



Instituto Politécnico Nacional

ESCOM

PLN

Islas Osorio Enrique

7CM3

Practice IV - Text classification

## Introducción.

En esta práctica para clasificar texto primero se carga un corpus de texto en este caso `fetch_20newsgroups`, luego se divide en conjuntos de entrenamiento y prueba. El texto se normaliza para eliminar el ruido, y luego se crea una representación del texto. Finalmente, se utiliza un método de aprendizaje automático para entrenar un modelo y predecir las instancias de prueba.

En la practica vamos a normalizar los textos con:

**Tokenización:** Es el proceso de dividir un texto en unidades significativas, llamadas tokens. Los tokens pueden ser palabras, números, símbolos o cualquier otra unidad de texto que tenga un significado específico para el contexto en el que se utiliza.

**Limpieza de texto:** Es el proceso de eliminar del texto elementos no deseados, como errores tipográficos, caracteres especiales, símbolos, stop words, etc. La limpieza de texto es importante para mejorar la calidad de los datos y la precisión de las tareas de procesamiento del lenguaje natural (NLP).

**Stop words:** Son palabras que aparecen con mucha frecuencia en un texto, pero que no tienen mucho significado para el contexto en el que se utilizan. Ejemplos de stop words en español son: "el", "la", "los", "las", "de", "a", "para", "en", "y", "que", etc. Las stop words suelen eliminarse del texto antes de realizar tareas de NLP, ya que pueden interferir con el rendimiento de los modelos.

**Lematización:** Es el proceso de reducir una palabra a su forma de base, o lexema. El lexema es la forma más básica de una palabra, sin sus afijos derivativos o flexivos. La lematización es importante para mejorar la precisión de las tareas de NLP, ya que permite agrupar las palabras que tienen el mismo significado, aunque estén flexionadas de forma diferente.

En la practica se hicieron pruebas sin normalizar, tokenización + stopwords, tokenización + lematización, tokenización + stopwords + lematización y , tokenización + stopwords + lematización + cleantext

También haremos uso de diferentes representaciones de texto.

**Binarización:** Representación de texto en la que cada palabra se representa como un número binario que indica si la palabra aparece en el texto o no.

**Frecuencia:** Representación de texto en la que cada palabra se representa como un número entero que indica la frecuencia con la que la palabra aparece en el texto.

**TF-IDF:** Representación de texto que combina la frecuencia y la importancia de las palabras.

Como vimos en la practica anterior, TF-IDF es el que mejor resultado dio

Finalmente, varios algoritmos de Machine Learning.

Logistic regression es un algoritmo de aprendizaje supervisado que se utiliza para problemas de clasificación binaria. Es un modelo lineal que predice la probabilidad de que una muestra pertenezca a una clase determinada.

Multinomial Naive Bayes es un algoritmo de aprendizaje supervisado que se utiliza para problemas de clasificación multiclase. Es un modelo probabilístico que asume que las variables predictoras son independientes entre sí.

K-nearest neighbors (KNN) es un algoritmo de aprendizaje no supervisado que se utiliza para problemas de clasificación, regresión y agrupamiento. KNN funciona comparando una nueva muestra con las muestras más cercanas en el conjunto de entrenamiento.

Random forest es un algoritmo de aprendizaje supervisado que combina múltiples árboles de decisión para mejorar la precisión. Random forest funciona entrenando un conjunto de árboles de decisión en subconjuntos aleatorios del conjunto de entrenamiento.

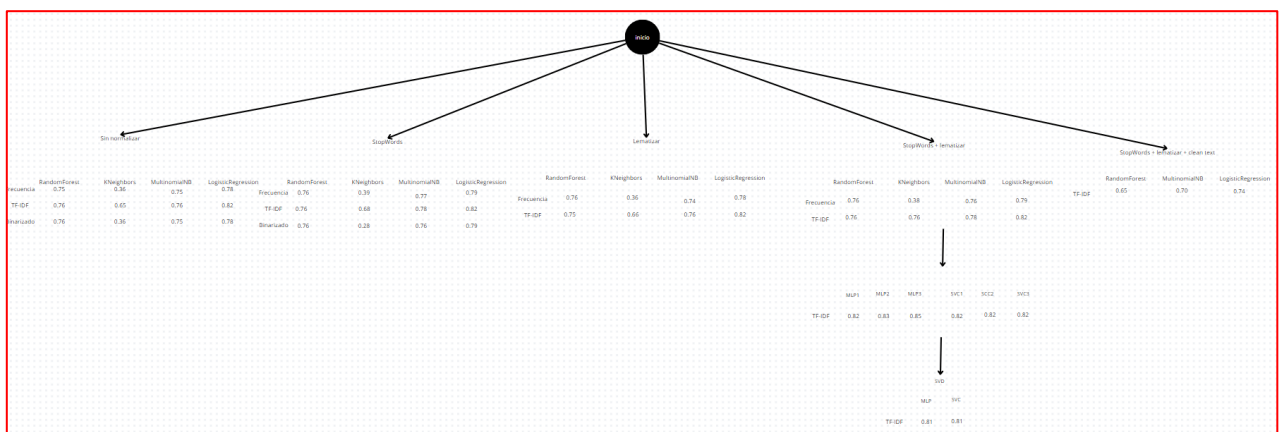
Multilayer perceptron (MLP) es un algoritmo de aprendizaje supervisado que se utiliza para problemas de clasificación y regresión. MLP es un tipo de red neuronal artificial que consta de una o más capas ocultas de neuronas.

## RESULTADOS

Este es mi intento de árbol como el que vimos en clase, primero probe sin normalizar y tokenizacion + stopwords con los 3 tipos de representaciones, como el binario era el peor, en los siguientes no lo tome en cuenta; lo siguiente que probe fue tokenizacion + lematizar y tokenizacion + stopwords + lematizar como KNN es el algoritmo de ML que peor resultados daba no lo tome en cuenta para probar con una normalización completa tokenizacion + stopwords + lematizar + clean text, no me dio buen resultado por lo que me quede con tokenizacion + stopwords + lematizar que fue lo que mejor resultado me dio.

Con tokenizacion + stopwords + lematizar probe los algoritmos de ML que más se tardaba que eran MLP y SVC estos los probe con diferentes parámetros y el que mejor resultado dio fue con MLP.

Finalmente use SVD con tokenizacion + stopwords + lematizar, TFIDF y MLP,SVC con los mejores parámetros



Sin normalizar					tokenizar + StopWords				
	RandomForest	KNeighbors	MultinomialNB	LogisticRegression		RandomForest	KNeighbors	MultinomialNB	LogisticRegression
Frecuencia	0.75	0.36	0.75	0.78	Frecuencia	0.76	0.39	0.77	0.79
TF-IDF	0.76	0.65	0.76	0.82	TF-IDF	0.76	0.68	0.78	0.82
Binarizado	0.76	0.36	0.75	0.78	Binarizado	0.76	0.28	0.76	0.79

tokenizar + Lematizar					tokenizar + StopWords + lematizar				
	RandomForest	KNeighbors	MultinomialNB	LogisticRegression		RandomForest	KNeighbors	MultinomialNB	LogisticRegression
Frecuencia	0.76	0.36	0.74	0.78	Frecuencia	0.76	0.38	0.76	0.79
TF-IDF	0.75	0.66	0.76	0.82	TF-IDF	0.76	0.76	0.78	0.82

tokenizar + StopWords + lematizar + clean text			
	RandomForest	MultinomialNB	LogisticRegression
TF-IDF	0.65	0.70	0.74

tokenizar + StopWords + lematizar

	RandomForest	KNeighbors	MultinomialNB	LogisticRegression
Frecuencia	0.76	0.38	0.76	0.79
TF-IDF	0.76	0.76	0.78	0.82



	MLP1	MLP2	MLP3	SVC1	SCC2	SVC3
TF-IDF	0.82	0.83	0.85	0.82	0.82	0.82



	SVC
TF-IDF	0.81

## Conclusión

Al realizar varias combinaciones, la mejor combinación fue tokenización + stopwords + lematizar, representación TF-IDF y MLP con estos parámetros

```
clf = MLPClassifier(  
    hidden_layer_sizes=(100,),  
    activation="relu",  
    solver="adam",  
    alpha=0.0001,  
    learning_rate="constant",  
    max_iter=200,  
    random_state=42,  
    early_stopping=True,  
    verbose=True
```

Nos dio un valor de 0.85, que fue el más alto. Como vimos en la práctica anterior, TF-IDF es una muy buena representación de texto, ya que le da valor a las palabras más importantes. Solo lematizar el texto no daba una gran mejora en comparación con solo quitar stopwords, pero al combinarlos dio un muy buen resultado, ya que se elimina el ruido. Cuando lo junté con cleantext, dio un peor resultado. Esto pudo ser porque tal vez quité cosas que eran importantes y no lo sabía. Finalmente, el algoritmo de MLP fue el que mejor resultado dio, porque es un buen algoritmo de aprendizaje automático.

