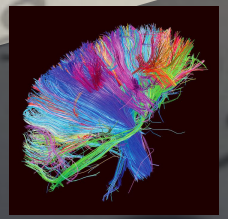
Análisis estilográfico de autores

Equipo:
Knights of Cydonia

Avalos Gonzalez Joel Sebastian
Castañon Maldonado Carlos Emilio
Reyes Ramos Luz María
Sánchez Castro Gustavo



Introducción e hipótesis

El análisis estilográfico es una rama de la lingüística forense que se ocupa de la identificación de la autoría de un texto basado en las características estilísticas del lenguaje utilizado. Tradicionalmente, el análisis estilográfico se ha realizado utilizando métodos manuales, pero con el advenimiento de la tecnología informática y el desarrollo de poderosas técnicas de aprendizaje automático, se ha hecho posible automatizar el proceso de análisis estilográfico.

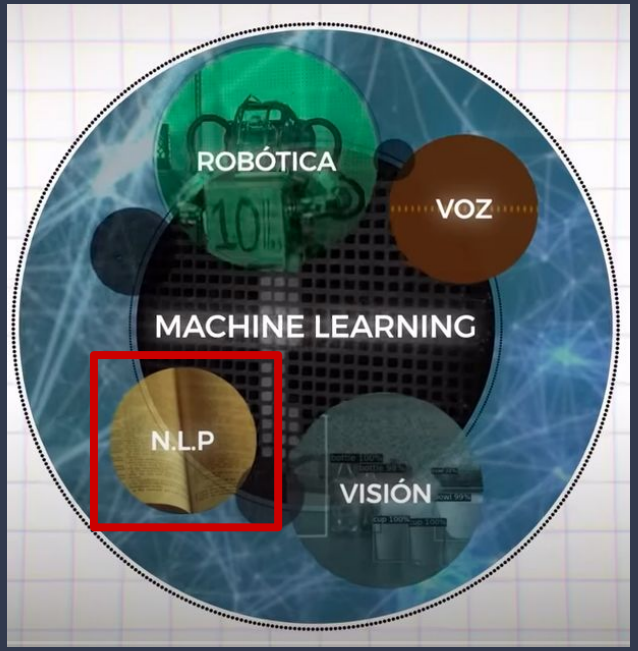
Hipótesis:

En este proyecto, nos proponemos determinar la autoría de un texto con un perceptrón multicapa a partir de textos de diferentes autores. La hipótesis que planteamos es que el análisis de características estilísticas del texto, como la longitud de las palabras, la frecuencia de las letras y la estructura de las oraciones, puede utilizarse para distinguir entre los estilos de diferentes autores.

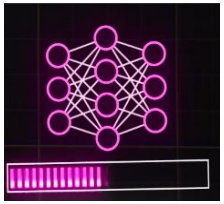
Objetivos:

El objetivo principal de este proyecto es desarrollar un modelo de aprendizaje automático que pueda ser utilizado para la clasificación automática de textos por autor. El modelo se basará en un conjunto de características estilísticas extraídas de los textos. Se espera que el modelo pueda clasificar correctamente los textos de diferentes autores con una alta precisión.

Motivación



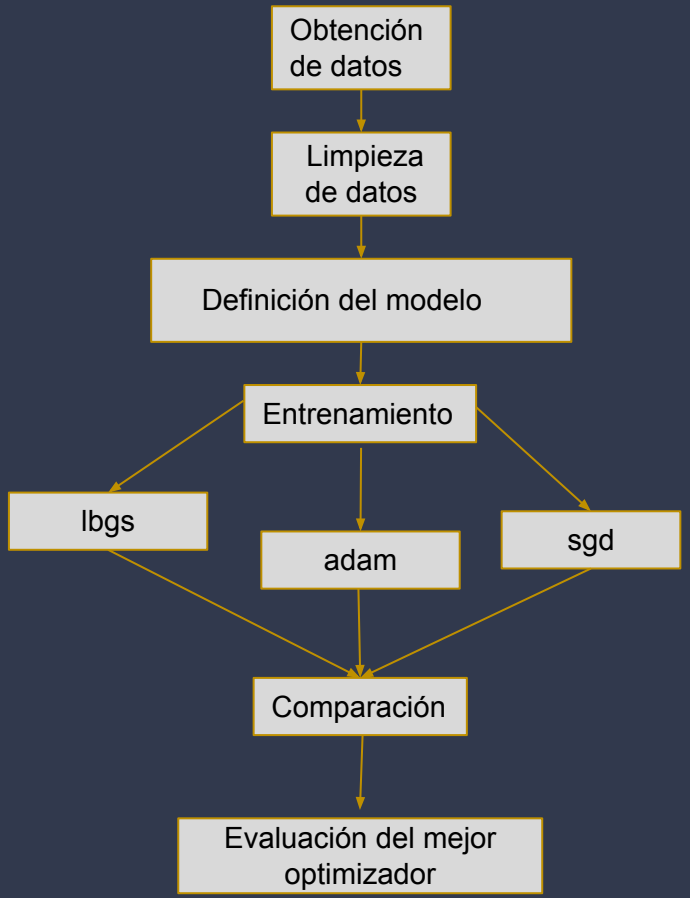
Actualmente el uso del procesamiento natural es de gran importancia en el mundo del ***deep learning.***



En áreas como la forense o la seguridad, el análisis estilográfico permite identificar a un autor desconocido basándose en el estilo de escritura. Esto puede ser útil para verificar la autenticidad de documentos o identificar posibles autores de textos anónimos, como en correos electrónicos anónimos o publicaciones en redes sociales.

El análisis estilográfico puede ayudar a detectar el plagio o la copia de textos, ya que permite identificar similitudes en el estilo de escritura entre diferentes documentos.

Resumen



Facultad de Ciencias UNAM

En este trabajo se tuvo como objetivo implementar un perceptrón multicapa para la asignación de autores de distintos textos.

