

Predicción de acciones usando modelado de tópicos y técnicas de análisis de sentimiento: Un caso de estudio de Machine Learning en Ecopetrol

Edwar Valenzuela Cortés
Santiago Puentes Núñez



UNIVERSIDAD
NACIONAL
DE COLOMBIA



STATDATA
—HUB—



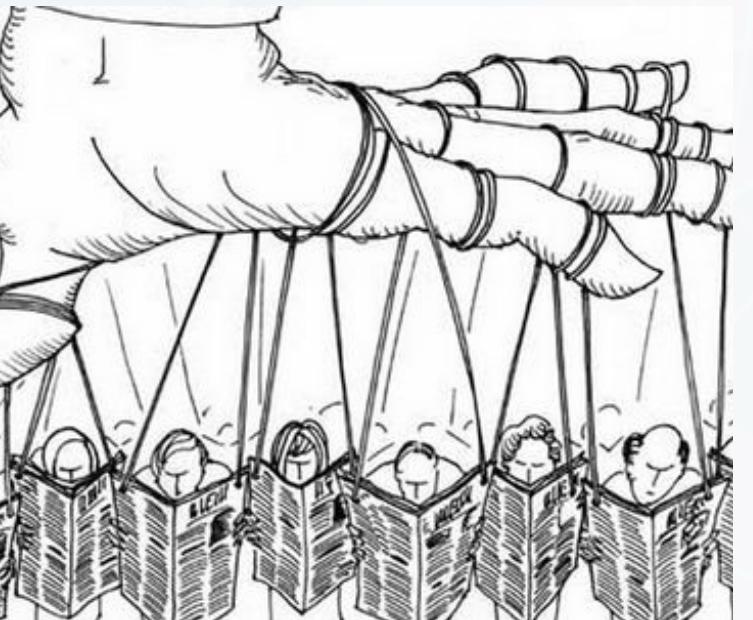
UACE

Problemáticas

Exceso de información



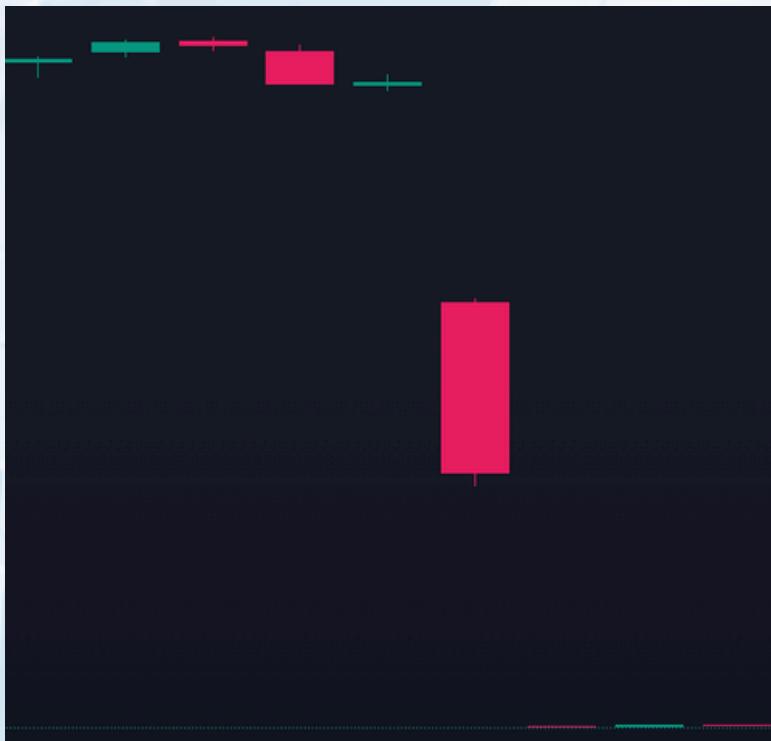
Manipulación



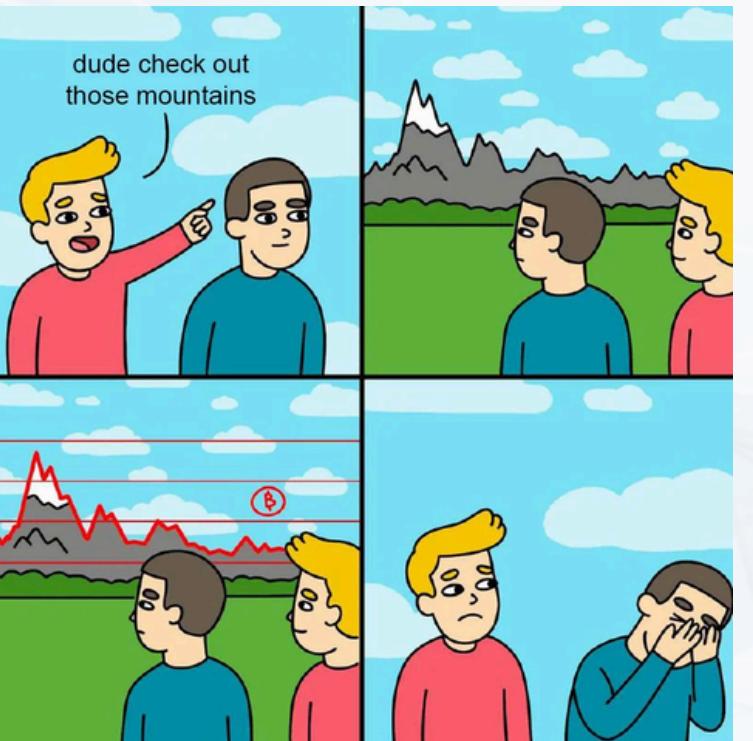
Tomar decisiones



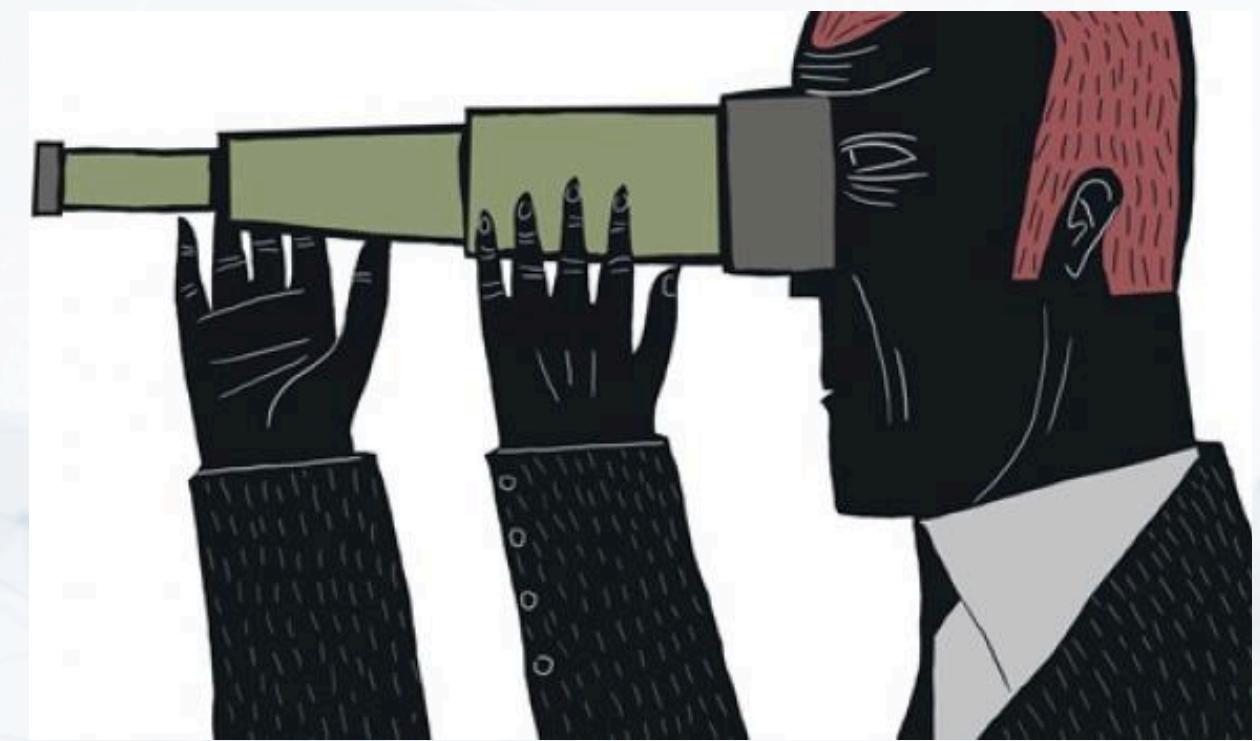
Sin tiempo para reaccionar



Emociones



Sesgo retrospectivo



Pregunta problema

¿Qué nos permite determinar el impacto de un evento en el precio de determinado activo sin dar a lugar interpretaciones subjetivas?