Estas son las capturas de los resultados para que no se me perdieran

```
*****Logistic Regression*****
[ 7  1  0 ...  9 12 15]
      precision    recall  f1-score   support

     0       0.80      0.74      0.77       319
     1       0.69      0.79      0.74       389
     2       0.75      0.73      0.74       394
     3       0.72      0.72      0.72       392
     4       0.81      0.83      0.82       385
     5       0.83      0.74      0.78       395
     6       0.76      0.90      0.82       390
     7       0.90      0.89      0.90       396
     8       0.95      0.95      0.95       398
     9       0.88      0.92      0.90       397
    10       0.94      0.95      0.95       399
    11       0.94      0.88      0.91       396
    12       0.76      0.80      0.78       393
    13       0.89      0.83      0.85       396
    14       0.91      0.92      0.91       394
    15       0.81      0.94      0.87       398
    16       0.72      0.88      0.79       364
    17       0.96      0.87      0.92       376
    18       0.76      0.59      0.66       310
    19       0.81      0.49      0.61       251

 accuracy                   0.83       7532
 macro avg                   0.83      0.82      0.82       7532
 weighted avg                 0.83      0.83      0.83       7532
```

Este primero fue sin normalizar ni nada, se uso representación de texto TF-IDF y Logistic regression

Igual sin nada con TF-IDF

```

*****MultinomialNB*****
[ 7 11 0 ... 9 3 15]
      precision    recall  f1-score   support

     0       0.80      0.52      0.63       319
     1       0.81      0.65      0.72       389
     2       0.82      0.65      0.73       394
     3       0.67      0.78      0.72       392
     4       0.86      0.77      0.81       385
     5       0.89      0.75      0.82       395
     6       0.93      0.69      0.80       390
     7       0.85      0.92      0.88       396
     8       0.94      0.93      0.93       398
     9       0.92      0.90      0.91       397
    10       0.89      0.97      0.93       399
    11       0.59      0.97      0.74       396
    12       0.84      0.60      0.70       393
    13       0.92      0.74      0.82       396
    14       0.84      0.89      0.87       394
    15       0.44      0.98      0.61       398
    16       0.64      0.94      0.76       364
    17       0.93      0.91      0.92       376
    18       0.96      0.42      0.58       310
    19       0.97      0.14      0.24       251

 accuracy                   0.77       7532
 macro avg                  0.83      0.76      0.76       7532
 weighted avg              0.82      0.77      0.77       7532

```



```

*****KNeighborsClassifier*****
[12 11  0 ...  2 19 15]
      precision    recall  f1-score   support

      0         0.43      0.76      0.55        319
      1         0.50      0.61      0.55        389
      2         0.56      0.57      0.57        394
      3         0.53      0.58      0.56        392
      4         0.59      0.56      0.57        385
      5         0.69      0.60      0.64        395
      6         0.58      0.45      0.51        390
      7         0.75      0.69      0.72        396
      8         0.84      0.81      0.82        398
      9         0.77      0.72      0.74        397
     10         0.85      0.84      0.84        399
     11         0.76      0.84      0.80        396
     12         0.70      0.50      0.58        393
     13         0.82      0.49      0.62        396
     14         0.79      0.76      0.78        394
     15         0.75      0.76      0.76        398
     16         0.70      0.73      0.72        364
     17         0.62      0.76      0.69        376
     18         0.55      0.61      0.58        310
     19         0.56      0.49      0.52        251

 accuracy                   0.66       7532
 macro avg                 0.67       7532
 weighted avg              0.67       7532

```

*****SVC*****				
	precision	recall	f1-score	support
0	0.83	0.71	0.76	319
1	0.62	0.82	0.71	389
2	0.80	0.70	0.75	394
3	0.73	0.78	0.76	392
4	0.82	0.83	0.82	385
5	0.83	0.73	0.77	395
6	0.73	0.91	0.81	390
7	0.90	0.87	0.89	396
8	0.96	0.93	0.95	398
9	0.88	0.91	0.90	397
10	0.97	0.91	0.94	399
11	0.96	0.85	0.90	396
12	0.65	0.85	0.74	393
13	0.88	0.78	0.82	396
14	0.93	0.88	0.90	394
15	0.79	0.92	0.85	398
16	0.75	0.88	0.81	364
17	0.97	0.81	0.88	376
18	0.75	0.59	0.66	310
19	0.79	0.55	0.65	251
accuracy			0.82	7532
macro avg			0.83	7532
weighted avg			0.83	7532

Frecuencia

*****Logistic Regression*****				
	precision	recall	f1-score	support
0	0.72	0.73	0.72	319
1	0.68	0.74	0.71	389
2	0.72	0.66	0.69	394
3	0.68	0.67	0.68	392
4	0.74	0.81	0.77	385
5	0.81	0.71	0.76	395
6	0.81	0.88	0.85	390
7	0.83	0.84	0.84	396
8	0.90	0.92	0.91	398
9	0.85	0.88	0.87	397
10	0.91	0.93	0.92	399
11	0.90	0.87	0.88	396
12	0.68	0.73	0.71	393
13	0.85	0.71	0.77	396
14	0.89	0.89	0.89	394
15	0.82	0.94	0.87	398
16	0.68	0.85	0.76	364
17	0.94	0.74	0.83	376
18	0.69	0.55	0.61	310
19	0.63	0.60	0.61	251
accuracy			0.79	7532
macro avg			0.79	7532
weighted avg			0.79	7532

```

*****MultinomialNB*****
      precision    recall  f1-score   support

     0       0.79       0.77       0.78        319
     1       0.67       0.74       0.70        389
     2       0.20       0.00       0.01        394
     3       0.56       0.77       0.65        392
     4       0.84       0.75       0.79        385
     5       0.65       0.84       0.73        395
     6       0.93       0.65       0.77        390
     7       0.87       0.91       0.89        396
     8       0.96       0.92       0.94        398
     9       0.96       0.87       0.91        397
    10       0.93       0.96       0.95        399
    11       0.67       0.95       0.78        396
    12       0.79       0.66       0.72        393
    13       0.87       0.82       0.85        396
    14       0.83       0.89       0.86        394
    15       0.70       0.96       0.81        398
    16       0.69       0.91       0.79        364
    17       0.85       0.94       0.89        376
    18       0.58       0.63       0.60        310
    19       0.89       0.33       0.49        251

 accuracy          0.77        7532
  macro avg       0.76        0.76       0.75        7532
 weighted avg     0.76        0.77       0.75        7532

```

\*\*\*\*\*KNeighborsClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.31	0.49	0.38	319
1	0.14	0.47	0.22	389
2	0.26	0.33	0.29	394
3	0.38	0.27	0.32	392
4	0.22	0.31	0.25	385
5	0.44	0.18	0.26	395
6	0.58	0.38	0.46	390
7	0.26	0.29	0.27	396
8	0.68	0.44	0.54	398
9	0.39	0.30	0.34	397
10	0.43	0.44	0.43	399
11	0.52	0.52	0.52	396
12	0.43	0.12	0.19	393
13	0.40	0.18	0.24	396
14	0.78	0.31	0.44	394
15	0.30	0.57	0.40	398
16	0.52	0.36	0.43	364
17	0.47	0.56	0.51	376
18	0.49	0.28	0.36	310
19	0.41	0.22	0.28	251
accuracy			0.35	7532
macro avg	0.42	0.35	0.36	7532
weighted avg	0.42	0.35	0.36	7532