Se busco una comparación de 3 diferentes funciones de optimización (lbfgs, adam y sgd) para los pesos del perceptrón, de esa forma quedarnos con el de mayor precisión al momento de clasificar los textos.

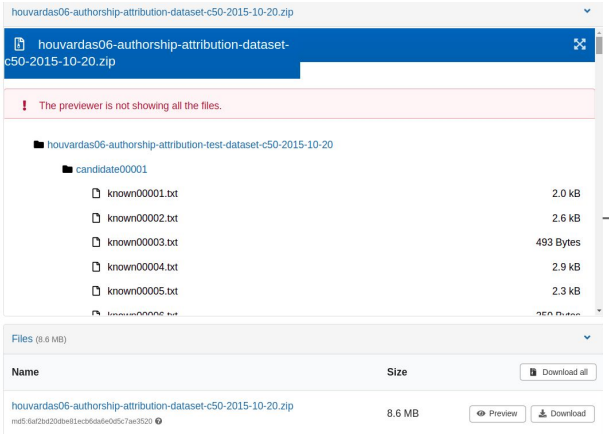
Adquisición de Datos

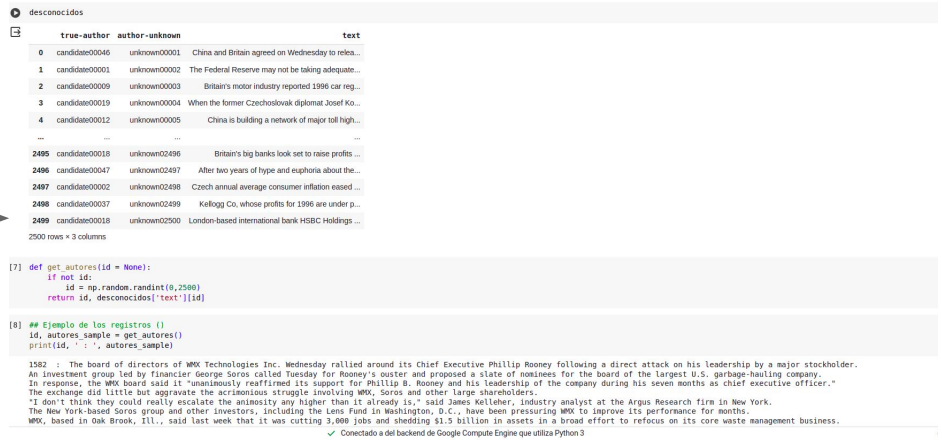
Los datos fueron recopilados del evento PAN-CLEF, Cross-Language Evaluation Forum (CLEF), es una iniciativa internacional que se enfoca en la evaluación y prueba de sistemas de información multilingües y multimodales.

El objetivo principal del evento PAN-CLEF es evaluar y promover el desarrollo de sistemas y herramientas para la recuperación de información, minería de textos, detección de plagio, análisis de autoría y tareas relacionadas en entornos multilingües y multinacionales. Se centra específicamente en el acceso personal a archivos web, lo que implica la búsqueda y recuperación de información en archivos web históricos y diversos.

Adquisición de Datos

El conjunto de datos contiene 2500 textos de 50 autores/candidatos diferentes (C50). La verdad fundamental se puede encontrar dentro de un archivo json.



