'Fail', 'DC', 'DD', 'CB', 'Fail', 'CC',  
          'Fail', 'Fail', 'CC', 'BA', 'Fail', 'CC', 'CC', 'CC', 'Fail',  
          'Fail', 'DC', 'CB', 'DD', 'CC', 'DD', 'DC', 'CB', 'CB', 'DD', 'CC',  
          'CB', 'CB', 'BB', 'DD', 'Fail', 'CC', 'BA', 'BA', 'DD', 'CC', 'BA',  
          'CB', 'CC', 'DD', 'CC', 'BA', 'CC', 'Fail', 'Fail', 'Fail', 'CC',  
          'CC', 'Fail', 'DC', 'DD', 'CC', 'DD', 'Fail', 'DC', 'Fail', 'DD',  
          'DC', 'DC', 'DD', 'CB', 'CC', 'Fail', 'DD', 'DD', 'Fail', 'CC'],  
          dtype='<U4')
```

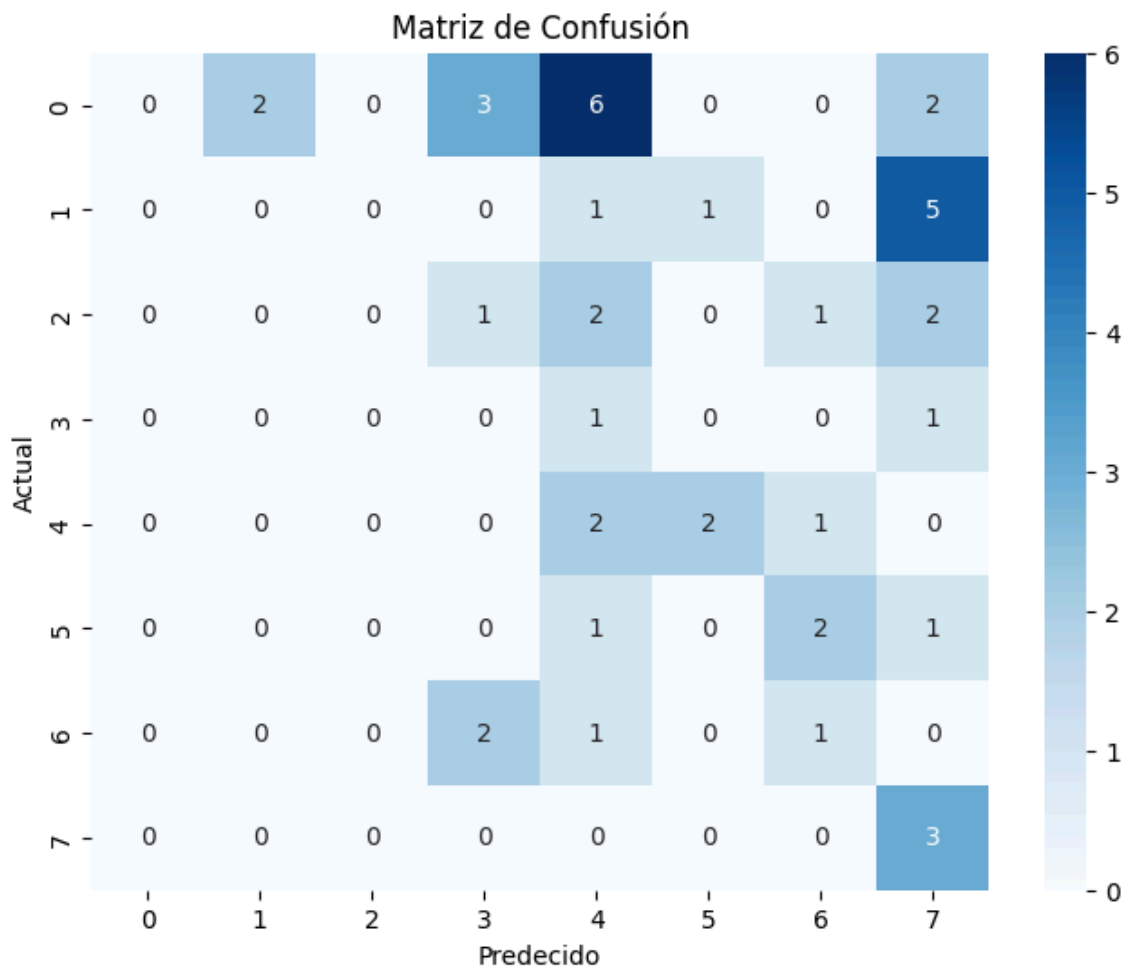
```
117]: acc = accuracy_score(y_train , y_pred_train)
```

```
print("La precision del modelo es de: ", acc)
```

```
La precision del modelo es de: 0.31683168316831684
```

El modelo creado presenta solo un 32% de eficacia, un rendimiento notablemente bajo, pero casi el doble de lo que se consiguió anteriormente, lo cual fue 14%. Se sugiere que el error principal podría residir en la cantidad de datos utilizados o en la construcción misma del modelo.

## Overfitting



### Reporte de Clasificación:

	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	8
1.0	0.00	0.00	0.00	17
2.0	0.00	0.00	0.00	13
3.0	0.00	0.00	0.00	17
4.0	0.06	0.90	0.12	10
5.0	0.00	0.00	0.00	21
6.0	0.00	0.00	0.00	24
7.0	0.00	0.00	0.00	35
accuracy			0.06	145
macro avg	0.01	0.11	0.01	145
weighted avg	0.00	0.06	0.01	145