*****RandomForestClassifier*****				
	precision	recall	f1-score	support
0	0.72	0.63	0.67	319
1	0.54	0.74	0.63	389
2	0.67	0.75	0.70	394
3	0.67	0.67	0.67	392
4	0.71	0.78	0.74	385
5	0.79	0.67	0.73	395
6	0.75	0.93	0.83	390
7	0.84	0.80	0.82	396
8	0.92	0.91	0.91	398
9	0.82	0.89	0.85	397
10	0.89	0.92	0.90	399
11	0.89	0.91	0.90	396
12	0.66	0.51	0.57	393
13	0.87	0.65	0.75	396
14	0.83	0.88	0.86	394
15	0.69	0.94	0.80	398
16	0.66	0.88	0.75	364
17	0.93	0.80	0.86	376
18	0.90	0.47	0.61	310
19	0.78	0.29	0.42	251
accuracy			0.76	7532
macro avg	0.78	0.75	0.75	7532
weighted avg	0.78	0.76	0.76	7532

BINARIO

*****Logistic Regression*****				
	precision	recall	f1-score	support
0	0.72	0.73	0.72	319
1	0.68	0.74	0.71	389
2	0.72	0.66	0.69	394
3	0.68	0.67	0.68	392
4	0.74	0.81	0.77	385
5	0.81	0.71	0.76	395
6	0.81	0.88	0.85	390
7	0.83	0.84	0.84	396
8	0.90	0.92	0.91	398
9	0.85	0.88	0.87	397
10	0.91	0.93	0.92	399
11	0.90	0.87	0.88	396
12	0.68	0.73	0.71	393
13	0.85	0.71	0.77	396
14	0.89	0.89	0.89	394
15	0.82	0.94	0.87	398
16	0.68	0.85	0.76	364
17	0.94	0.74	0.83	376
18	0.69	0.55	0.61	310
19	0.63	0.60	0.61	251
accuracy			0.79	7532
macro avg			0.79	7532
weighted avg			0.79	7532

```

*****MultinomialNB*****
      precision    recall  f1-score   support

     0       0.79      0.77      0.78        319
     1       0.67      0.74      0.70        389
     2       0.20      0.00      0.01        394
     3       0.56      0.77      0.65        392
     4       0.84      0.75      0.79        385
     5       0.65      0.84      0.73        395
     6       0.93      0.65      0.77        390
     7       0.87      0.91      0.89        396
     8       0.96      0.92      0.94        398
     9       0.96      0.87      0.91        397
    10       0.93      0.96      0.95        399
    11       0.67      0.95      0.78        396
    12       0.79      0.66      0.72        393
    13       0.87      0.82      0.85        396
    14       0.83      0.89      0.86        394
    15       0.70      0.96      0.81        398
    16       0.69      0.91      0.79        364
    17       0.85      0.94      0.89        376
    18       0.58      0.63      0.60        310
    19       0.89      0.33      0.49        251

 accuracy          0.77      7532
  macro avg       0.76      0.76      0.75      7532
 weighted avg     0.76      0.77      0.75      7532

```



\*\*\*\*\*KNeighborsClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.31	0.49	0.38	319
1	0.14	0.47	0.22	389
2	0.26	0.33	0.29	394
3	0.38	0.27	0.32	392
4	0.22	0.31	0.25	385
5	0.44	0.18	0.26	395
6	0.58	0.38	0.46	390
7	0.26	0.29	0.27	396
8	0.68	0.44	0.54	398
9	0.39	0.30	0.34	397
10	0.43	0.44	0.43	399
11	0.52	0.52	0.52	396
12	0.43	0.12	0.19	393
13	0.40	0.18	0.24	396
14	0.78	0.31	0.44	394
15	0.30	0.57	0.40	398
16	0.52	0.36	0.43	364
17	0.47	0.56	0.51	376
18	0.49	0.28	0.36	310
19	0.41	0.22	0.28	251
accuracy			0.35	7532
macro avg	0.42	0.35	0.36	7532
weighted avg	0.42	0.35	0.36	7532

```

*****RandomForestClassifier*****
      precision    recall  f1-score   support

     0       0.69      0.65      0.67       319
     1       0.57      0.69      0.62       389
     2       0.67      0.76      0.71       394
     3       0.66      0.66      0.66       392
     4       0.75      0.78      0.76       385
     5       0.79      0.70      0.74       395
     6       0.74      0.92      0.82       390
     7       0.82      0.81      0.82       396
     8       0.91      0.91      0.91       398
     9       0.79      0.93      0.86       397
    10       0.91      0.90      0.90       399
    11       0.86      0.91      0.89       396
    12       0.68      0.53      0.60       393
    13       0.84      0.69      0.76       396
    14       0.84      0.88      0.86       394
    15       0.72      0.92      0.81       398
    16       0.68      0.87      0.76       364
    17       0.95      0.80      0.87       376
    18       0.86      0.48      0.61       310
    19       0.84      0.37      0.51       251

 accuracy          0.77       7532
 macro avg         0.78       0.76      0.76       7532
 weighted avg      0.78       0.77      0.76       7532

```

Sin stopwords y TFIDF

*****Logistic Regression*****				
	precision	recall	f1-score	support
0	0.79	0.74	0.77	319
1	0.69	0.79	0.74	389
2	0.75	0.74	0.75	394
3	0.72	0.73	0.72	392
4	0.81	0.83	0.82	385
5	0.84	0.74	0.78	395
6	0.75	0.89	0.82	390
7	0.91	0.89	0.90	396
8	0.95	0.95	0.95	398
9	0.89	0.93	0.91	397
10	0.95	0.95	0.95	399
11	0.94	0.88	0.90	396
12	0.75	0.80	0.77	393
13	0.89	0.83	0.86	396
14	0.90	0.92	0.91	394
15	0.81	0.94	0.87	398
16	0.72	0.88	0.79	364
17	0.96	0.87	0.91	376
18	0.77	0.59	0.67	310
19	0.81	0.50	0.62	251
accuracy			0.83	7532
macro avg			0.83	7532
weighted avg			0.83	7532

*****MultinomialNB*****				
	precision	recall	f1-score	support
0	0.80	0.52	0.63	319
1	0.81	0.65	0.72	389
2	0.82	0.65	0.72	394
3	0.67	0.78	0.72	392
4	0.86	0.78	0.82	385
5	0.90	0.76	0.82	395
6	0.93	0.70	0.80	390
7	0.85	0.92	0.88	396
8	0.93	0.93	0.93	398
9	0.93	0.91	0.92	397
10	0.89	0.97	0.93	399
11	0.60	0.97	0.74	396
12	0.84	0.60	0.70	393
13	0.92	0.74	0.82	396
14	0.85	0.90	0.87	394
15	0.44	0.98	0.61	398
16	0.65	0.93	0.76	364
17	0.93	0.91	0.92	376
18	0.96	0.42	0.58	310
19	0.97	0.14	0.24	251
accuracy			0.78	7532
macro avg	0.83	0.76	0.76	7532
weighted avg	0.82	0.78	0.77	7532

```

*****KNeighborsClassifier*****
      precision    recall  f1-score   support

     0         0.42      0.76      0.54        319
     1         0.52      0.62      0.57        389
     2         0.56      0.56      0.56        394
     3         0.54      0.59      0.56        392
     4         0.59      0.56      0.57        385
     5         0.69      0.59      0.64        395
     6         0.58      0.46      0.51        390
     7         0.77      0.69      0.73        396
     8         0.84      0.81      0.82        398
     9         0.77      0.72      0.75        397
    10         0.84      0.84      0.84        399
    11         0.76      0.84      0.80        396
    12         0.69      0.49      0.57        393
    13         0.82      0.49      0.62        396
    14         0.79      0.77      0.78        394
    15         0.75      0.75      0.75        398
    16         0.70      0.73      0.71        364
    17         0.63      0.76      0.69        376
    18         0.54      0.61      0.58        310
    19         0.55      0.49      0.51        251

 accuracy          0.66        7532
  macro avg         0.67        0.66      0.66        7532
 weighted avg         0.67        0.66      0.66        7532

