



# **PROYECTO DE MACHINE LEARNING**

# **ANÁLISIS DE SENTIMIENTO**

**ANDRADA\_MARCUSANU**  
**DATA\_SCIENCE\_BOOTCAMP\_\_\_THE\_BRIDGE\_\_\_MADRID\_2024**



# ANÁLISIS DE SENTIMIENTO



Una tarea básica en análisis de sentimientos es clasificar la polaridad de un texto dado — si la opinión expresada en un documento es **positiva, negativa, o neutra**.



El análisis de sentimiento tiene múltiples aplicaciones, entre las cuales realizar análisis de reputación para empresas, estimaciones de popularidad en campañas electorales y ayuda en **identificar tendencias en los mercados financieros**.



Like



Love



Haha



Wow



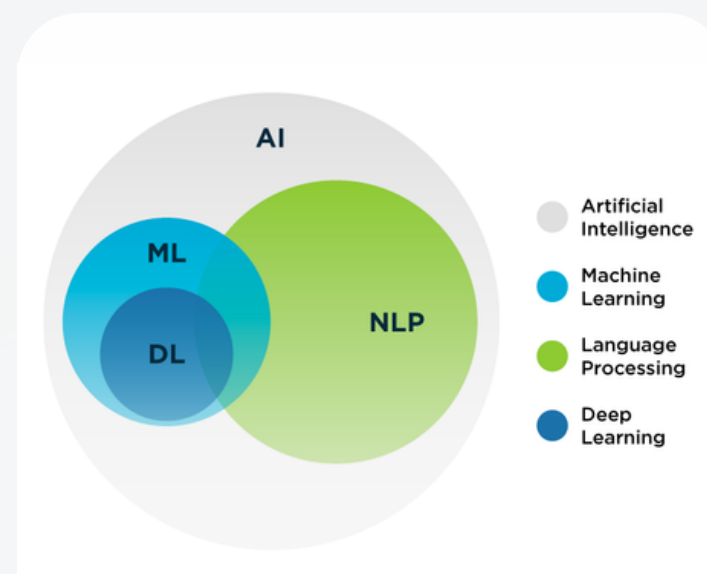
Sad



Angry

# NLP & MODELOS DE ML

Analizamos un dataset de **tweets** con temática financiera, previamente etiquetados.