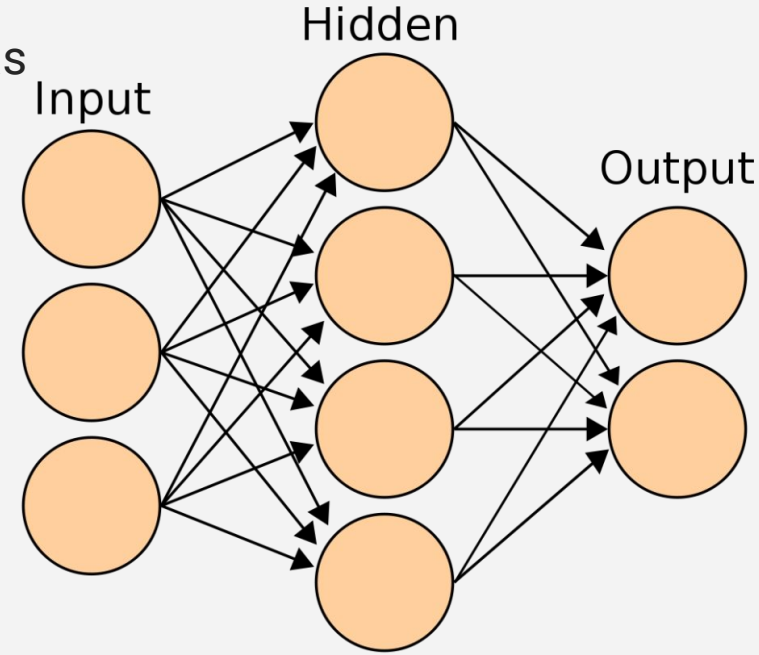


https://zenodo.org/records/3759068

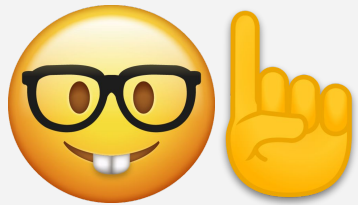
Perceptrón Multicapa

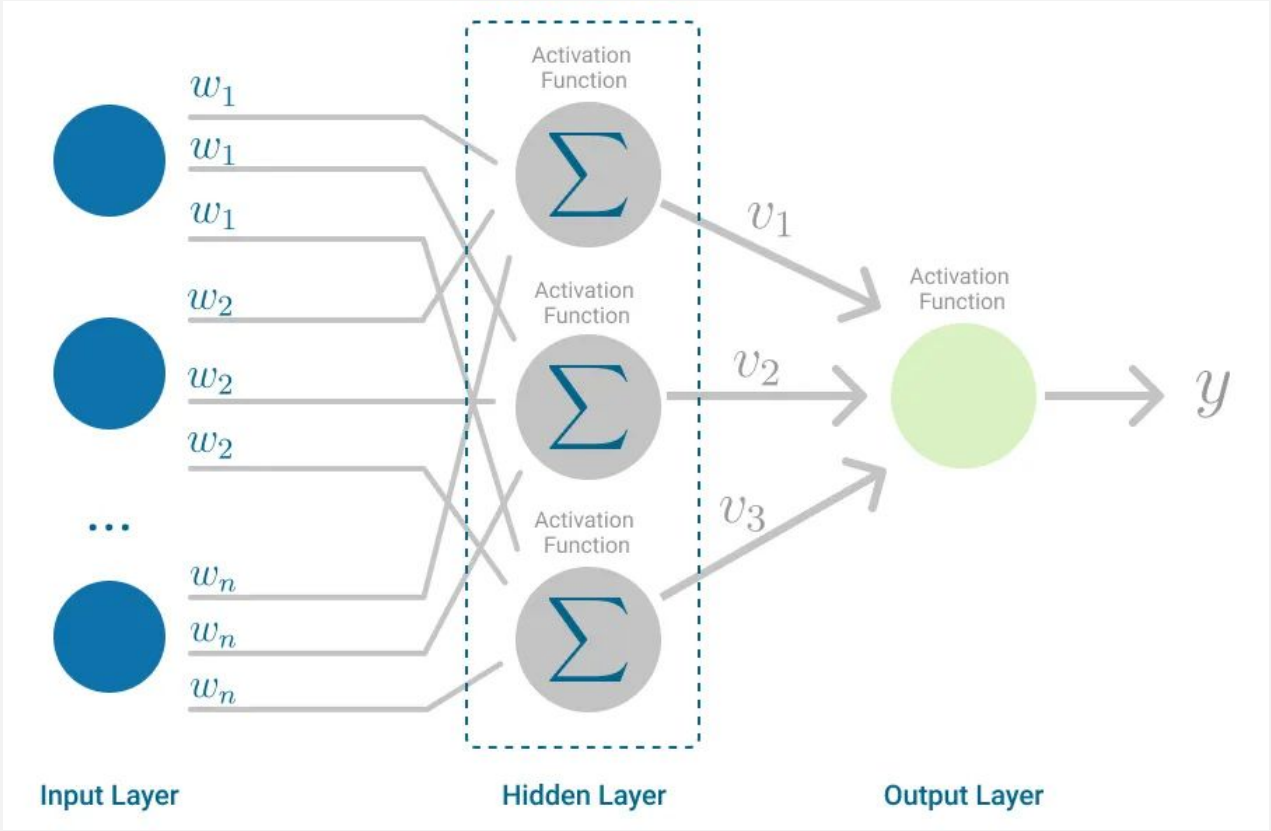
Un perceptrón multicapa posee capas de entrada y salida, y una o más capas ocultas con muchas neuronas apiladas.

Mientras que en el Perceptrón la neurona debe tener una función de activación que imponga un umbral, como ReLU o sigmoide, las neuronas en un Perceptrón multicapa pueden usar cualquier función de activación arbitraria.



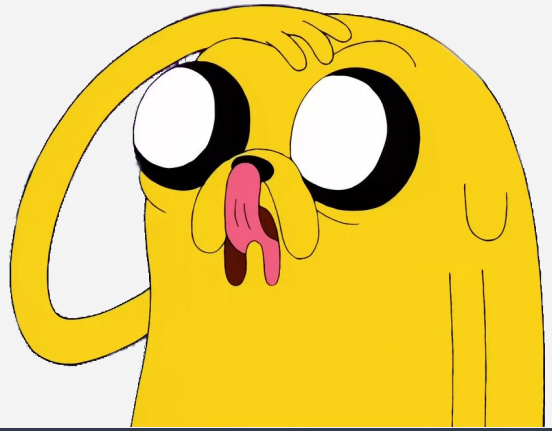
El Perceptrón multicapa se clasifica como un algoritmo de avance, ya que combina las entradas con los pesos iniciales mediante una suma ponderada y aplica una función de activación, similar al Perceptrón. La diferencia clave radica en que cada combinación lineal se propaga a la siguiente capa, creando una representación interna de los datos que fluye a través de capas ocultas hasta la salida. Sin embargo, es crucial destacar que si el algoritmo se limitará a calcular sumas ponderadas y propagar resultados solo hasta la capa de salida, no sería capaz de aprender los pesos que minimizan la función de costos. El aprendizaje real implica que este proceso se repite a lo largo de varias iteraciones para ajustar y mejorar continuamente los pesos.