```

```

*****RandomForestClassifier*****
      precision    recall  f1-score   support

     0       0.71      0.64      0.67       319
     1       0.58      0.69      0.63       389
     2       0.65      0.76      0.70       394
     3       0.65      0.64      0.64       392
     4       0.71      0.73      0.72       385
     5       0.74      0.69      0.71       395
     6       0.74      0.92      0.82       390
     7       0.82      0.80      0.81       396
     8       0.89      0.89      0.89       398
     9       0.81      0.90      0.85       397
    10       0.89      0.93      0.91       399
    11       0.88      0.92      0.90       396
    12       0.64      0.48      0.55       393
    13       0.85      0.68      0.75       396
    14       0.84      0.89      0.86       394
    15       0.70      0.91      0.79       398
    16       0.67      0.86      0.75       364
    17       0.95      0.80      0.86       376
    18       0.86      0.48      0.61       310
    19       0.78      0.33      0.46       251

 accuracy          0.76       7532
 macro avg         0.77       0.75      0.75       7532
 weighted avg      0.77       0.76      0.75       7532

```

Stopwrodsenglish and TFIDF

*****Logistic Regression*****					
	precision	recall	f1-score	support	
0	0.80	0.73	0.76	319	
1	0.69	0.79	0.74	389	
2	0.76	0.75	0.76	394	
3	0.70	0.73	0.71	392	
4	0.81	0.83	0.82	385	
5	0.82	0.75	0.78	395	
6	0.78	0.88	0.83	390	
7	0.91	0.89	0.90	396	
8	0.95	0.95	0.95	398	
9	0.89	0.93	0.91	397	
10	0.94	0.96	0.95	399	
11	0.95	0.90	0.92	396	
12	0.75	0.79	0.77	393	
13	0.89	0.84	0.86	396	
14	0.90	0.91	0.90	394	
15	0.80	0.93	0.86	398	
16	0.73	0.90	0.81	364	
17	0.97	0.88	0.92	376	
18	0.79	0.59	0.68	310	
19	0.78	0.49	0.60	251	
accuracy			0.83	7532	
macro avg	0.83	0.82	0.82	7532	
weighted avg	0.83	0.83	0.83	7532	

*****MultinomialNB*****				
	precision	recall	f1-score	support
0	0.81	0.63	0.71	319
1	0.79	0.69	0.74	389
2	0.80	0.69	0.74	394
3	0.66	0.80	0.73	392
4	0.86	0.81	0.83	385
5	0.88	0.76	0.82	395
6	0.92	0.75	0.83	390
7	0.87	0.93	0.90	396
8	0.94	0.95	0.95	398
9	0.92	0.93	0.92	397
10	0.89	0.97	0.93	399
11	0.67	0.97	0.79	396
12	0.84	0.64	0.72	393
13	0.92	0.76	0.83	396
14	0.85	0.92	0.89	394
15	0.53	0.97	0.69	398
16	0.65	0.94	0.77	364
17	0.93	0.92	0.93	376
18	0.95	0.45	0.62	310
19	0.96	0.17	0.29	251
accuracy			0.80	7532
macro avg	0.83	0.78	0.78	7532
weighted avg	0.83	0.80	0.79	7532



\*\*\*\*\*KNeighborsClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.58	0.76	0.66	319
1	0.48	0.66	0.56	389
2	0.54	0.61	0.58	394
3	0.52	0.61	0.56	392
4	0.57	0.56	0.57	385
5	0.68	0.57	0.62	395
6	0.51	0.46	0.48	390
7	0.74	0.71	0.73	396
8	0.82	0.85	0.83	398
9	0.76	0.75	0.76	397
10	0.82	0.86	0.84	399
11	0.79	0.84	0.81	396
12	0.70	0.54	0.61	393
13	0.80	0.56	0.66	396
14	0.78	0.80	0.79	394
15	0.80	0.78	0.79	398
16	0.72	0.73	0.72	364
17	0.72	0.76	0.74	376
18	0.76	0.60	0.67	310
19	0.63	0.55	0.59	251
accuracy			0.68	7532
macro avg	0.69	0.68	0.68	7532
weighted avg	0.69	0.68	0.68	7532

*****RandomForestClassifier*****				
	precision	recall	f1-score	support
0	0.73	0.64	0.68	319
1	0.60	0.72	0.66	389
2	0.66	0.77	0.71	394
3	0.62	0.64	0.63	392
4	0.75	0.77	0.76	385
5	0.80	0.70	0.75	395
6	0.75	0.92	0.82	390
7	0.81	0.82	0.82	396
8	0.91	0.90	0.91	398
9	0.86	0.89	0.88	397
10	0.89	0.94	0.91	399
11	0.89	0.93	0.91	396
12	0.73	0.56	0.63	393
13	0.83	0.70	0.76	396
14	0.83	0.89	0.85	394
15	0.72	0.93	0.81	398
16	0.66	0.88	0.76	364
17	0.93	0.82	0.87	376
18	0.89	0.50	0.64	310
19	0.80	0.36	0.50	251
accuracy			0.78	7532
macro avg	0.78	0.76	0.76	7532
weighted avg	0.78	0.78	0.77	7532

STOPWORDS y BINARIO

```

*****Logistic Regression*****
      precision    recall  f1-score   support

     0         0.76      0.73      0.74         319
     1         0.66      0.76      0.71         389
     2         0.74      0.71      0.72         394
     3         0.66      0.70      0.68         392
     4         0.73      0.81      0.77         385
     5         0.81      0.71      0.76         395
     6         0.83      0.91      0.86         390
     7         0.88      0.84      0.86         396
     8         0.93      0.92      0.93         398
     9         0.87      0.90      0.88         397
    10         0.93      0.94      0.93         399
    11         0.93      0.86      0.89         396
    12         0.69      0.69      0.69         393
    13         0.82      0.74      0.78         396
    14         0.87      0.88      0.87         394
    15         0.83      0.94      0.88         398
    16         0.70      0.86      0.77         364
    17         0.96      0.79      0.87         376
    18         0.71      0.56      0.63         310
    19         0.67      0.64      0.65         251

 accuracy          0.80          7532
  macro avg         0.80          0.79          0.79          7532
 weighted avg         0.80          0.80          0.80          7532

```

\*\*\*\*\*MultinomialNB\*\*\*\*\*

	precision	recall	f1-score	support
0	0.81	0.71	0.76	319
1	0.71	0.67	0.69	389
2	0.90	0.14	0.24	394
3	0.55	0.80	0.65	392
4	0.90	0.74	0.81	385
5	0.72	0.84	0.78	395
6	0.92	0.71	0.80	390
7	0.85	0.90	0.88	396
8	0.96	0.94	0.95	398
9	0.96	0.90	0.93	397
10	0.93	0.97	0.95	399
11	0.62	0.96	0.75	396
12	0.80	0.66	0.73	393
13	0.86	0.85	0.86	396
14	0.80	0.93	0.86	394
15	0.67	0.97	0.80	398
16	0.66	0.95	0.78	364
17	0.89	0.95	0.92	376
18	0.83	0.56	0.67	310
19	0.96	0.29	0.45	251
accuracy			0.78	7532
macro avg	0.81	0.77	0.76	7532
weighted avg	0.81	0.78	0.77	7532

\*\*\*\*\*KNeighborsClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.43	0.41	0.42	319
1	0.08	0.66	0.14	389
2	0.45	0.19	0.27	394
3	0.41	0.16	0.23	392
4	0.10	0.39	0.16	385
5	0.77	0.18	0.30	395
6	0.53	0.28	0.37	390
7	0.33	0.17	0.23	396
8	0.90	0.27	0.41	398
9	0.43	0.19	0.27	397
10	0.43	0.18	0.26	399
11	0.95	0.20	0.33	396
12	0.64	0.08	0.14	393
13	0.70	0.10	0.18	396
14	0.89	0.18	0.30	394
15	0.48	0.31	0.38	398
16	0.91	0.20	0.33	364
17	0.42	0.44	0.43	376
18	0.37	0.22	0.27	310
19	0.61	0.16	0.26	251
accuracy			0.25	7532
macro avg	0.54	0.25	0.28	7532
weighted avg	0.54	0.25	0.28	7532