Hipotesis: El impacto de estas situaciones puede ser medido relacionando distintos eventos o topicos similares entre sí a lo largo del tiempo con el precio de un activo

Trabajo clave: Predicción

Stock market prediction using machine learning classifiers and social media, news

Wasiat Khan¹ · Mustansar Ali Ghazanfar² · Muhammad Awais Azam³ · Amin Karami² · Khaled H. Alyoubi⁴ · Ahmed S. Alfakeeh⁴

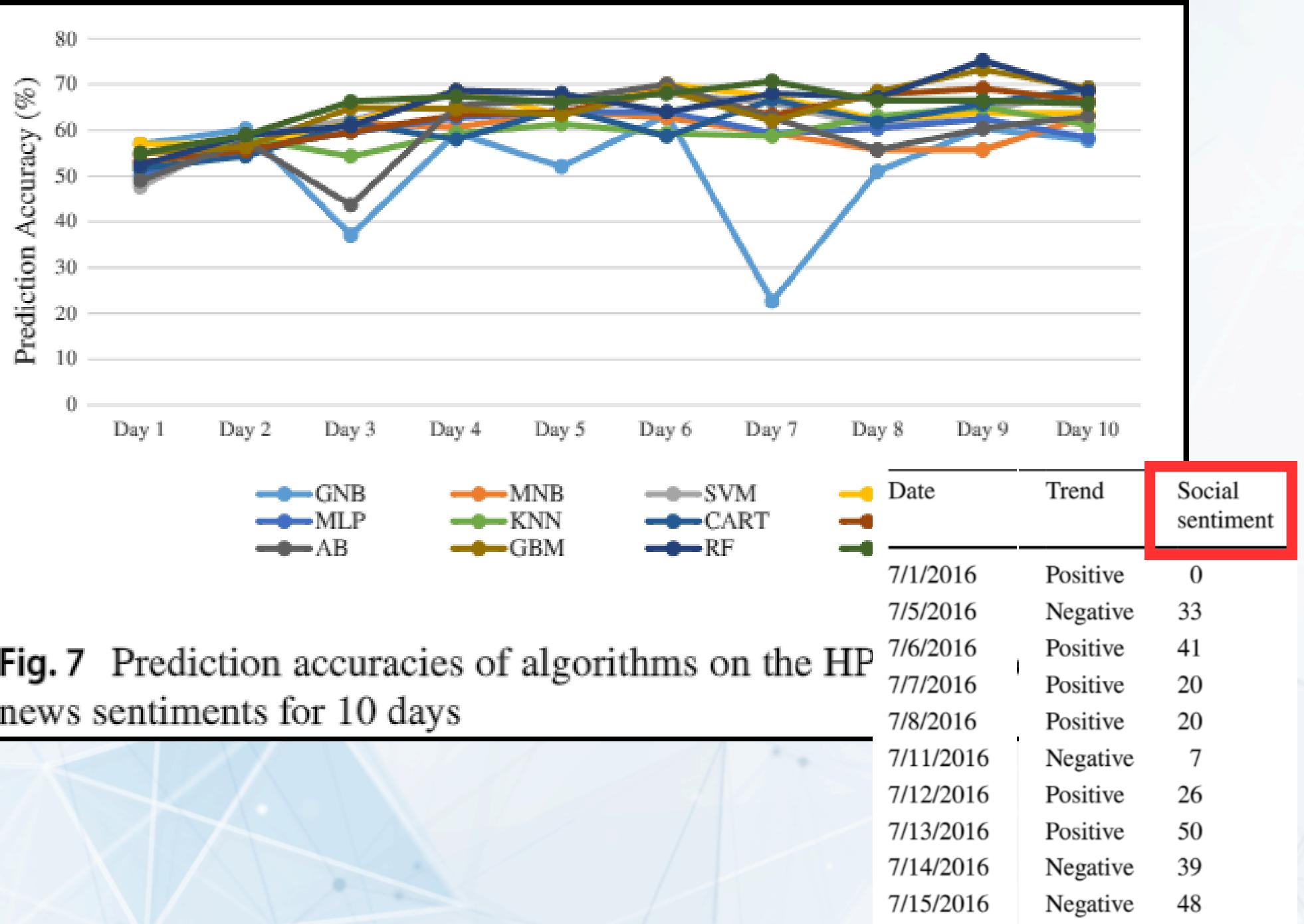


Fig. 7 Prediction accuracies of algorithms on the HP news sentiments for 10 days

MODELO DE RECOMENDACIÓN PARA INVERSIÓN EN ACCIONES COLOMBIANAS PERTENECIENTES AL ÍNDICE COLCAP BASADO EN ANÁLISIS TÉCNICO Y SENTIMIENTO DEL MERCADO LOCAL