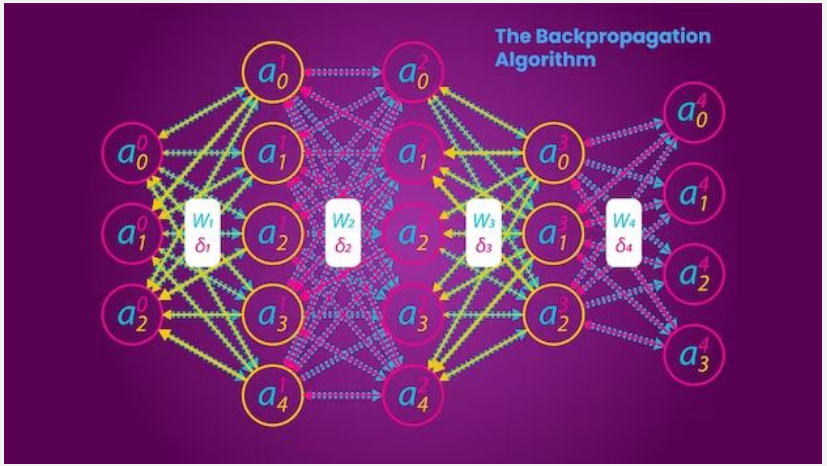


Modelo: Perceptrón Multicapa

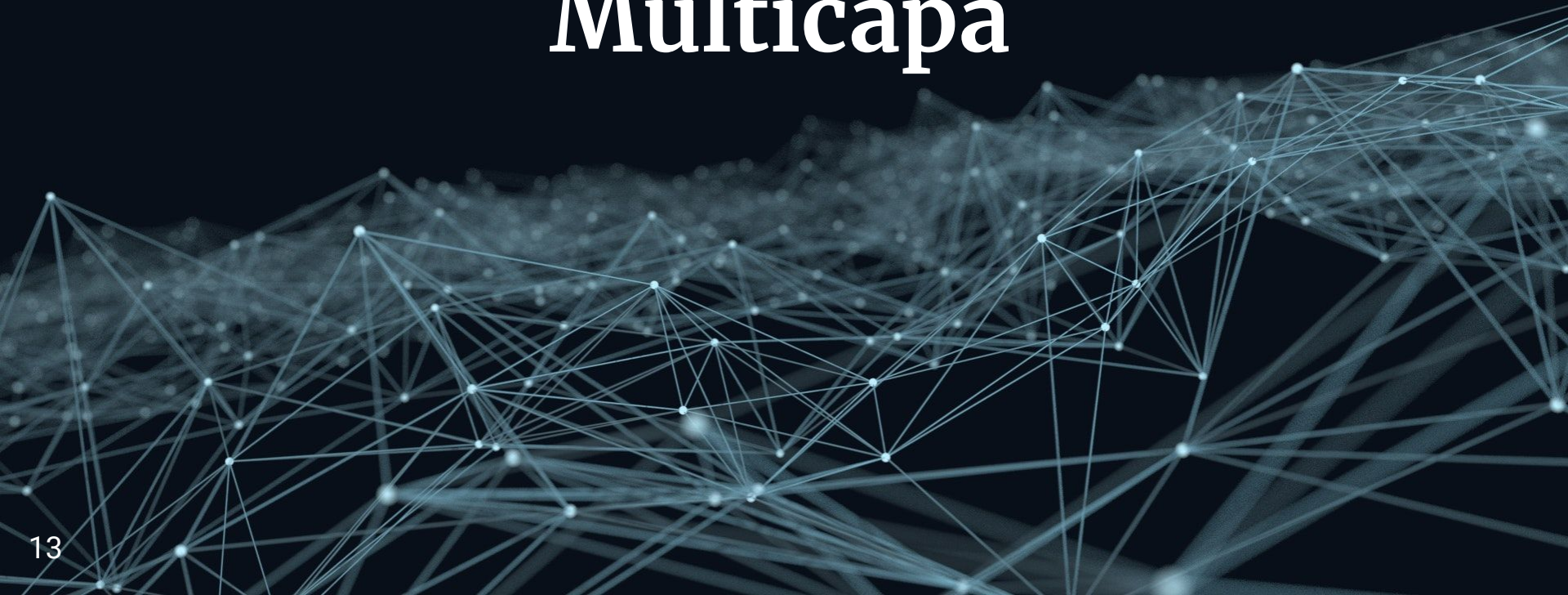
Si el algoritmo sólo calculará las sumas ponderadas en cada neurona, propagara los resultados a la capa de salida y se detuviera allí, no podría aprender los pesos que minimizan la función de costos. Si el algoritmo solo calculara una iteración, no habría aprendizaje real.



La retropropagación es clave para que el Perceptrón multicapa aprenda eficientemente. Permite ajustar los pesos de la red en cada iteración para minimizar la función de costos. Las funciones en cada neurona deben ser diferenciables para que funcione correctamente. En cada paso, se calcula el gradiente del error y se actualizan los pesos de la primera capa oculta. Este proceso se repite hasta que el gradiente converge para todos los pares de entrada-salida, asegurando un ajuste efectivo de los pesos en la red neuronal.



Implementando el Perceptrón Multicapa



```
] from sklearn.feature_extraction.text import TfidfVectorizer

  vectorizer = TfidfVectorizer(stop_words='english',
                               max_features= 10000, # máximo número de términos
                               max_df = 0.5,
                               smooth_idf=True)

  X = vectorizer.fit_transform(desconocidos['text_limp2'])
```

```
y = desconocidos['true-author']
# División de datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Lista de solvers
solvers = ['lbfgs', 'sgd', 'adam']
accuracies = []

for solver in solvers:
    # Creación y entrenamiento del modelo de perceptrón multicapa con el solver actual
    model = MLPClassifier(hidden_layer_sizes=(128, 64), max_iter=1000, activation='relu', solver=solver, random_state=1)
    model.fit(X_train, y_train)

    # Evaluación del modelo
    accuracy = model.score(X_test, y_test)
    accuracies.append(accuracy)
```