*****RandomForestClassifier*****				
	precision	recall	f1-score	support
0	0.71	0.65	0.68	319
1	0.57	0.69	0.63	389
2	0.66	0.80	0.72	394
3	0.67	0.69	0.68	392
4	0.75	0.79	0.77	385
5	0.78	0.70	0.74	395
6	0.76	0.92	0.83	390
7	0.85	0.81	0.83	396
8	0.91	0.90	0.90	398
9	0.83	0.91	0.87	397
10	0.89	0.93	0.91	399
11	0.87	0.92	0.90	396
12	0.76	0.53	0.62	393
13	0.88	0.69	0.77	396
14	0.83	0.89	0.86	394
15	0.71	0.94	0.81	398
16	0.66	0.87	0.75	364
17	0.94	0.79	0.86	376
18	0.87	0.50	0.63	310
19	0.79	0.38	0.51	251
accuracy			0.78	7532
macro avg	0.79	0.76	0.76	7532
weighted avg	0.79	0.78	0.77	7532

FRECUENCIA Y STOPWORS

*****Logistic Regression*****				
	precision	recall	f1-score	support
0	0.72	0.71	0.72	319
1	0.69	0.76	0.72	389
2	0.74	0.68	0.70	394
3	0.64	0.67	0.65	392
4	0.75	0.79	0.77	385
5	0.82	0.70	0.76	395
6	0.79	0.88	0.83	390
7	0.84	0.86	0.85	396
8	0.93	0.92	0.92	398
9	0.84	0.89	0.86	397
10	0.92	0.94	0.93	399
11	0.92	0.87	0.90	396
12	0.67	0.71	0.69	393
13	0.82	0.75	0.78	396
14	0.94	0.89	0.91	394
15	0.83	0.92	0.87	398
16	0.72	0.88	0.79	364
17	0.94	0.79	0.86	376
18	0.72	0.54	0.62	310
19	0.62	0.61	0.62	251
accuracy			0.79	7532
macro avg	0.79	0.79	0.79	7532
weighted avg	0.80	0.79	0.79	7532

\*\*\*\*\*MultinomialNB\*\*\*\*\*

	precision	recall	f1-score	support
0	0.79	0.80	0.79	319
1	0.65	0.77	0.70	389
2	0.60	0.02	0.03	394
3	0.55	0.80	0.65	392
4	0.84	0.78	0.81	385
5	0.69	0.84	0.76	395
6	0.93	0.71	0.81	390
7	0.87	0.92	0.90	396
8	0.96	0.94	0.95	398
9	0.95	0.91	0.93	397
10	0.93	0.97	0.95	399
11	0.74	0.96	0.84	396
12	0.80	0.70	0.75	393
13	0.89	0.84	0.86	396
14	0.84	0.91	0.88	394
15	0.75	0.96	0.84	398
16	0.70	0.92	0.80	364
17	0.92	0.93	0.92	376
18	0.64	0.63	0.64	310
19	0.90	0.38	0.54	251
accuracy			0.79	7532
macro avg	0.80	0.78	0.77	7532
weighted avg	0.80	0.79	0.77	7532



\*\*\*\*\*KNeighborsClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.34	0.51	0.40	319
1	0.12	0.58	0.20	389
2	0.26	0.36	0.30	394
3	0.43	0.30	0.35	392
4	0.13	0.39	0.20	385
5	0.61	0.17	0.27	395
6	0.59	0.34	0.43	390
7	0.38	0.34	0.36	396
8	0.73	0.43	0.54	398
9	0.54	0.34	0.42	397
10	0.76	0.36	0.49	399
11	0.81	0.46	0.59	396
12	0.61	0.15	0.23	393
13	0.60	0.20	0.31	396
14	0.81	0.40	0.54	394
15	0.45	0.51	0.47	398
16	0.75	0.34	0.47	364
17	0.65	0.48	0.55	376
18	0.70	0.27	0.39	310
19	0.45	0.23	0.30	251
accuracy			0.36	7532
macro avg	0.54	0.36	0.39	7532
weighted avg	0.54	0.36	0.39	7532

*****RandomForestClassifier*****				
	precision	recall	f1-score	support
0	0.74	0.65	0.69	319
1	0.58	0.71	0.64	389
2	0.66	0.77	0.71	394
3	0.67	0.69	0.68	392
4	0.73	0.75	0.74	385
5	0.76	0.66	0.71	395
6	0.76	0.93	0.84	390
7	0.82	0.81	0.81	396
8	0.93	0.90	0.92	398
9	0.82	0.91	0.86	397
10	0.90	0.92	0.91	399
11	0.89	0.93	0.91	396
12	0.71	0.55	0.62	393
13	0.87	0.69	0.77	396
14	0.86	0.88	0.87	394
15	0.72	0.95	0.82	398
16	0.67	0.87	0.76	364
17	0.94	0.82	0.87	376
18	0.88	0.51	0.65	310
19	0.78	0.35	0.49	251
accuracy			0.77	7532
macro avg			0.78	7532
weighted avg			0.77	7532

LEMATIZADO FRECUENCIA

```

*****Logistic Regression*****
      precision    recall  f1-score   support

     0       0.73      0.74      0.74       319
     1       0.70      0.75      0.72       389
     2       0.72      0.63      0.67       394
     3       0.63      0.69      0.66       392
     4       0.78      0.80      0.79       385
     5       0.83      0.73      0.78       395
     6       0.80      0.86      0.83       390
     7       0.85      0.84      0.84       396
     8       0.91      0.90      0.90       398
     9       0.85      0.90      0.87       397
    10       0.95      0.91      0.93       399
    11       0.90      0.86      0.88       396
    12       0.65      0.71      0.68       393
    13       0.81      0.77      0.79       396
    14       0.92      0.87      0.89       394
    15       0.82      0.92      0.87       398
    16       0.69      0.82      0.75       364
    17       0.92      0.81      0.86       376
    18       0.65      0.55      0.60       310
    19       0.60      0.57      0.59       251

 accuracy          0.79       7532
  macro avg       0.79      0.78      0.78       7532
 weighted avg     0.79      0.79      0.79       7532

```

*****MultinomialNB*****				
	precision	recall	f1-score	support
0	0.79	0.78	0.78	319
1	0.67	0.74	0.70	389
2	0.20	0.00	0.01	394
3	0.56	0.76	0.64	392
4	0.83	0.74	0.78	385
5	0.65	0.84	0.73	395
6	0.94	0.62	0.75	390
7	0.85	0.91	0.88	396
8	0.97	0.92	0.95	398
9	0.97	0.88	0.92	397
10	0.93	0.96	0.95	399
11	0.65	0.95	0.77	396
12	0.78	0.65	0.71	393
13	0.88	0.82	0.85	396
14	0.82	0.89	0.85	394
15	0.71	0.96	0.82	398
16	0.69	0.91	0.78	364
17	0.85	0.95	0.89	376
18	0.60	0.63	0.62	310
19	0.90	0.34	0.50	251
accuracy			0.77	7532
macro avg	0.76	0.76	0.74	7532
weighted avg	0.76	0.77	0.75	7532

\*\*\*\*\*KNeighborsClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.23	0.49	0.31	319
1	0.16	0.46	0.24	389
2	0.34	0.30	0.32	394
3	0.43	0.29	0.35	392
4	0.24	0.30	0.27	385
5	0.42	0.20	0.27	395
6	0.61	0.39	0.47	390
7	0.27	0.34	0.30	396
8	0.66	0.47	0.55	398
9	0.38	0.38	0.38	397
10	0.48	0.52	0.50	399
11	0.54	0.54	0.54	396
12	0.49	0.12	0.20	393
13	0.35	0.20	0.25	396
14	0.79	0.27	0.41	394
15	0.29	0.58	0.39	398
16	0.52	0.37	0.43	364
17	0.42	0.50	0.46	376
18	0.58	0.28	0.38	310
19	0.40	0.20	0.27	251
accuracy			0.36	7532
macro avg	0.43	0.36	0.36	7532
weighted avg	0.43	0.36	0.37	7532

*****RandomForestClassifier*****				
	precision	recall	f1-score	support
0	0.73	0.62	0.67	319
1	0.60	0.76	0.67	389
2	0.66	0.77	0.71	394
3	0.66	0.66	0.66	392
4	0.72	0.79	0.75	385
5	0.79	0.70	0.74	395
6	0.73	0.92	0.82	390
7	0.84	0.82	0.83	396
8	0.90	0.91	0.91	398
9	0.83	0.91	0.87	397
10	0.91	0.93	0.92	399
11	0.86	0.90	0.88	396
12	0.72	0.54	0.61	393
13	0.86	0.66	0.75	396
14	0.86	0.88	0.87	394
15	0.69	0.94	0.79	398
16	0.65	0.86	0.74	364
17	0.94	0.78	0.85	376
18	0.90	0.48	0.63	310
19	0.83	0.33	0.48	251
accuracy			0.77	7532
macro avg			0.78	7532
weighted avg			0.77	7532