Julian Palacio Roldan

	Entrenamiento	Validación
Accuracy	70.80%	66.41%

Columna	Descripción
date	Fecha de referencia
equity_id	Identificador de la acción
close	Precio de cierre
MM5	Media móvil 5 días
MM20	Media móvil 20 días
positive	# Noticias positivas
neutral	# Noticias neutrales
negative	# Noticias negativas

Trabajo clave: Extracción de información

FINBERT: FINANCIAL SENTIMENT ANALYSIS WITH PRE-TRAINED LANGUAGE MODELS
SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE
DOGU ARACI

Leveraging State-of-the-Art Topic Modeling for News Impact Analysis on Financial Markets: A Comparative Study
by Weisi Chen ^{1,*}  , Fethi Rabhi ² , Wenqi Liao ¹  and Islam Al-Qudah ³ 

Table 2: Experimental Results on the Financial PhraseBank dataset

Model	All data			Data with 100% agreement		
	Loss	Accuracy	F1 Score	Loss	Accuracy	F1 Score
LSTM	0.81	0.71	0.64	0.57	0.81	0.74
LSTM with ELMo	0.72	0.75	0.7	0.50	0.84	0.77
ULMFit	0.41	0.83	0.79	0.20	0.93	0.91
LPS	-	0.71	0.71	-	0.79	0.80
HSC	-	0.71	0.76	-	0.83	0.86
FinSSLX	-	-	-	-	0.91	0.88
FinBERT	0.37	0.86	0.84	0.13	0.97	0.95

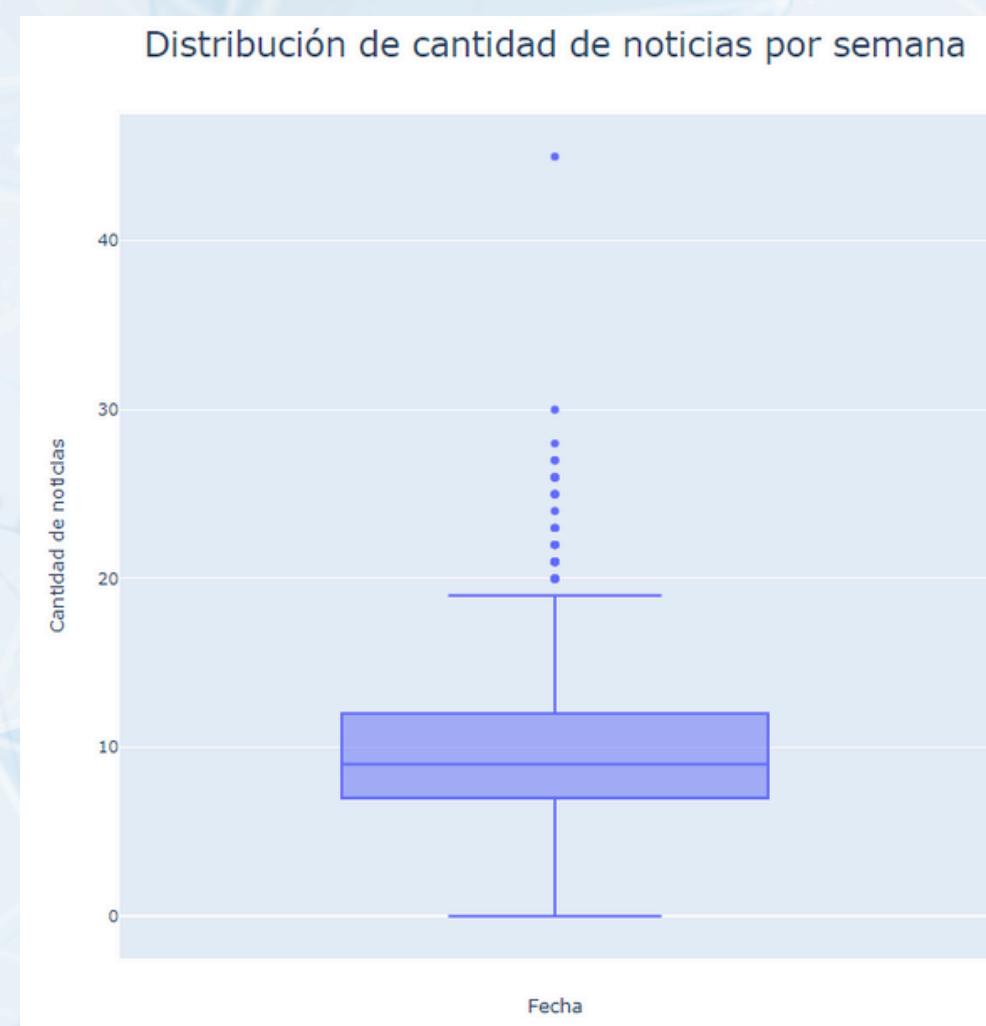
Table 9. Comparison between selected topic modeling models.

Topic Model	No. of Topics	Training Time (s)	Coherence (c_v)	Coherence (u_mass)
LDA 20-1	20	69.44	0.575	-16.939
LDA 50-1	50	331.99	0.582	-16.942
LDA 100-1	100	472.75	0.590	-17.197
LDA 500-1	500	777.06	0.591	-17.246
LDA 20-10	20	237.78	0.586	-17.230
LDA 50-10	50	1363.26	0.586	-17.140
LDA 100-10	100	2052.66	0.586	-17.217
LDA 500-10	500	4244.94	0.591	-17.222
Top2Vec	444	3321.50	0.545	-6.957
BERTopic	608	2632.65	0.823	-1.156