NLP

Natural Language  
Processing

## Entrenados con nuestros datos:

1. Logistic Regression
2. Random Forest Classifier
3. NNs:
  - a. Feed-forward NN
  - b. CNN
  - c. LSTM

## Modelos pre-entrenados de NLP:

1. VADER
2. roBERTa

ALGORITMOS ML

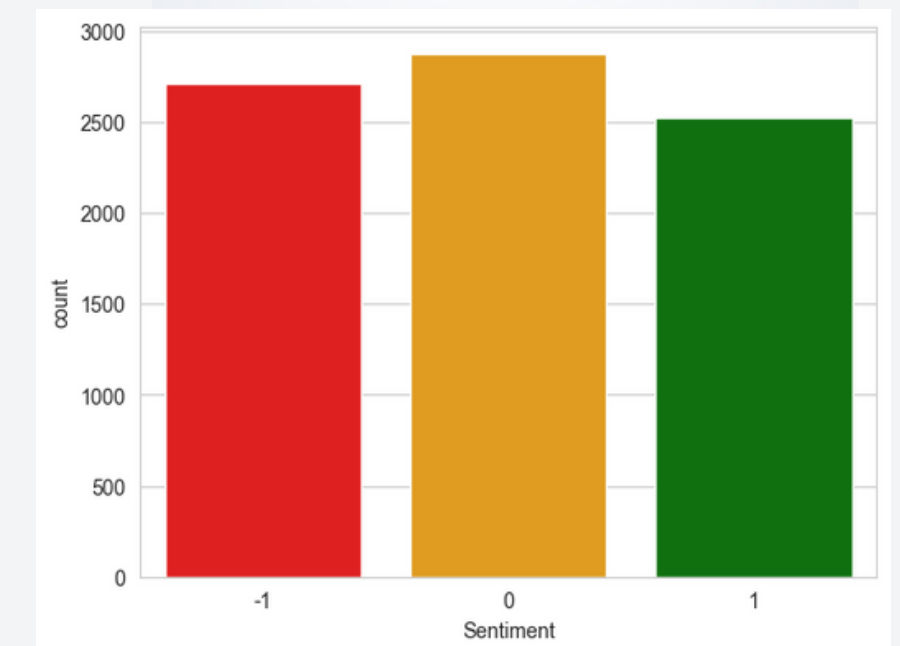
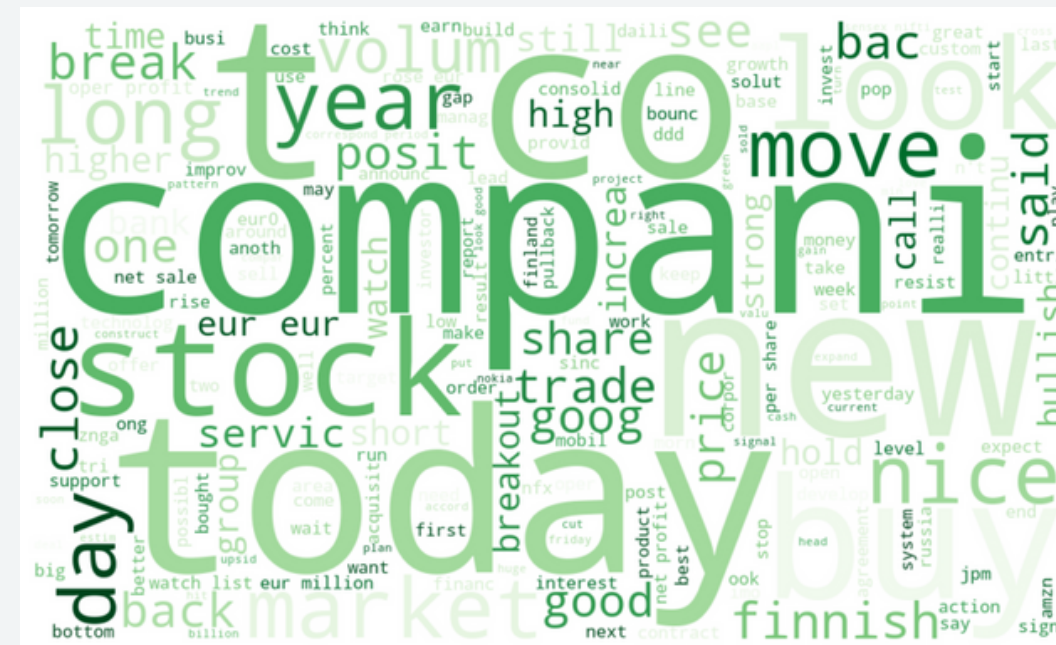


NEW DATA



Text	Sentiment
ere the company is growing	neutral

	Text	Sentiment
0	According to Gran , the company has no plans to move all production to Russia , although that is where the company is growing	neutral
1	Technopolis plans to develop in stages an area of no less than 100,000 square meters in order to host companies working in computer technologies and telecommunications , the statement said	neutral
2	The international electronic industry company Elcoteq has laid off tens of employees from its Tallinn facility ; contrary to earlier layoffs the company contracted the ranks of its office workers , the daily Postimees reported	negative
3	With the new production plant the company would increase its capacity to meet the expected increase in demand and would improve the use of raw materials and therefore increase the production profitability	positive
4	According to the company 's updated strategy for the years 2009-2012 , Basware targets a long-term net sales growth in the range of 20 % -40 % with an operating profit margin of 10 % -20 % of net sales	positive



# DATASET - NLP PREPROCESSING

## NLP PRE-PROCESSING

- Sentence Segmentation
- Words Tokenization
- Text Lemmatization
- Stop Words
- Dependency Parsing in NLP
- Named Entity Recognition
- Coreference Resolution

*"Jenna went back to University."*

Normalize

→ *"jenna went back to university"*

Tokenize

→ *<"jenna", "went", "back", "to", "university">*

Remove  
Stop Words

→ *<"jenna", "went", "university">*

Stem /  
Lemmatize

→ *<"jenna", "go", "univers">*

# ALGORITMOS DE ML

**CountVectorizer()**

**GloVe Embedding**

Log Reg	Random Forest Classifier	Forward-Feed NN	CNN	LSTM	VADER	roBERTa
76.52 %	74.80 %	36.61 %	37.12 %	26.79 %	46.62 %	41.87 %