LEMATIZAR TFIDF

*****Logistic Regression*****					
	precision	recall	f1-score	support	
0	0.77	0.74	0.76	319	
1	0.70	0.80	0.75	389	
2	0.78	0.74	0.76	394	
3	0.72	0.72	0.72	392	
4	0.82	0.84	0.83	385	
5	0.84	0.77	0.80	395	
6	0.73	0.89	0.80	390	
7	0.92	0.87	0.89	396	
8	0.96	0.93	0.94	398	
9	0.88	0.92	0.90	397	
10	0.95	0.94	0.95	399	
11	0.93	0.90	0.91	396	
12	0.75	0.78	0.76	393	
13	0.89	0.85	0.87	396	
14	0.89	0.91	0.90	394	
15	0.80	0.92	0.85	398	
16	0.74	0.88	0.80	364	
17	0.96	0.86	0.91	376	
18	0.77	0.61	0.68	310	
19	0.79	0.48	0.60	251	
accuracy			0.83	7532	
macro avg	0.83	0.82	0.82	7532	
weighted avg	0.83	0.83	0.83	7532	

\*\*\*\*\*MultinomialNB\*\*\*\*\*

	precision	recall	f1-score	support
0	0.81	0.51	0.62	319
1	0.81	0.66	0.73	389
2	0.83	0.64	0.73	394
3	0.68	0.78	0.72	392
4	0.85	0.79	0.82	385
5	0.88	0.75	0.81	395
6	0.94	0.65	0.77	390
7	0.84	0.93	0.88	396
8	0.93	0.94	0.93	398
9	0.92	0.90	0.91	397
10	0.91	0.97	0.94	399
11	0.57	0.97	0.72	396
12	0.84	0.61	0.71	393
13	0.93	0.75	0.83	396
14	0.85	0.89	0.87	394
15	0.45	0.98	0.62	398
16	0.65	0.94	0.77	364
17	0.92	0.92	0.92	376
18	0.96	0.42	0.58	310
19	1.00	0.14	0.24	251
accuracy			0.77	7532
macro avg	0.83	0.76	0.76	7532
weighted avg	0.82	0.77	0.77	7532



```

*****KNeighborsClassifier*****
      precision    recall  f1-score   support

     0         0.47      0.75      0.58        319
     1         0.51      0.61      0.55        389
     2         0.57      0.56      0.56        394
     3         0.55      0.60      0.57        392
     4         0.60      0.55      0.58        385
     5         0.69      0.63      0.66        395
     6         0.60      0.43      0.50        390
     7         0.77      0.74      0.75        396
     8         0.83      0.83      0.83        398
     9         0.76      0.76      0.76        397
    10         0.86      0.84      0.85        399
    11         0.72      0.83      0.77        396
    12         0.72      0.50      0.59        393
    13         0.81      0.49      0.62        396
    14         0.80      0.77      0.78        394
    15         0.75      0.78      0.76        398
    16         0.63      0.76      0.69        364
    17         0.59      0.75      0.66        376
    18         0.58      0.56      0.57        310
    19         0.55      0.45      0.50        251

 accuracy                   0.66       7532
 macro avg                 0.67       0.66      0.66       7532
 weighted avg              0.67       0.66      0.66       7532

```

*****RandomForestClassifier*****					
	precision	recall	f1-score	support	
0	0.73	0.65	0.69	319	
1	0.56	0.68	0.61	389	
2	0.66	0.75	0.70	394	
3	0.66	0.64	0.65	392	
4	0.74	0.77	0.75	385	
5	0.79	0.69	0.74	395	
6	0.73	0.91	0.81	390	
7	0.82	0.82	0.82	396	
8	0.91	0.91	0.91	398	
9	0.79	0.89	0.84	397	
10	0.91	0.92	0.91	399	
11	0.89	0.92	0.90	396	
12	0.70	0.52	0.60	393	
13	0.85	0.67	0.75	396	
14	0.81	0.90	0.85	394	
15	0.68	0.92	0.78	398	
16	0.66	0.87	0.75	364	
17	0.94	0.80	0.86	376	
18	0.86	0.48	0.61	310	
19	0.77	0.30	0.43	251	
accuracy			0.76	7532	
macro avg	0.77	0.75	0.75	7532	
weighted avg	0.77	0.76	0.76	7532	

LEM STOP TFD FIDF

*****Logistic Regression*****				
	precision	recall	f1-score	support
0	0.79	0.74	0.76	319
1	0.69	0.80	0.74	389
2	0.78	0.75	0.76	394
3	0.70	0.72	0.71	392
4	0.81	0.83	0.82	385
5	0.85	0.77	0.81	395
6	0.74	0.88	0.80	390
7	0.92	0.89	0.90	396
8	0.96	0.94	0.95	398
9	0.90	0.92	0.91	397
10	0.95	0.95	0.95	399
11	0.94	0.91	0.92	396
12	0.74	0.79	0.76	393
13	0.89	0.85	0.87	396
14	0.90	0.90	0.90	394
15	0.81	0.92	0.86	398
16	0.74	0.89	0.81	364
17	0.96	0.87	0.91	376
18	0.78	0.60	0.68	310
19	0.79	0.49	0.61	251
accuracy			0.83	7532
macro avg	0.83	0.82	0.82	7532
weighted avg	0.83	0.83	0.83	7532

```

*****MultinomialNB*****
      precision    recall  f1-score   support

     0         0.81         0.61         0.70         319
     1         0.79         0.70         0.74         389
     2         0.82         0.68         0.74         394
     3         0.67         0.79         0.73         392
     4         0.86         0.82         0.84         385
     5         0.88         0.77         0.82         395
     6         0.92         0.71         0.80         390
     7         0.85         0.93         0.89         396
     8         0.93         0.95         0.94         398
     9         0.92         0.92         0.92         397
    10         0.90         0.98         0.94         399
    11         0.65         0.97         0.78         396
    12         0.85         0.64         0.73         393
    13         0.92         0.76         0.83         396
    14         0.84         0.94         0.89         394
    15         0.54         0.97         0.70         398
    16         0.65         0.94         0.77         364
    17         0.93         0.93         0.93         376
    18         0.97         0.45         0.62         310
    19         0.94         0.18         0.30         251

 accuracy                   0.80         7532
 macro avg         0.83         0.78         0.78         7532
 weighted avg      0.83         0.80         0.79         7532

```

```

*****KNeighborsClassifier*****
      precision    recall  f1-score   support

     0         0.57      0.76      0.65       319
     1         0.49      0.63      0.55       389
     2         0.55      0.59      0.57       394
     3         0.51      0.59      0.55       392
     4         0.60      0.57      0.58       385
     5         0.68      0.59      0.63       395
     6         0.55      0.44      0.49       390
     7         0.75      0.73      0.74       396
     8         0.81      0.85      0.83       398
     9         0.75      0.74      0.75       397
    10         0.83      0.86      0.85       399
    11         0.77      0.84      0.80       396
    12         0.71      0.52      0.60       393
    13         0.80      0.56      0.66       396
    14         0.78      0.81      0.80       394
    15         0.79      0.78      0.78       398
    16         0.70      0.77      0.73       364
    17         0.63      0.76      0.69       376
    18         0.75      0.59      0.66       310
    19         0.64      0.50      0.56       251

 accuracy          0.68       7532
 macro avg         0.68      0.67      0.67       7532
 weighted avg      0.68      0.68      0.68       7532

```

```

*****RandomForestClassifier*****
      precision    recall  f1-score   support

     0       0.74       0.66       0.70       319
     1       0.60       0.72       0.65       389
     2       0.68       0.78       0.72       394
     3       0.67       0.67       0.67       392
     4       0.73       0.75       0.74       385
     5       0.76       0.72       0.74       395
     6       0.76       0.91       0.83       390
     7       0.85       0.82       0.83       396
     8       0.88       0.91       0.90       398
     9       0.84       0.91       0.87       397
    10       0.90       0.93       0.92       399
    11       0.88       0.91       0.90       396
    12       0.73       0.54       0.62       393
    13       0.85       0.72       0.78       396
    14       0.84       0.87       0.86       394
    15       0.71       0.94       0.81       398
    16       0.66       0.88       0.75       364
    17       0.96       0.81       0.88       376
    18       0.84       0.51       0.63       310
    19       0.77       0.31       0.44       251

 accuracy          0.78          0.78          0.78       7532
  macro avg       0.78       0.76       0.76       7532
 weighted avg     0.78       0.78       0.77       7532