Datos de noticias

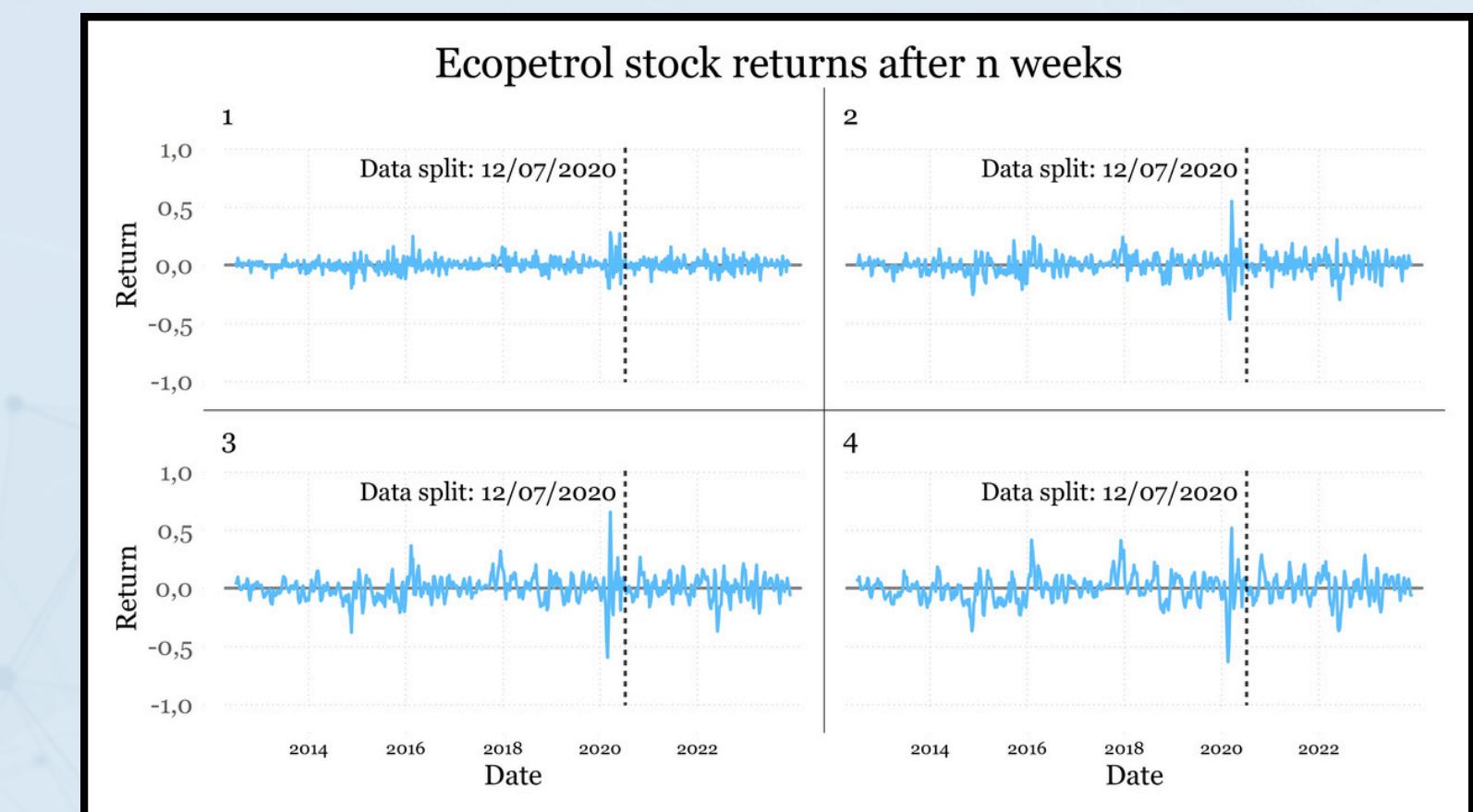
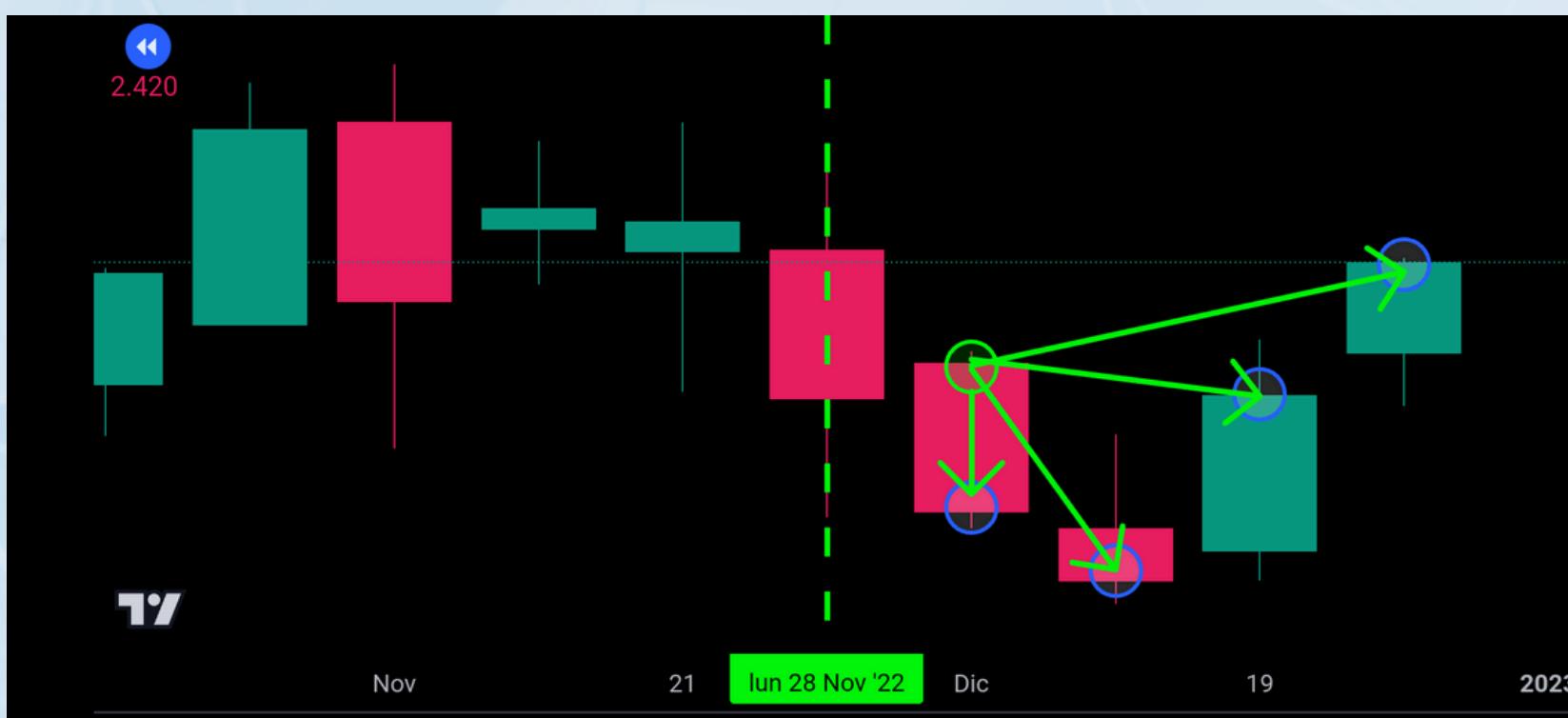
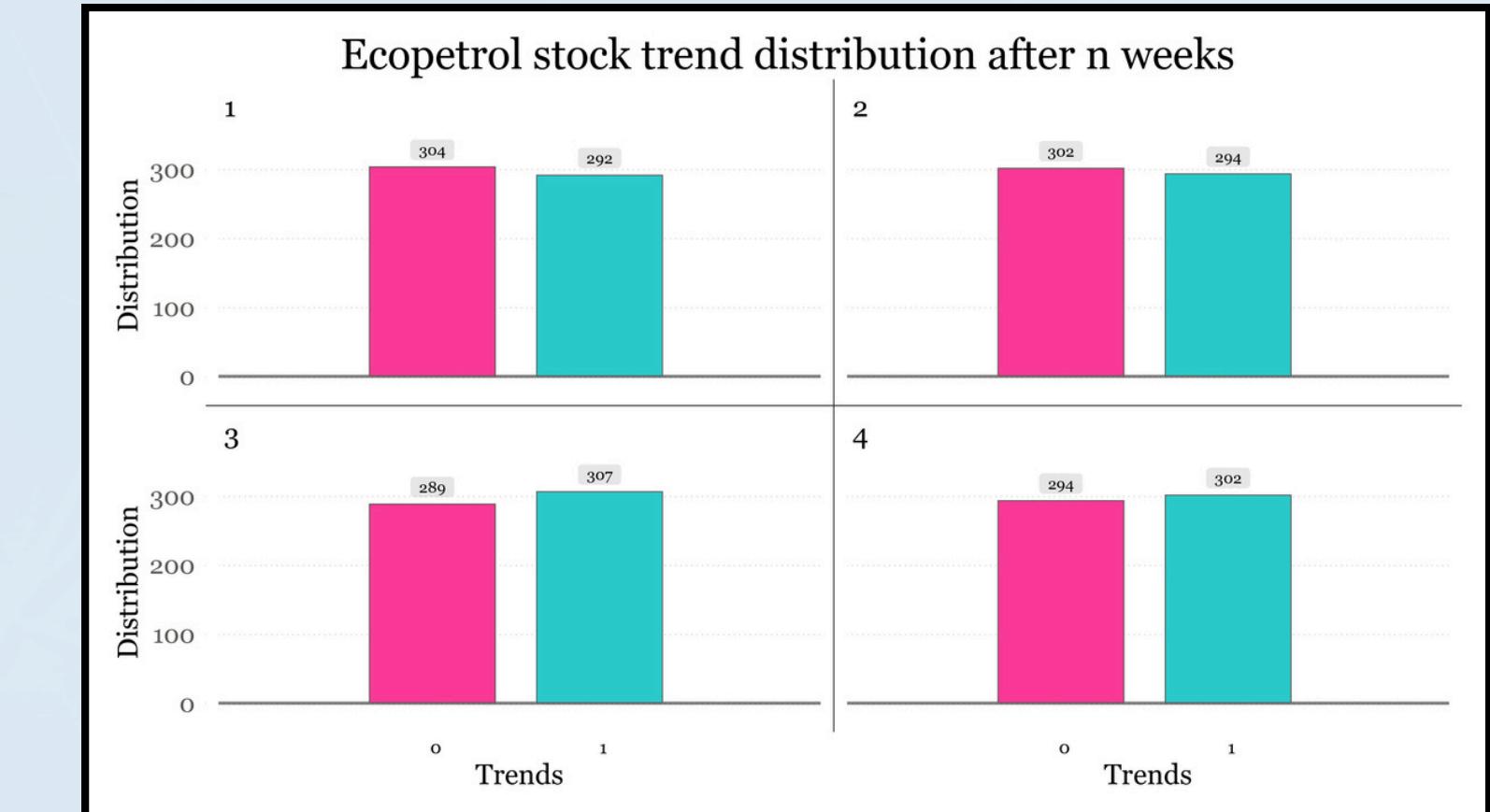
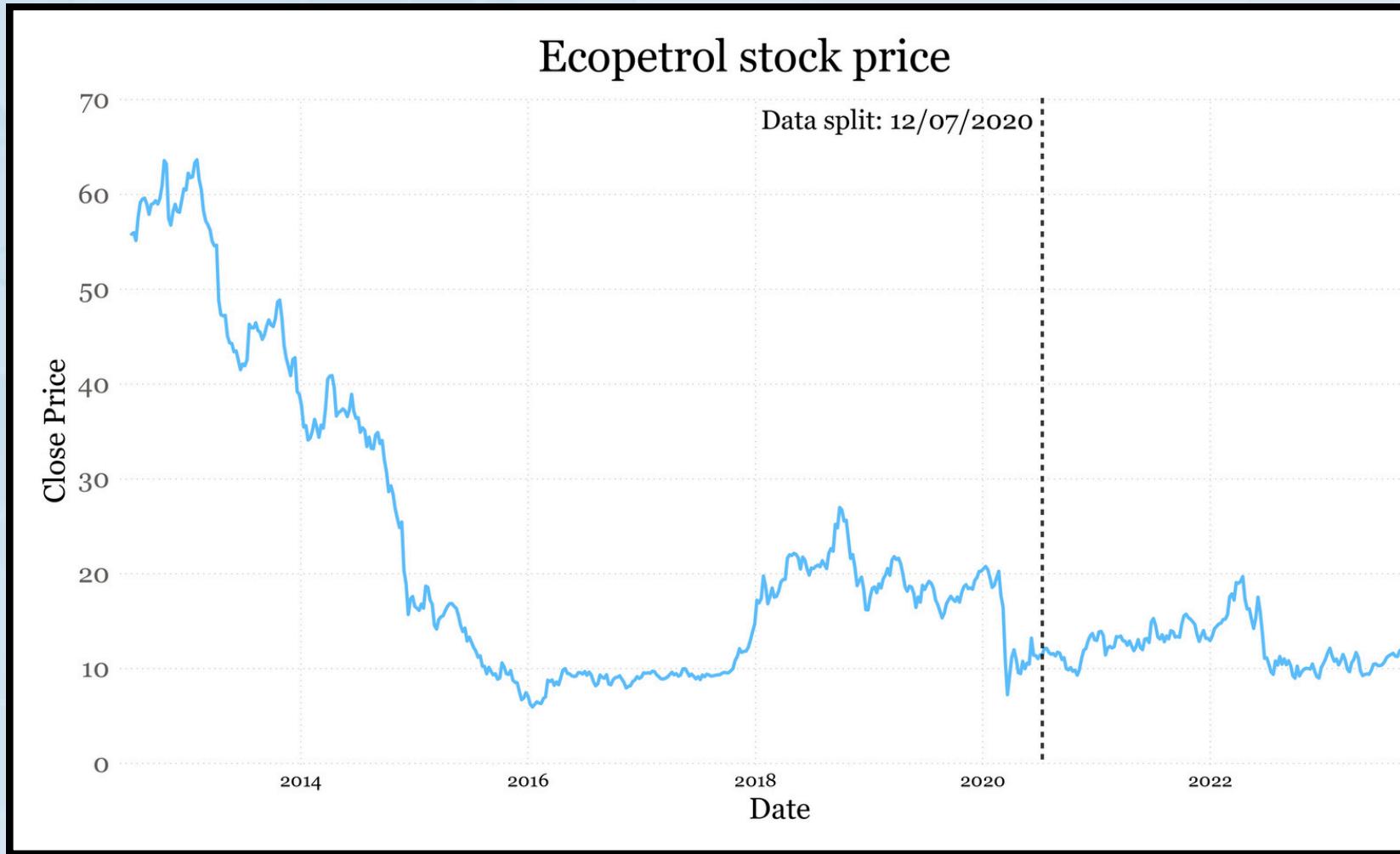
ENERGÍA | Ago. 01, 2024
Junta de Ecopetrol desistió de comprar activos de CrownRock, de propiedad de OXY
Brayan Xavier Becerra
Ecopetrol CrownRock Ricardo Roa