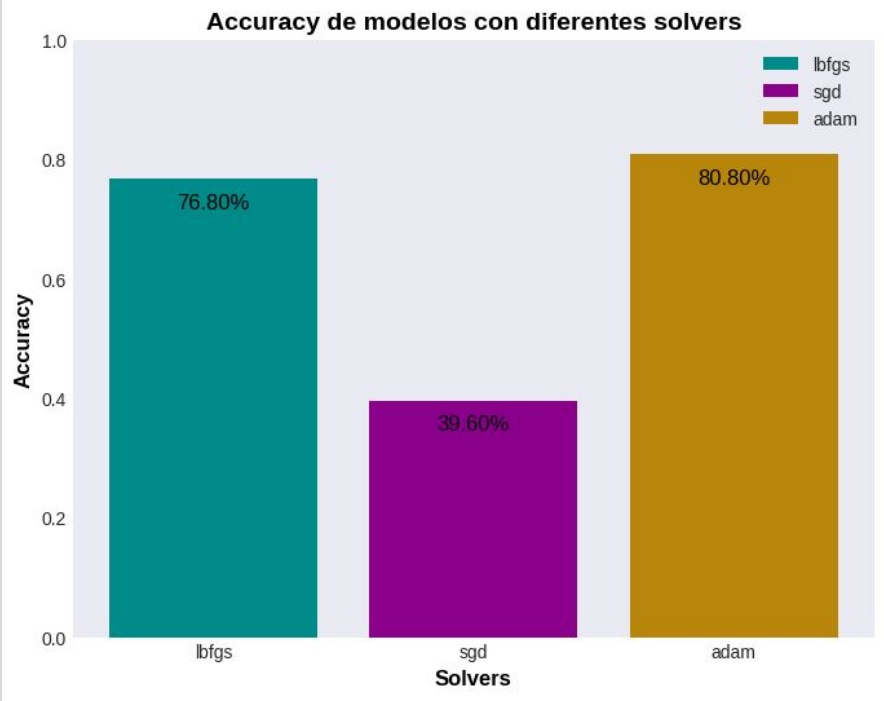
Modelo: Perceptrón Multicapa

Comparación de Resultados

Precisión del modelo

- ***Adam (Adaptive Moment Estimation):***
  Utiliza el concepto de momentos (primero y segundo) para adaptar la tasa de aprendizaje de cada parámetro de la red, suele ser rápido y eficiente en la convergencia, especialmente en conjuntos de datos grandes.
- ***L-BFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno):***
  Tiene la capacidad para manejar problemas con muchas variables y por su convergencia rápida en problemas que no son muy grandes en escala.
- ***SGD (Stochastic Gradient Descent):***
  Actualiza los parámetros de la red en función del gradiente de la función de pérdida con respecto a un ejemplo de datos a la vez, en lugar de usar el conjunto completo de datos.

Adam tiene la mejor precisión con un 80.80% de precisión, justa automáticamente la tasa de aprendizaje para cada parámetro de la red neuronal, lo que puede ser ventajoso en problemas con datos no uniformes o con gradientes variables. En el ***análisis de texto***, donde la estructura y la frecuencia de las palabras pueden variar mucho, esta adaptabilidad puede ayudar a encontrar un equilibrio adecuado entre explorar y explotar el espacio de búsqueda.



Reporte de clasificación del modelo Adam

Precisión: Es la medida de la fracción de identificaciones positivas correctas entre todas las identificaciones positivas hechas por el modelo para un autor en particular. En promedio, la precisión es del 80%.

Recall: Representa la fracción de las instancias relevantes que han sido recuperadas sobre el total de instancias relevantes. El recall promedio es del 80%, lo que significa que el modelo identifica correctamente alrededor del 80% de los textos de cada autor.

El modelo parece tener un rendimiento bastante sólido en la identificación de los autores basado en los textos, con valores de precisión, recall y F1-score alrededor del 80%. Sin embargo, algunas clases específicas pueden tener un rendimiento inferior, como se refleja en los puntajes más bajos para ciertos autores. (unos ejemplos están en color azul).

Reporte de Clasificación:

	precision	recall	f1-score	support
candidate00001	1.00	0.93	0.97	15
candidate00002	0.75	1.00	0.86	9
candidate00003	0.43	0.50	0.46	6
candidate00004	0.50	0.50	0.50	6
candidate00005	0.75	0.75	0.75	12
candidate00006	0.69	0.75	0.72	12
candidate00007	1.00	0.91	0.95	11
candidate00008	0.67	0.86	0.75	7
candidate00009	0.89	1.00	0.94	8
candidate00010	0.79	0.85	0.81	13
candidate00011	0.93	1.00	0.96	13
candidate00012	0.73	0.85	0.79	13
candidate00013	0.92	1.00	0.96	12
candidate00014	1.00	0.56	0.71	9
candidate00015	0.43	0.55	0.48	11
candidate00016	0.91	1.00	0.95	10
candidate00017	0.90	1.00	0.95	9
candidate00018	0.92	0.92	0.92	12
candidate00019	0.86	0.92	0.89	13
candidate00020	0.75	0.90	0.82	10
candidate00021	1.00	1.00	1.00	8
candidate00022	0.71	0.71	0.71	7
candidate00023	0.93	0.81	0.87	16
candidate00024	0.73	0.73	0.73	11
candidate00025	0.86	0.60	0.71	10
candidate00026	0.93	0.87	0.90	15
candidate00027	1.00	0.89	0.94	9
candidate00028	0.86	0.80	0.83	15
candidate00029	1.00	1.00	1.00	8
candidate00030	0.73	0.89	0.80	9
candidate00031	1.00	0.83	0.91	12
candidate00032	1.00	0.56	0.72	16
candidate00033	1.00	1.00	1.00	8
candidate00034	1.00	1.00	1.00	10
candidate00035	0.00	0.00	0.00	5
candidate00036	1.00	0.83	0.91	6
candidate00037	0.62	1.00	0.77	5
candidate00038	0.63	0.80	0.71	15
candidate00039	0.80	0.80	0.80	5
candidate00040	0.69	1.00	0.82	9
candidate00041	1.00	0.88	0.93	8
candidate00042	0.62	0.89	0.73	9
candidate00043	0.60	0.60	0.60	10
candidate00044	0.62	0.62	0.62	13
candidate00045	0.88	0.88	0.88	8
candidate00046	0.67	0.50	0.57	8
candidate00047	0.88	0.88	0.88	8
candidate00048	1.00	0.70	0.82	10
candidate00049	0.88	0.88	0.88	8
candidate00050	0.50	0.25	0.33	8
accuracy			0.81	500
macro avg	0.80	0.80	0.79	500
weighted avg	0.82	0.81	0.80	500

