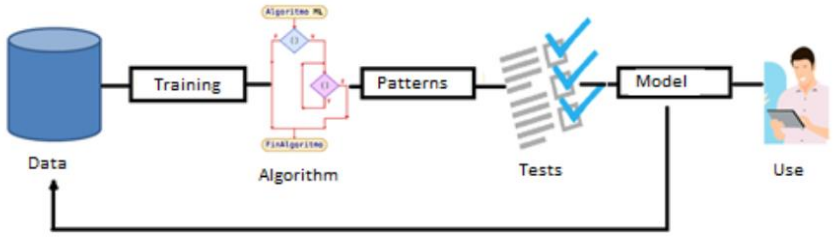
Aprendizaje de Máquina (Machine Learning)

Profesor: Víctor Viera Balanta

Clasificador Naive Bayes



Google Colab

```
[40] #Adaptado por Víctor Viera Balanta
     #se importan las bibliotecas
     #para datos números vectores,matrices
     import numpy as np
     #manejo de data Frames
     import pandas as pd
```

```
#biblioteca de extracción de características del texto
from sklearn.feature_extraction.text import CountVectorizer
#para medir la precisión del modelo, matriz de confusión
from sklearn.metrics import precision_score,confusion_matrix
#divide el conjunto de datos de entrenamiento
from sklearn.model_selection import train_test_split
#Modelo de Naive Bayes, El clasificador multinomial Naive Bayes es adecuado para la clasificación
#con características discretas (por ejemplo, recuento de palabras para la clasificación de texto)
from sklearn.naive_bayes import MultinomialNB
```

```
[42] #carga los datos
     data = pd.read_csv('https://raw.githubusercontent.com/AiDevNepal/ai-saturdays-workshop-8/master/data/spam.csv')
```

#Copiar esta ruta, donde están los datos de entrenamiento
'https://raw.githubusercontent.com/AiDevNepal/ai-saturdays-workshop-8/master/data/spam.csv'

```
#convierte los datos de texto a número (Spam 1, NoSpam 0)
data['target'] = np.where(data['target']=='spam',1, 0)
```

```
print(data)
```

Debe salir

```
                                                   text  target
0     Go until jurong point, crazy.. Available only ...       0
1                         Ok lar... Joking wif u oni...       0
2     Free entry in 2 a wkly comp to win FA Cup fina...       1
3     U dun say so early hor... U c already then say...       0
4     Nah I don't think he goes to usf, he lives aro...       0
...                                                 ...     ...
5567  This is the 2nd time we have tried 2 contact u...       1
5568              Will Ì_ b going to esplanade fr home?       0
5569  Pity, * was in mood for that. So...any other s...       0
5570  The guy did some bitching but I acted like i'd...       0
5571                        Rofl. Its true to its name       0

[5572 rows x 2 columns]
```

```
#civde el conjunto de datos en entranemiento(X_train,X_test) y prueba(X_test,Y_test)
X_train, X_test, Y_train, Y_test = train_test_split(data['text'],
                                                    data['target'],
                                                    random_state=0)
```

```
# extrear las características
vectorizer = CountVectorizer(ngram_range=(1, 2)).fit(X_train)
X_train_vectorized = vectorizer.transform(X_train)
```

```
#se acomodan los datos para introducirlos
#el parámetro alfa es lo que se conoce como hiperparámetro; es decir,
#un parámetro que controla la forma del modelo en sí
#El suavizado de Laplace es una técnica utilizada para evitar problemas de probabilidad nula en el modelo de Naive Bayes.
#puede influir en el rendimiento y la generalización del clasificador.(laplace)
model = MultinomialNB(alpha=0.1)
model.fit(X_train_vectorized, Y_train)
```

```
#calcula la exactitud del modelo
predictions = model.predict(vectorizer.transform(X_test))
print("Accuracy:", 100 * sum(predictions == print("Accuracy:", 100 * sum(predictions == Y_test) / len(predictions), '%')
) / len(predictions), '%')
```

Debe salir

```
Accuracy: 98.99497487437186 %
Accuracy: 0.0 %
```

```
#cómo se comporta de acuerdo a la matriz de confusión
#la diagonal principal indica los aciertos, los datos equivocados (3,11)
confusion_matrix(predictions, Y_test)
```

Debe salir

```
array([[1193,   11],
       [   3,  186]])
```

Abrir una nueva celda de código y copiar, para probar

```
#se colocan textos de ejemplo
model.predict(vectorizer.transform(
  [
    "Thank you, ABC. Can you also share your LinkedIn profile? As you are a good at programming at pyhthon, would be willing to see your personal/college projects.",
    "Hi y'all, We have a Job Openings in the positions of software engineer, IT officer at ABC Company.Kindly, send us your resume and the cover letter as soon as possible if you think you are an eligible candidate and meet the criteria.",
    "Dear ABC, Congratulations! You have been selected as a SOftware Developer at XYZ Company. We were really happy to see your enthusiasm for this vision and mission. We are impressed with your background and we think you would make an excellent addition to the team.",
  ]))
```

Debe salir

```
array([0, 0, 0])
```

Abrir una nueva línea de código y copiar

```
#texto de ejemplo
model.predict(vectorizer.transform(
  [
    "congratulations, you became today's lucky winner",
    "1-month unlimited calls offer Activate now",
    "Ram wants your phone number",

  ]))
```

Debe salir

```
array([1, 1, 1])
```