```

LEM STOPWORDS FRECUENCIA

*****Logistic Regression*****				
	precision	recall	f1-score	support
0	0.73	0.72	0.73	319
1	0.70	0.77	0.73	389
2	0.75	0.67	0.71	394
3	0.66	0.70	0.68	392
4	0.76	0.81	0.78	385
5	0.80	0.74	0.77	395
6	0.79	0.90	0.84	390
7	0.86	0.84	0.85	396
8	0.92	0.92	0.92	398
9	0.84	0.89	0.86	397
10	0.92	0.92	0.92	399
11	0.92	0.89	0.90	396
12	0.70	0.72	0.71	393
13	0.88	0.77	0.82	396
14	0.91	0.89	0.90	394
15	0.83	0.93	0.88	398
16	0.69	0.84	0.76	364
17	0.94	0.78	0.85	376
18	0.73	0.55	0.63	310
19	0.61	0.61	0.61	251
accuracy			0.80	7532
macro avg	0.80	0.79	0.79	7532
weighted avg	0.80	0.80	0.80	7532

*****MultinomialNB*****				
	precision	recall	f1-score	support
0	0.80	0.80	0.80	319
1	0.66	0.77	0.71	389
2	0.43	0.01	0.01	394
3	0.56	0.80	0.66	392
4	0.82	0.77	0.80	385
5	0.68	0.84	0.75	395
6	0.93	0.67	0.78	390
7	0.85	0.92	0.88	396
8	0.97	0.94	0.95	398
9	0.97	0.90	0.93	397
10	0.93	0.97	0.95	399
11	0.72	0.96	0.82	396
12	0.79	0.70	0.74	393
13	0.88	0.82	0.85	396
14	0.83	0.90	0.87	394
15	0.75	0.96	0.84	398
16	0.71	0.92	0.80	364
17	0.89	0.94	0.91	376
18	0.65	0.64	0.64	310
19	0.90	0.38	0.53	251
accuracy			0.79	7532
macro avg	0.79	0.78	0.76	7532
weighted avg	0.79	0.79	0.77	7532



\*\*\*\*\*KNeighborsClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.19	0.59	0.29	319
1	0.16	0.47	0.24	389
2	0.33	0.31	0.32	394
3	0.42	0.30	0.35	392
4	0.21	0.31	0.25	385
5	0.50	0.19	0.28	395
6	0.59	0.29	0.39	390
7	0.35	0.38	0.37	396
8	0.73	0.47	0.58	398
9	0.44	0.43	0.43	397
10	0.81	0.39	0.52	399
11	0.74	0.51	0.60	396
12	0.60	0.13	0.22	393
13	0.63	0.24	0.35	396
14	0.85	0.35	0.49	394
15	0.32	0.66	0.43	398
16	0.75	0.32	0.45	364
17	0.27	0.49	0.35	376
18	0.71	0.26	0.38	310
19	0.43	0.17	0.25	251
accuracy			0.37	7532
macro avg	0.50	0.36	0.38	7532
weighted avg	0.50	0.37	0.38	7532

*****RandomForestClassifier*****				
	precision	recall	f1-score	support
0	0.71	0.69	0.70	319
1	0.63	0.74	0.68	389
2	0.65	0.76	0.70	394
3	0.64	0.68	0.66	392
4	0.76	0.76	0.76	385
5	0.78	0.72	0.75	395
6	0.74	0.91	0.81	390
7	0.81	0.82	0.81	396
8	0.90	0.92	0.91	398
9	0.84	0.92	0.88	397
10	0.89	0.94	0.92	399
11	0.88	0.93	0.91	396
12	0.72	0.50	0.59	393
13	0.90	0.69	0.78	396
14	0.85	0.90	0.87	394
15	0.74	0.94	0.82	398
16	0.66	0.87	0.75	364
17	0.94	0.81	0.87	376
18	0.87	0.48	0.62	310
19	0.77	0.36	0.49	251
accuracy			0.78	7532
macro avg			0.78	7532
weighted avg			0.78	7532

LEM SW CT TFIDF

*****Logistic Regression*****				
	precision	recall	f1-score	support
0	0.71	0.63	0.67	319
1	0.66	0.71	0.69	389
2	0.68	0.63	0.65	394
3	0.69	0.66	0.68	392
4	0.76	0.75	0.76	385
5	0.79	0.72	0.75	395
6	0.68	0.82	0.74	390
7	0.82	0.78	0.80	396
8	0.83	0.92	0.87	398
9	0.78	0.88	0.83	397
10	0.92	0.90	0.91	399
11	0.88	0.78	0.83	396
12	0.60	0.64	0.62	393
13	0.76	0.78	0.77	396
14	0.78	0.80	0.79	394
15	0.70	0.85	0.77	398
16	0.67	0.76	0.71	364
17	0.92	0.81	0.86	376
18	0.66	0.58	0.62	310
19	0.65	0.34	0.45	251
accuracy			0.75	7532
macro avg	0.75	0.74	0.74	7532
weighted avg	0.75	0.75	0.75	7532

```

*****MultinomialNB*****
      precision    recall  f1-score   support

0         0.81      0.30      0.43       319
1         0.76      0.68      0.72       389
2         0.75      0.56      0.64       394
3         0.62      0.76      0.68       392
4         0.81      0.74      0.77       385
5         0.83      0.75      0.79       395
6         0.90      0.70      0.79       390
7         0.84      0.83      0.83       396
8         0.85      0.91      0.88       398
9         0.86      0.85      0.86       397
10        0.87      0.93      0.90       399
11        0.55      0.91      0.69       396
12        0.80      0.52      0.63       393
13        0.87      0.73      0.79       396
14        0.78      0.84      0.81       394
15        0.39      0.97      0.55       398
16        0.60      0.82      0.70       364
17        0.86      0.84      0.85       376
18        0.95      0.32      0.48       310
19        1.00      0.09      0.17       251

accuracy          0.72       7532
macro avg         0.79      0.70      0.70       7532
weighted avg      0.78      0.72      0.71       7532

```

\*\*\*\*\*KNeighborsClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.62	0.56	0.59	319
1	0.08	0.94	0.15	389
2	0.81	0.19	0.31	394
3	0.77	0.28	0.41	392
4	0.88	0.25	0.39	385
5	0.90	0.33	0.48	395
6	0.96	0.18	0.30	390
7	0.91	0.36	0.52	396
8	0.95	0.58	0.72	398
9	0.93	0.41	0.57	397
10	0.97	0.59	0.73	399
11	0.89	0.59	0.71	396
12	0.74	0.20	0.31	393
13	0.85	0.31	0.46	396
14	0.88	0.47	0.62	394
15	0.89	0.37	0.52	398
16	0.84	0.34	0.49	364
17	0.69	0.54	0.61	376
18	0.91	0.35	0.50	310
19	0.73	0.18	0.28	251
accuracy			0.40	7532
macro avg	0.81	0.40	0.48	7532
weighted avg	0.81	0.40	0.49	7532

*****RandomForestClassifier*****					
	precision	recall	f1-score	support	
0	0.58	0.49	0.53	319	
1	0.42	0.65	0.51	389	
2	0.56	0.61	0.59	394	
3	0.65	0.57	0.60	392	
4	0.69	0.66	0.67	385	
5	0.73	0.64	0.68	395	
6	0.70	0.81	0.75	390	
7	0.69	0.69	0.69	396	
8	0.82	0.83	0.83	398	
9	0.67	0.83	0.74	397	
10	0.79	0.85	0.82	399	
11	0.78	0.77	0.77	396	
12	0.57	0.44	0.50	393	
13	0.71	0.64	0.67	396	
14	0.74	0.76	0.75	394	
15	0.59	0.86	0.70	398	
16	0.59	0.68	0.63	364	
17	0.92	0.76	0.83	376	
18	0.78	0.40	0.53	310	
19	0.56	0.18	0.28	251	
accuracy			0.67	7532	
macro avg	0.68	0.66	0.65	7532	
weighted avg	0.68	0.67	0.66	7532	