Resultados

```
Texto original:
Australia's anti-monopolies watchdog, examining claims of collusion against the four major banks, has asked them to explain why they all cut their mortgage rates by the same amount within hours of each other.
Australian Competition and Consumer Commission (ACCC) Chairman Allan Fels said it was unusual that the banks would all move so quickly to cut their rates by the same amount.
"We don't claim to have evidence of collusion. What we are doing is seeking an explanation from the banks," Fels said in a radio interview broadcast on Monday.
National Australia Bank, Australia and New Zealand Banking Group Ltd, Commonwealth Bank of Australia and Westpac Banking Corp all cut their standard variable mortgage rates by 0.35 percent to 7.2 percent on Friday, hours after the Reserve Bank cut official
Government and opposition politicians and consumer groups have since suggested the banks colluded to ensure they did not pass all of the official rate cut on to their customers.
Fels said the commission would ask for details of any communications between the banks on Friday and whether there had been any previous discussion about their approach to possible official rate cuts.
"The reason why we're seeking an explanation from the banks is that it is unusual for them all to have moved so quickly within a few hours of one and other and to have reduced their rates by the same limited amount of 0.35 (points) as against a Reserve Bank
"Occasionally, one player will do something a bit different," Fels said.
"But in this case the four big players, all of them, moved by exactly the same amount, despite having told so many of us over time that their circumstances, their costs and their strategies all differ."
Commonwealth Bank of Australia, the first big bank to cut mortgage rates after Friday's official rate cut, denied on Monday that it had colluded with the other major banks.
"I haven't discussed interest rates with the other banks," CBA Managing Director David Murray said.
"We have a very careful discipline within the bank, because we don't believe in collusion, we believe in competition," Murray said, adding he would cooperate fully with the ACCC.
Australian Prime Minister John Howard on Friday questioned the level of competition amoung the banks.
"It is pretty hard to believe we have a highly competitive banking system at the present time when there can be delays of the kind that we are apparently about to witness in the passing on in the reduction in official interest rates," Howard said.
Opposition politicians also criticised the banks, who have also come under heavy public fire this year for increasing bank fees to offset a squeeze on home lending margins.
Opposition Treasury spokesman Gareth Evans said it was hard to believe the banks were not colluding on rates. "It's a little like a line of chorus girls, all with impeccable timing raising their leg exactly so high," Evans told reporters.

Autor verdadero: candidate00001
Autor predicho: candidate00001
```

```
Texto original:
TVX Gold Inc is taking steps to join the ranks of major mining companies, bringing aboard three senior executives who pledge to turn the company into a million-ounce-a-year gold producer.
"We're here to help TVX grow," said Cliff Davis, the new senior vice-president of North American operations, among three officers who joined TVX in the past month.
"We understand the need for organization. We understand realistic plans. I think that's something we can help TVX with," Davis told Reuters in an interview.
Davis's next plan is to speed production at the New Britannia mine in Manitoba and develop more reserves at Casa Berardi in Quebec.
Joining Davis at TVX are David Murray, the recently appointed president and chief operating officer, and Ken Sangster as senior vice-president of European operations.
The three have worked together before. They did stints at mining giant RTZ-CRA Corp Plc and have years of experience with huge projects.
"For us it was a perfect fit," said TVX spokesman Ed Baer. "One of the things that was a common concern to the market was our lack of management."
Toronto-based TVX split off from Inco Ltd in 1993 and has concentrated on expanding its assets.
With its interests now growing in Canada, Greece, Brazil, Ecuador, Chile, the Czech Republic and Russia, the company is ready to switch its focus to production, he said.
TVX has promised its shareholders it will be producing 1.1 million ounces of gold a year by the year 2000 from 423,000 ounces in 1995.
To help keep its promise, the company has placed its bets on the Kassandra Mines in Greece. "If everything works for us, the story really lies in Greece," said Baer.
The property so far is expected to produce 300,000 ounces of gold a year at a cash cost of US$150 an ounce. TVX was still drilling and would not have a clear idea until next spring what the deposit offered, said Baer.
Uncertainty over Kassandra has unsettled TVX stock. Its shares peaked at 14.85 in Toronto during a road show and high gold bullion prices last winter. But the stock has since slid.
With its stock now trading around 10 on the Toronto Stock Exchange and around 7.50 on the New York Stock Exchange, TVX has been rumored as a takeover target for some major gold producers.
But as a growing company, TVX is always looking for joint ventures and is on the prowl for small mining companies with proven reserves, Baer said.
Davis's first move at TVX was to shut the company's gold mine in Montana earlier this month.
The Mineral Hill mine drained money from the company for years. TVX hoped to revive it with additional ore found at nearby Crevice Mountain, but ore grades at Mineral Hill grew worse and the company could not keep its mill supp
Although the company continues to explore at Crevice, Davis did not hold out much hope for a resurrection of mining on the property.
"If it's what it appears to be, we won't reopen," he said.

Autor verdadero: candidate00021
Autor predicho: candidate00021
```