Date	Title
2024-07-31	Analysts do not see any failure in gas imports...
2024-08-01	Ecopetrol Board decided not to buy CrownRock a...
2024-08-01	Ecopetrol announced changes in senior manageme...

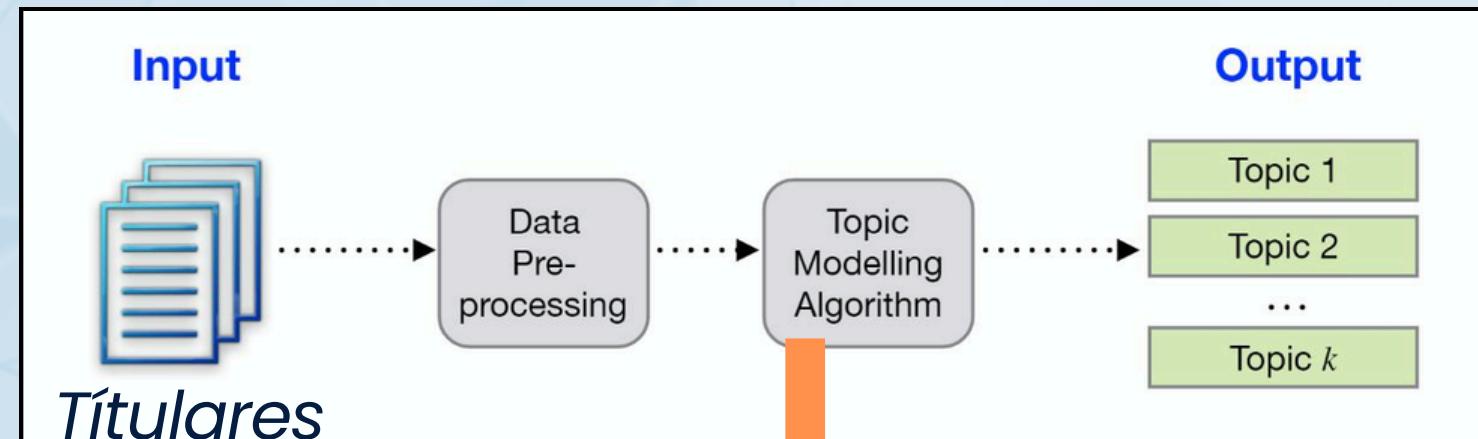


- 6300 artículos procesados entre 2012 y 2024 de **La República e Hydrocarbons**
- Se hace la traducción para hacer más fácil los siguientes pasos. Se eliminan palabras cortas y caracteres especiales
- Se asigna el 70% para **entrenamiento (2012-07-01 y 2020-07-12)**, y el 30% para **prueba (2020-07-12 y 2023-12-31)**

Variables de precios



Modelado de temas: BERTopic



Fecha	Títtular	Topico
2023-12-12	Ecopetrol suffers its biggest drop on Wall Street since 2020	0
2023-12-12	Natural gas industry best on reducing CO2 emissions	4

Date	Topic ₁	Topic ₂	...	Topic _n
1/07/2012	1	0	...	0
8/07/2012	0	1	...	1
15/07/2012	1	0	...	0
22/07/2012	1	0	...	1
29/07/2012	0	0	...	0

Topic	Name
0	Ecopetrol Financial Performance Analysis
1	Colombia's Fracking Policy and Stock Market Impact
2	Oil Refinery Operations and Environmental Concerns
3	ECP Echeverry Plan for Export Strategy
4	Fuel Price Fluctuations and Natural Gas Market
5	USO Strike and Labor Disputes in Putumayo
6	Corporate Social Responsibility Projects Overview
7	Rubiales and Pacific Field Reversion Decisions
8	Pipeline Security and Contingency Planning
9	Continuing Corruption Scandal at Reficar
10	Perception and Outlook on Progress
11	ANLA and ANH License Issues
12	Geopark's Role in Industry and Community Relations
13	Security Challenges in Cao, ELN, and Limn Regions

14	Environmental Concerns along the Magdalena River
15	Security and Environmental Issues Reporting
16	Blockades in Barrancabermeja and Mininterior Response
17	Analysis of Colcap Index and Market Performance
18	Bayn, ECP, and Bayon Perspective Analysis
19	Parex Reserves Growth and Agreements
20	Tax Reform Impact and Fiscal Works
21	Energy and Mining Sector Transition to Renewables
22	Lizama Outcrop Scandal and Incident Investigation
23	Monthly Reports from Inner Circle
24	Local Hiring Practices and Labor Trends
25	Referendum in Tauramena and Popular Court Decisions
26	Reviewing Roads Infrastructure in Casanare

27	Royalty System Projects and Departamental Issues
28	Minminas and EITI Efforts Addressing Challenges
29	Campetrol's Exploration Forecasts for 2016
30	ACIPET's Educational Initiatives and Layoff Information
31	Bioenergy and Ethanol Production Challenges
32	Hocol's Operations in Macarena, Ballena, and Chuchupa
33	Smuggling and Fuel Theft Issues
34	Seismic Exploration and Water Regulation
35	Canacol's Bet on Shale Exploration in 3Q13
36	Tierra Gran Llanos Energy Asset Analysis
37	Field Suspension Decisions and Inactivity
38	Cenit's Achievements and CNE Sale Status

Análisis de sentimiento: FinBERT y Vader

Date	Title	Finbert sentiment	Vader sentiment
2023-10-10	Significant surge in natural gas vehicle installations	positive	0,5106
2023-09-22	Moody's maintained its rating for Ecopetrol at Baa3, but with a negative outlook	negative	-0,7351
2023-11-29	ECP announces new appointments	neutral	0



Date	neg_vader	neu_vader	pos_vader	neg_finbert	neu_finbert	pos_finbert	avg_vader	avg_finbert	avg_vader_10periods	avg_finbert_10periods
2023-12-03	0	4	9	1	7	5	0.14	0.25	0.13	0.22
2023-12-10	3	2	6	3	3	5	0.04	0.31	0.12	0.18
2023-12-17	1	0	9	2	2	6	0.11	0.37	0.13	0.21
2023-12-24	2	3	1	1	4	1	0.04	-0.17	0.12	0.15
2023-12-31	0	0	1	0	0	1	0.00	1.00	0.11	0.25

Selección de variables: SelectKBest

- Se seleccionan las mejores variables de tópico y sentimiento para las 8 variables dependientes.
- Para retornos, se utiliza el **test ANOVA**.
- Para tendencias, se emplea el test **Chi-cuadrado**.
- Se eligen variables con una puntuación mayor a 1.