LEM + STOP MLP

\*\*\*\*\*MLPClassifier\*\*\*\*\*

	precision	recall	f1-score	support
0	0.85	0.78	0.81	319
1	0.75	0.80	0.77	389
2	0.76	0.75	0.76	394
3	0.78	0.67	0.72	392
4	0.87	0.79	0.82	385
5	0.89	0.79	0.84	395
7	0.84	0.92	0.88	396
8	0.98	0.94	0.96	398
9	0.91	0.97	0.94	397
10	0.98	0.97	0.98	399
11	0.90	0.94	0.92	396
12	0.68	0.86	0.76	393
13	0.88	0.84	0.86	396
14	0.89	0.92	0.90	394
15	0.86	0.93	0.90	398
16	0.77	0.92	0.84	364
17	0.98	0.86	0.92	376
18	0.84	0.59	0.70	310
19	0.73	0.67	0.70	251
accuracy			0.85	7532
macro avg	0.85	0.84	0.84	7532
weighted avg	0.85	0.85	0.85	7532

```

*****MLPClassifier*****
      precision    recall  f1-score   support

     0       0.73      0.76      0.75      319
     1       0.71      0.81      0.76      389
     2       0.81      0.68      0.74      394
     3       0.70      0.73      0.71      392
     4       0.77      0.84      0.80      385
     5       0.84      0.76      0.80      395
     6       0.78      0.86      0.82      390
     7       0.90      0.91      0.90      396
     8       0.91      0.94      0.93      398
     9       0.92      0.90      0.91      397
    10       0.94      0.98      0.96      399
    11       0.89      0.90      0.90      396
    12       0.76      0.77      0.76      393
    13       0.89      0.81      0.85      396
    14       0.88      0.89      0.88      394
    15       0.88      0.90      0.89      398
    16       0.73      0.86      0.79      364
    17       0.97      0.81      0.88      376
    18       0.78      0.60      0.68      310
    19       0.62      0.62      0.62      251

 accuracy          0.82      7532
 macro avg         0.82      0.82      0.82      7532
 weighted avg      0.83      0.82      0.82      7532

```



*****MLPClassifier*****				
	precision	recall	f1-score	support
0	0.81	0.76	0.79	319
1	0.75	0.79	0.77	389
2	0.72	0.74	0.73	394
3	0.66	0.76	0.71	392
4	0.81	0.83	0.82	385
5	0.87	0.75	0.81	395
6	0.81	0.84	0.82	390
7	0.92	0.87	0.89	396
8	0.87	0.97	0.92	398
9	0.90	0.94	0.92	397
10	0.97	0.95	0.96	399
11	0.86	0.93	0.90	396
12	0.76	0.72	0.74	393
13	0.94	0.78	0.85	396
14	0.90	0.87	0.89	394
15	0.86	0.92	0.89	398
16	0.75	0.91	0.82	364
17	0.95	0.88	0.91	376
18	0.82	0.63	0.72	310
19	0.70	0.68	0.69	251
accuracy			0.83	7532
macro avg			0.83	7532
weighted avg			0.83	7532

*****MLPClassifier*****				
	precision	recall	f1-score	support
0	0.85	0.77	0.81	319
1	0.76	0.78	0.77	389
2	0.77	0.75	0.76	394
3	0.70	0.74	0.72	392
4	0.85	0.85	0.85	385
5	0.88	0.82	0.85	395
6	0.82	0.87	0.85	390
7	0.91	0.92	0.92	396
8	0.98	0.95	0.97	398
9	0.93	0.97	0.95	397
10	0.97	0.98	0.98	399
11	0.93	0.93	0.93	396
12	0.80	0.80	0.80	393
13	0.90	0.88	0.89	396
14	0.91	0.90	0.91	394
15	0.83	0.94	0.88	398
16	0.77	0.91	0.83	364
17	0.97	0.90	0.94	376
18	0.82	0.63	0.71	310
19	0.73	0.64	0.68	251
accuracy			0.86	7532
macro avg			0.85	7532
weighted avg			0.86	7532

AHORA CON SVC

1.0

*****SVC*****				
	precision	recall	f1-score	support
0	0.84	0.73	0.78	319
1	0.63	0.81	0.71	389
2	0.81	0.72	0.76	394
3	0.71	0.78	0.74	392
4	0.83	0.82	0.83	385
5	0.83	0.76	0.79	395
6	0.71	0.90	0.79	390
7	0.92	0.87	0.89	396
8	0.97	0.93	0.95	398
9	0.90	0.92	0.91	397
10	0.97	0.93	0.95	399
11	0.96	0.87	0.91	396
12	0.65	0.84	0.73	393
13	0.87	0.79	0.83	396
14	0.91	0.87	0.89	394
15	0.79	0.92	0.85	398
16	0.76	0.88	0.82	364
17	0.98	0.81	0.89	376
18	0.78	0.60	0.68	310
19	0.80	0.54	0.64	251
accuracy			0.82	7532
macro avg			0.83	7532
weighted avg			0.83	7532

*****SVC*****					
	precision	recall	f1-score	support	
0	0.84	0.73	0.78	319	
1	0.65	0.81	0.73	389	
2	0.80	0.72	0.76	394	
3	0.72	0.78	0.74	392	
4	0.82	0.83	0.83	385	
5	0.83	0.76	0.79	395	
6	0.73	0.90	0.80	390	
7	0.92	0.88	0.90	396	
8	0.97	0.93	0.95	398	
9	0.90	0.91	0.91	397	
10	0.97	0.94	0.96	399	
11	0.96	0.87	0.91	396	
12	0.68	0.84	0.75	393	
13	0.87	0.81	0.84	396	
14	0.91	0.88	0.89	394	
15	0.81	0.92	0.86	398	
16	0.77	0.89	0.82	364	
17	0.98	0.82	0.90	376	
18	0.78	0.61	0.68	310	
19	0.77	0.57	0.65	251	
accuracy			0.83	7532	
macro avg		0.83	0.82	0.82	7532
weighted avg		0.84	0.83	0.83	7532

*****SVC*****					
	precision	recall	f1-score	support	
0	0.84	0.73	0.78	319	
1	0.65	0.81	0.72	389	
2	0.80	0.72	0.76	394	
3	0.71	0.77	0.74	392	
4	0.82	0.83	0.83	385	
5	0.83	0.76	0.79	395	
6	0.72	0.90	0.80	390	
7	0.92	0.88	0.90	396	
8	0.97	0.93	0.95	398	
9	0.90	0.91	0.91	397	
10	0.97	0.94	0.96	399	
11	0.96	0.87	0.91	396	
12	0.68	0.84	0.75	393	
13	0.87	0.81	0.84	396	
14	0.91	0.88	0.89	394	
15	0.81	0.92	0.86	398	
16	0.77	0.89	0.82	364	
17	0.98	0.82	0.90	376	
18	0.78	0.61	0.68	310	
19	0.77	0.57	0.65	251	
accuracy			0.83	7532	
macro avg	0.83	0.82	0.82	7532	
weighted avg	0.84	0.83	0.83	7532	

100

MLP con SVD

	precision	recall	f1-score	support
0	0.80	0.77	0.79	319
1	0.74	0.77	0.75	389
2	0.74	0.72	0.73	394
3	0.65	0.70	0.67	392
4	0.78	0.83	0.81	385
5	0.82	0.76	0.79	395
6	0.78	0.86	0.82	390
7	0.87	0.88	0.87	396
8	0.97	0.92	0.94	398
9	0.89	0.92	0.90	397
10	0.96	0.96	0.96	399
11	0.93	0.89	0.91	396
12	0.69	0.76	0.73	393
13	0.89	0.81	0.85	396
14	0.91	0.89	0.90	394
15	0.83	0.89	0.86	398
16	0.74	0.88	0.80	364
17	0.96	0.82	0.89	376
18	0.78	0.62	0.69	310
19	0.63	0.59	0.61	251
accuracy			0.82	7532
macro avg	0.82	0.81	0.81	7532
weighted avg	0.82	0.82	0.82	7532