Autor 00001 y 00021

Resultados

```
Texto original:
When thousands of auto dealers rolled into Las Vegas for their annual convention last year, there was much clamour over the looming onslaught of no-haggle superstores.
As dealers prepare for this year's gathering, which starts Saturday in Atlanta, the juggernaught is upon them.
Republic Industries Inc., the company controlled by billionaire investor Wayne Huizenga of Blockbuster Video fame, is on a buying spree that may have already made it the largest auto dealer group in the nation.
Within two months, Republic, which includes the AutoNation USA superstore chain, has gone from owning no new-car franchises to six dealer groups.
In the meantime, officials at CarMax, the auto superstore chain owned by Circuit City Stores Inc., said earlier this month they would consider buying existing dealerships and adding more new-car outlets as part of their expansi
"I think it's got a lot of dealers rattled," said John Sinclair, owner of Sinclair Ford of Festus in Festus, Mo.
The sweeping changes prompted industry trade magazine Automotive News to ask in its most recent edition: "Apocalpyse Now?"
Before its latest purchase, Montgomery Securities estimated Republic's 1997 franchised dealership revenues at $2.5 billion. That puts it ahead of the previous largest group, Hendrick Automotive Group, which has estimated 1997 r
The explosive growth of Republic, CarMax and others comes as Detroit's Big Three automakers have backed off their traditional resistance to publicly owned dealerships.
Ford Motor Co. Chairman Alex Trotman told dealers in Las Vegas a year ago that the No. 2 automaker had no plans to follow Chrysler Corp. and award a new car franchise to an auto superstore.
But the automaker reversed its position in December when it formed an alliance with Republic that resulted in the sale of Ohio-based Mullinax Management, one of the country's biggest Ford dealers, to Republic.
Ford Chief Financial Officer John Devine said earlier this week the automaker still believes the private enterprise system is the best for the industry.
But he added, "We also have to recognise that changes are here and more changes are going to come."
How profitable and successful the superstores ultimately become remains to be seen. But in the short term, they are helping turn the tradition-bound retail car industry upside down. Armed with capital, the auto superstores have
The results are huge lots of high-quality, pre-certified used cars. Buyers who dread walking into an old-fashioned car dealerships enter the new superstores to find user-friendly computer kiosks and childrens' play areas. Salar
"I think there's going to be a revolution in the car business," Fred Ricart, owner of Ricart Ford in Columbus, Ohio, said earlier this month during a panel discussion at the Automotive News World Congress.
Ricart, whose own dealership is structured in an auto mall format, believes traditional dealers can overcome their negative sterotypes and compete. But he says they have to find new ways of improving and measuring customer loya
"The advantage the superstores have is they've got a clean slate going in," he said.
The National Automobile Dealers Association's 80th Annual Convention and Equipment Exposition will be held at the Georgia World Congress in Atlanta, from Feb. 1-4.

Autor verdadero: candidate00003
Autor predicho: candidate00018
```

```
Texto original:
British soldiers fought their way through jungle, attacked an enemy camp and took prisoners in Hong Kong on Tuesday, but it was only a mock battle and the last such exercise by the army after 156 years of colonial rule.
With only 70 days left before Britain hands Hong Kong back to China, the Black Watch regiment headed to the hills of the New Territories to play jungle war games and ignore, for a few hours, the fast-approaching end of British
Shots ripped through the air and red flares spewed smoke as sweat-drenched soldiers laden with gear waded across rivers, charged up and down hills and crawled through lush and heavy undergrowth in a practice assault on an enem
The Black Watch -- dubbed the "Highland Furies" by the French after the Battle of Fontenoy in 1745 -- are nearly half-way through an historic tour of duty in Hong Kong.
On June 30, the pipes and drums of the regiment will sound the end of British colonial rule.
But on Tuesday, sweat and slog rather than ceremony was the order of the day.
The camp and two prisoners-of-war were successfully taken, marking the end to the British army's permanent access to a valued training ground for jungle warfare skills.
"I'll miss it. We'll miss it, because you don't get training like this on a regular basis on a tour," said Corporal John Lyon, 28. "This area you can train throughout the whole time you're here."
Lieutenant-Colonel Alasdair Loudon, commanding officer of the Black Watch in Hong Kong, said the New Territories have become something of a secret weapon.
Few would guess that less than one-hour's drive from the concrete jungle of Hong Kong Island lie stretches of hot, humid and heavily overgrown countryside which represents a challenge for British troops.
"For us, it's harsh terrain. We're not used to the heat, we're not used to the hills and when you get away from tracks the undergrowth is very difficult. So it's demanding," he said.
The Black Watch also had a busy day on Monday -- they were escorting an advance party of China's People's Liberation Army (PLA) soldiers as they drove from the border to the Prince of Wales barracks in the territory's Central b
The 40-strong PLA party are to lay the ground for up to 10,000 Chinese troops likely to be stationed in Hong Kong after the handover.
Loudon said his regiment's main task as June 30 approaches is to ensure the handover goes smoothly.
"Our most important task is to get out on June 30 and have a fine parade so that the handover of British sovereignty has gone in a smooth and dignified way," he told reporters.
At least 100 of the Black Watch soldiers will be involved in the ceremonies with another 100 providing logistical support. The regiment flies back to Inverness in northern Scotland after the ceremonies.
"To a man, they will have had a very interesting time here. They will have done some good training, they've done a good job," Loudon said.
"And I'm sure when they get back to Inverness, where it's warm only two days a year, they're going to miss Hong Kong very much," he said.

Autor verdadero: candidate00015
Autor predicho: candidate00050
```

Modelo: Autor 00003 y 00015

Conclusiones:

Discusión:

- ¿Cuál es la complicación en esta aplicación?

- ¿En qué medida éste algoritmo que escogimos funciona para el problema planteado?

- Una de las principales complicaciones en esta aplicación es la diversidad de estilos literarios. Cada autor tiene su propio estilo único, que se puede manifestar en una variedad de características estilísticas, como la longitud de las oraciones, el uso de verbos, adjetivos y adverbios, la complejidad de la sintaxis, la frecuencia de las palabras, etc.

- Otra complicación es la variabilidad del estilo de un autor a lo largo del tiempo. Los autores pueden cambiar su estilo a medida que maduran o experimentan cambios en su vida. Por ejemplo, un autor puede escribir de forma más formal en sus primeros trabajos y de forma más informal en sus trabajos posteriores.

- Por último, la complicación de esta aplicación es la presencia de ruido en los datos. El ruido puede ser causado por una variedad de factores, como la presencia de errores tipográficos, la ambigüedad del lenguaje, o la influencia de otros autores.

En el caso de la determinación de la autoría de textos, el perceptrón multicapa puede utilizarse para aprender patrones en el estilo de escritura de diferentes autores. Estos patrones pueden incluir la longitud de las palabras, la frecuencia de las letras y la estructura de las oraciones.
En nuestras pruebas, obtuvimos resultados prometedores con el perceptrón multicapa. Logramos una precisión de aproximadamente el 80% en un conjunto de datos de prueba. Esto sugiere que el perceptrón multicapa es un modelo eficaz para la determinación de la autoría de textos.
Sin embargo, es importante tener en cuenta que la precisión del perceptrón multicapa puede variar en función del conjunto de datos utilizado. Si el conjunto de datos es pequeño o si los autores tienen estilos similares, la precisión del modelo puede ser menor.