**ACCURACY SCORES**



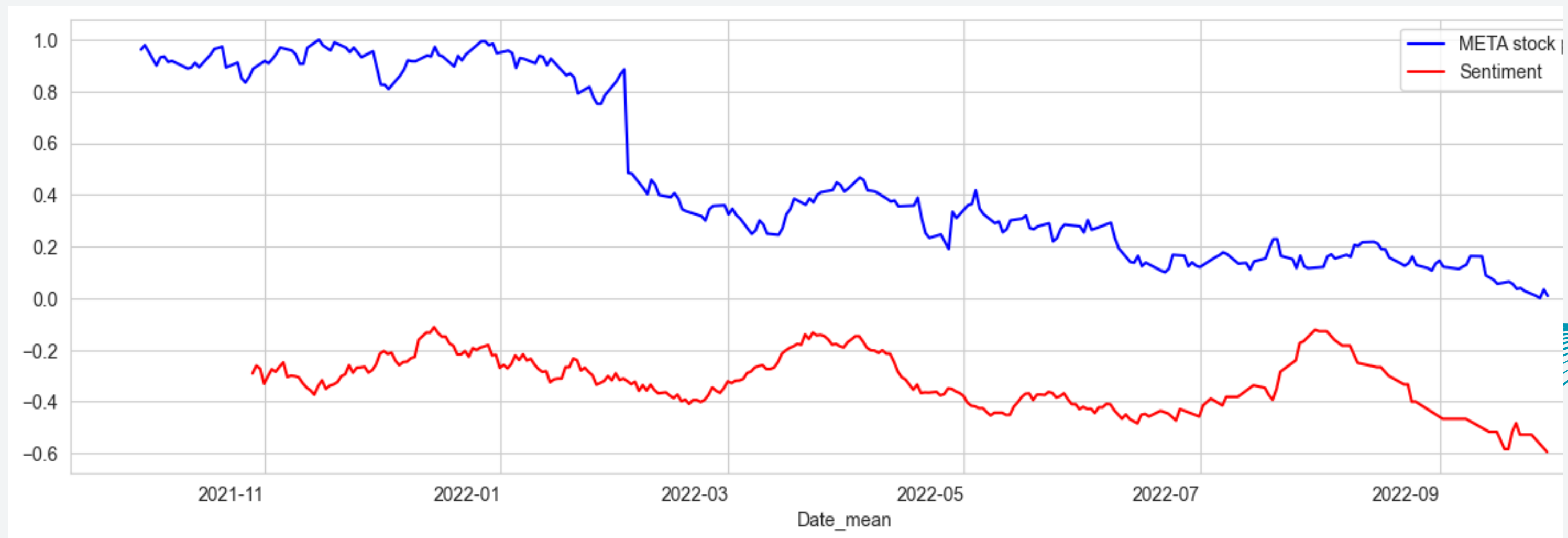
# PREDICCIONES CON NUEVOS DATOS

Dataset que contiene tweets scrapeados, publicados entre septiembre 2021 y septiembre 2022, sobre las 25 más importantes empresas de EEUU

	Date	Tweet	Stock Name	Company Name
0	2022-09-29 23:41:16+00:00	Mainstream media has done an amazing job at brainwashing people. Today at work, we were asked what companies we believe in & I said @Tesla because they make the safest cars & EVERYONE disagreed with me because they heard "they catch on fire & the batteries cost 20k to replace"	TSLA	Tesla, Inc.
1	2022-09-29 23:24:43+00:00	Tesla delivery estimates are at around 364k from the analysts. \$tsla	TSLA	Tesla, Inc.
2	2022-09-29 23:18:08+00:00	3/ Even if I include 63.0M unvested RSUs as of 6/30, additional equity needed for the RSUs is $63.0M \times \$54.20 = \$3.4B$ . If the deal closed tomorrow at \$54.20, Elon would need \$2.0B for existing shares plus \$3.4B for RSUs, so \$5.4B new equity. \$twtr \$tsla	TSLA	Tesla, Inc.
3	2022-09-29 22:40:07+00:00	@RealDanODowd @WholeMarsBlog @Tesla Hahaha why are you still trying to stop Tesla FSD bro! Get your shit together and make something better? Thats how companies work, they competed. Crying big old ass fart clown!	TSLA	Tesla, Inc.
4	2022-09-29 22:27:05+00:00	@RealDanODowd @Tesla Stop trying to kill kids, you sad deranged old man	TSLA	Tesla, Inc.

# SENTIMIENTO VS. REALIDAD

Recuperando el gráfico del valor histórico de las acciones de estas empresas, intentaremos averiguar si hay algún tipo de relación entre el sentimiento detectado en la opinión pública y el valor de las acciones - **ejemplo: META**





# SENTIMIENTO VS. REALIDAD

Recuperando el gráfico del valor histórico de las acciones de estas empresas, intentaremos averiguar si hay algún tipo de relación entre el sentimiento detectado en la opinión pública y el valor de las acciones - **ejemplo: NETFLIX**



# PRÓXIMOS PASOS

- Analizar con más detalle **la semántica de los tweets** de este tópico para realizar una vectorización superior
- Intentar **ampliar el dataset de tweets** para abarcar un período más amplio (**webscraping**)
- Recopilar información de otras fuentes (otras redes sociales, **portales de noticias**, etc.)
- Profundizar el **estudio de NLP y Análisis de sentimiento** en el ámbito financiero
- **Probar más modelos**, tanto entrenados con nuestros datos como pre-entrenados