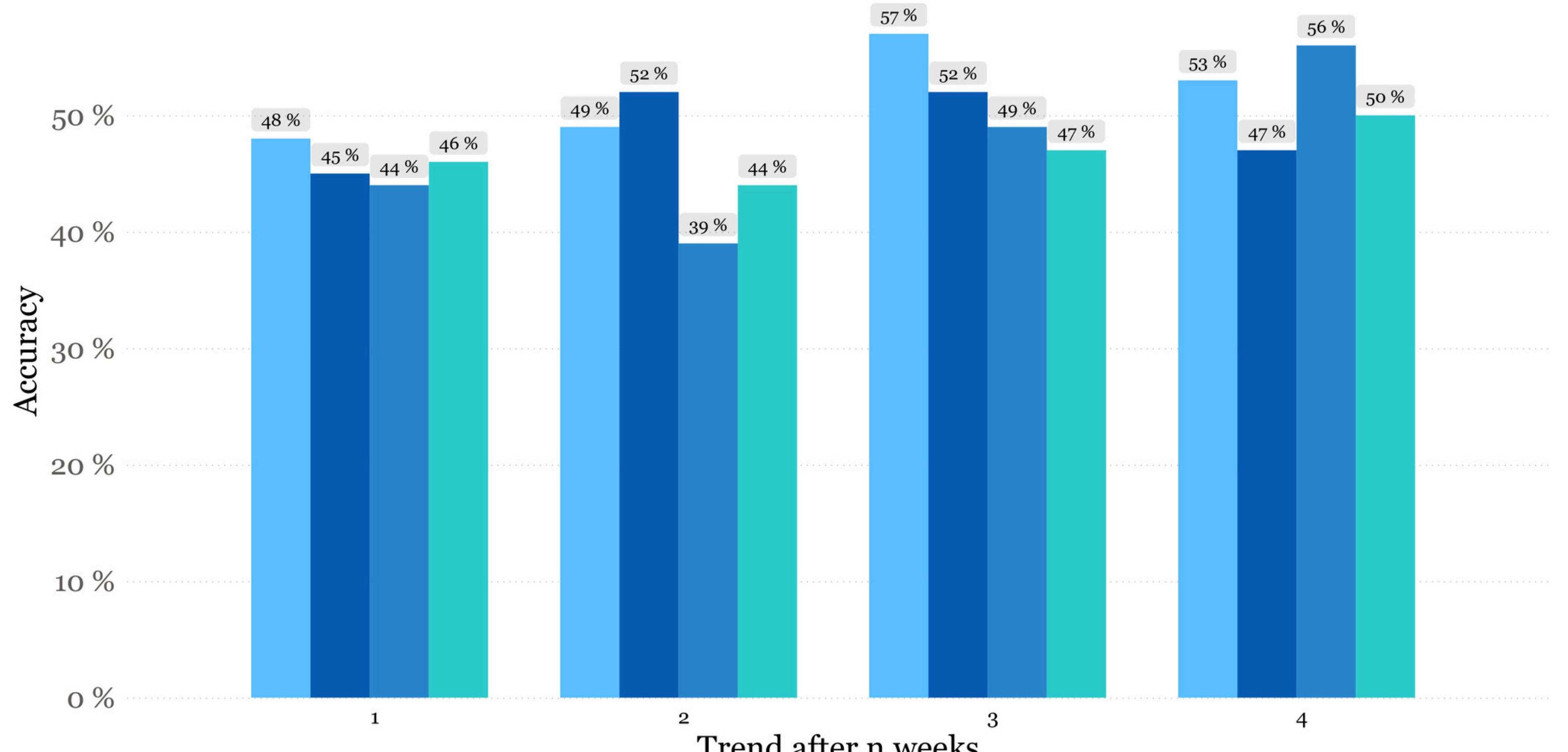
Table 3. Relevant features for each target

Targets	Features
trend1	Average Headline Sentiment Vader, Headline Sentiment Vader, T7, T24, T10, T33, T19, T31, T17
trend2	Average Headline Sentiment Vader, Headline Sentiment Vader, Negative Vader, Neutral Finbert, Neutral Vader, T5, T2, T28, T7, T24, T16, T23, T29, T9, T19, T13, T20, T17
trend3	Average Headline Sentiment Vader, Headline Sentiment Vader, Neutral Finbert, Neutral Vader, T2, T16, T4, T29, T32, T18, T22
trend4	Average Headline Sentiment Finbert, Average Headline Sentiment Vader, Neutral Finbert, Neutral Vader, T14, T24, T16, T11, T23, T26, T29, T19, T37, T18
return1	Headline Sentiment Vader, T5, T7, T21, T14, T24, T12, T4, T23, T29, T33, T19, T20, T30, T32, T17
return2	Headline Sentiment Vader, Negative Finbert, Negative Vader, Neutral Vader, T2, T8, T14, T24, T16, T12, T23, T29, T20, T30, T18, T17
return3	Average Headline Sentiment Vader, Negative Finbert, Negative Vader, Neutral Vader, T2, T7, T14, T24, T16, T12, T11, T23, T29, T38, T33, T19, T20, T17, T22
return4	Average Headline Sentiment Vader, Headline Sentiment Vader, Negative Finbert, Negative Vader, Neutral Vader, T1, T2, T36, T7, T14, T24, T16, T12, T4, T23, T26, T29, T33, T19, T37, T17, T22