Conclusiones:

Trabajo futuro:

Para entender un poco más los textos que no fueron correctamente clasificados podemos aplicar otras técnicas como ***la búsqueda de elementos similares,*** *e ir variando la métrica de estas similitudes (Coseno, Jaccard, euclidiana, minmax),* también podemos hacer un modelado de tópicos con ***SVD*** o ***LDA*** que nos ayude a identificar clusters de palabras, de esa forma ver si las características del texto es semejante con las de otro texto.

Ejemplo de búsqueda de elementos similares

Definimos la función para la similitud coseno

```
def similitud_coseno(x, y):
  x = x.toarray()[0]
  y = y.toarray()[0]
  pnorma = (np.sqrt(x @ x) * np.sqrt(y @ y))

  if pnorma > 0:
    return (x @ y) / pnorma
  else:
    return np.nan
```

```
sims = fuerza_bruta(bolsas[1:], bolsa_dc, similitud_coseno)
```

```
print('Similitud máxima es {0} de documento {1}'.format(np.nanmax(sims), np.nanargmax(sims)+ 1))
```

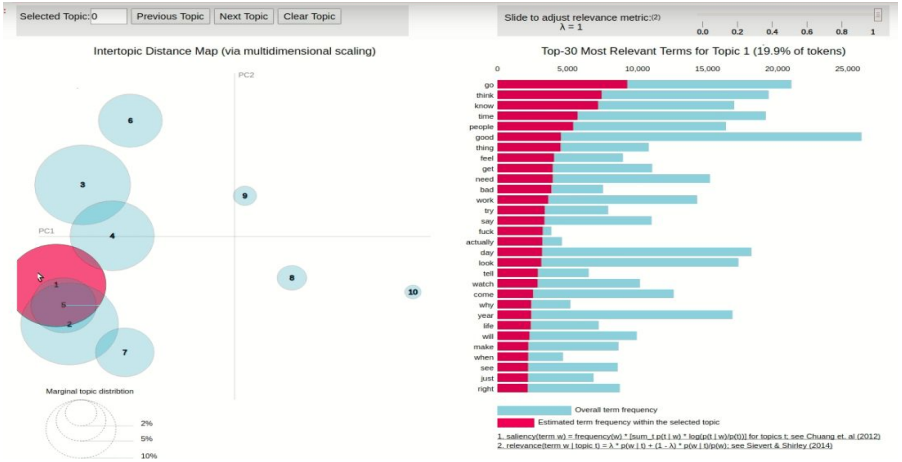
```
Similitud máxima es 0.4589534811637672 de documento 6997
```

Revisamos documento más similar

```
print(db.data[np.nanargmax(sims) + 1])
```

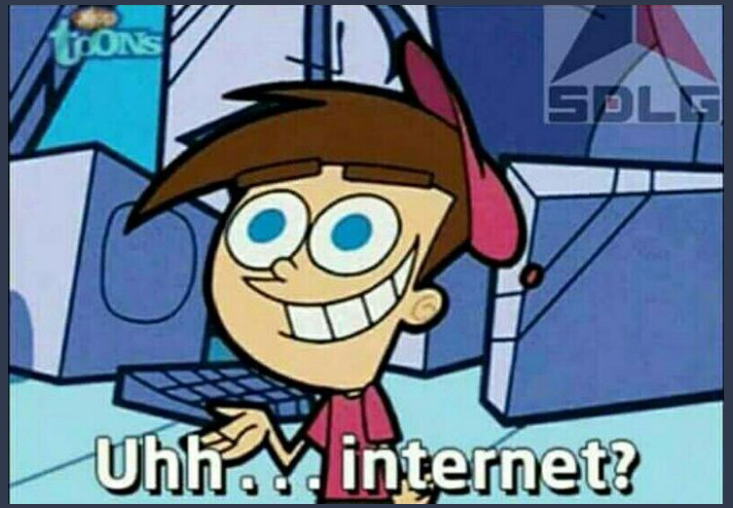
```
Perhaps instead of this silly argument about what backup lights
are for, couldn't we agree that they serve the dual purpose of
letting people behind your car know that you have it in reverse
and that they can also light up the area behind your car while
you're backing up so you can see?
```

Ejemplo de modelado de tópicos LDA



Referencias:

- Houvardas, J., & Stamatatos, E. (2006). Houvardas06 Author Identification: C50-Attribution [Conjunto de datos]. Zenodo. https://zenodo.org/records/3759068
- scikit-learn. (2023). sklearn.linear_model.Perceptron. Recuperado de https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Perceptron.html
- scikit-learn. (2023). sklearn.feature_extraction.text.TfidfVectorizer. Recuperado de https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html
- Bento, C. (2021, 21 de septiembre). Multilayer Perceptron Explained with a Real Life Example and Python Code: Sentiment Analysis. Towards Data Science. Recuperado de https://towardsdatascience.com/multilayer-perceptron-explained-with-a-real-life-example-and-python-code-sentiment-analysis-cb408ee93141
- González, M. V. (2014-15). El uso del Perceptrón Multicapa para la clasificación de patrones en conductas adictivas. Facultat de Psicologia. Recuperado de https://dspace.uib.es/xmlui/bitstream/handle/11201/1126/TFG%20Marta%20Vidal%20Gonz%C3%A1lez.pdf?sequence=1



Referencias:

- IBM. (s/f). Perceptrón multicapa. Recuperado de https://www.ibm.com/docs/es/spss-statistics/saas?topic=trees-multilayer-perceptron
- LibreTexts Español. (s/f). Perceptrón Multicapa. En Las matemáticas de la inteligencia artificial. Recuperado de https://espanol.libretexts.org/Matematicas/Las_matematicas_de_la_inteligencia_artificial/06%3A_Redes_Neuronales/6.3%3A_Perceptron_Multicapa
- Tech Lib. (2023). Perceptrón multicapa (MLP) - Definición y explicación. Recuperado de https://techlib.net/techedu/perceptron-multicapa-mlp/