Resultados: Precisión

Accuracy by model after n weeks

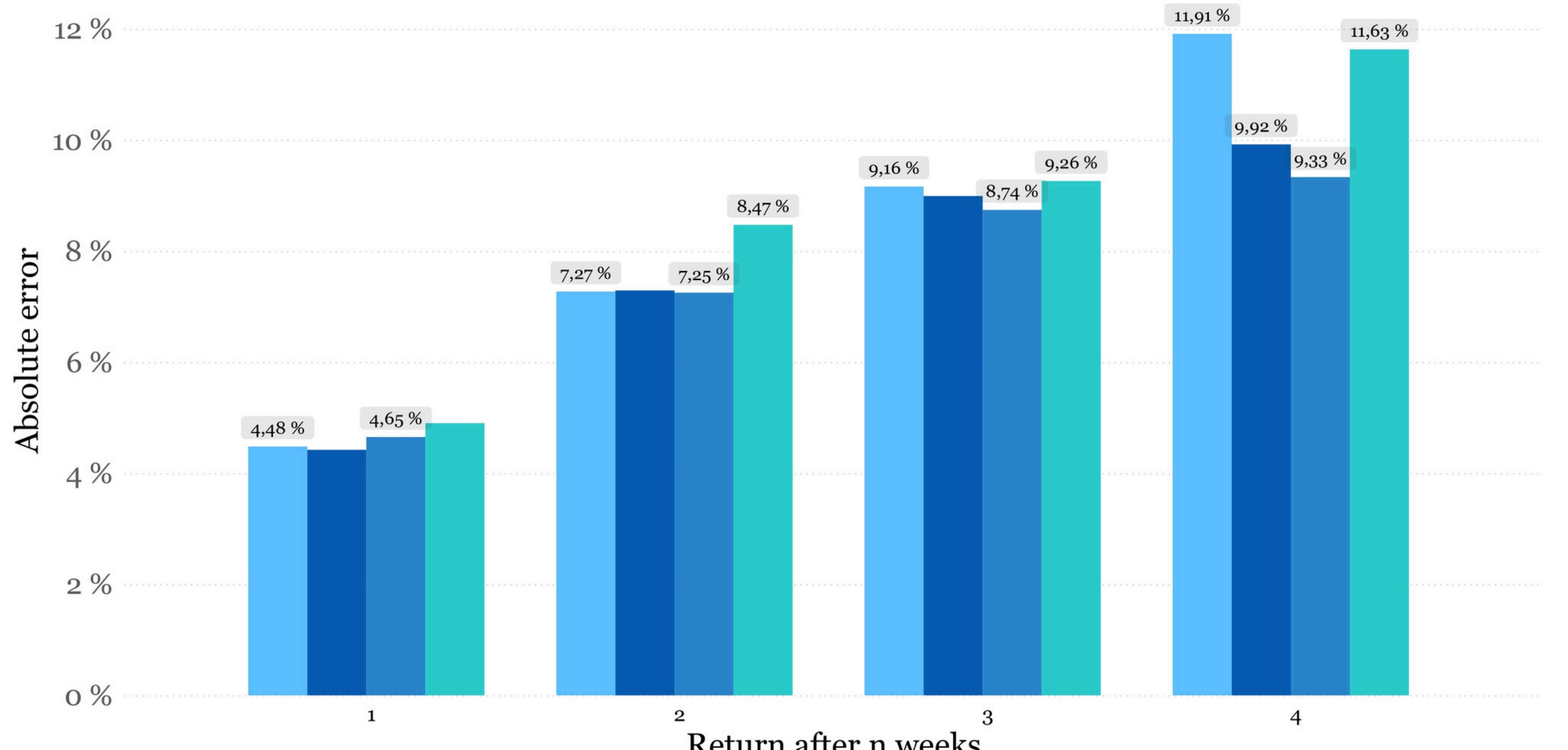
Models ● GradientBoostingClassifier ● KNeighborsClassifier ● RandomForestClassifier ● XGBClassifier



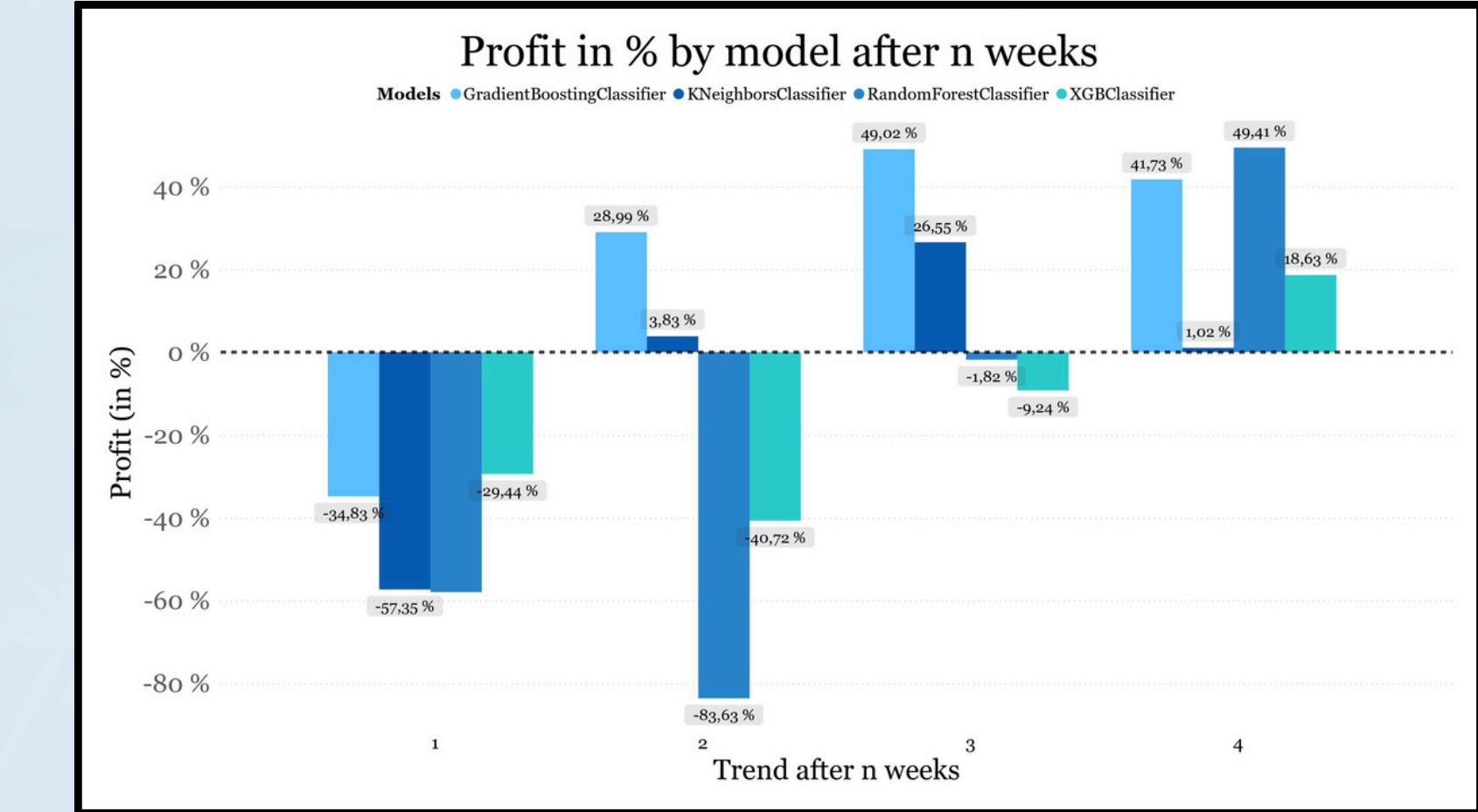
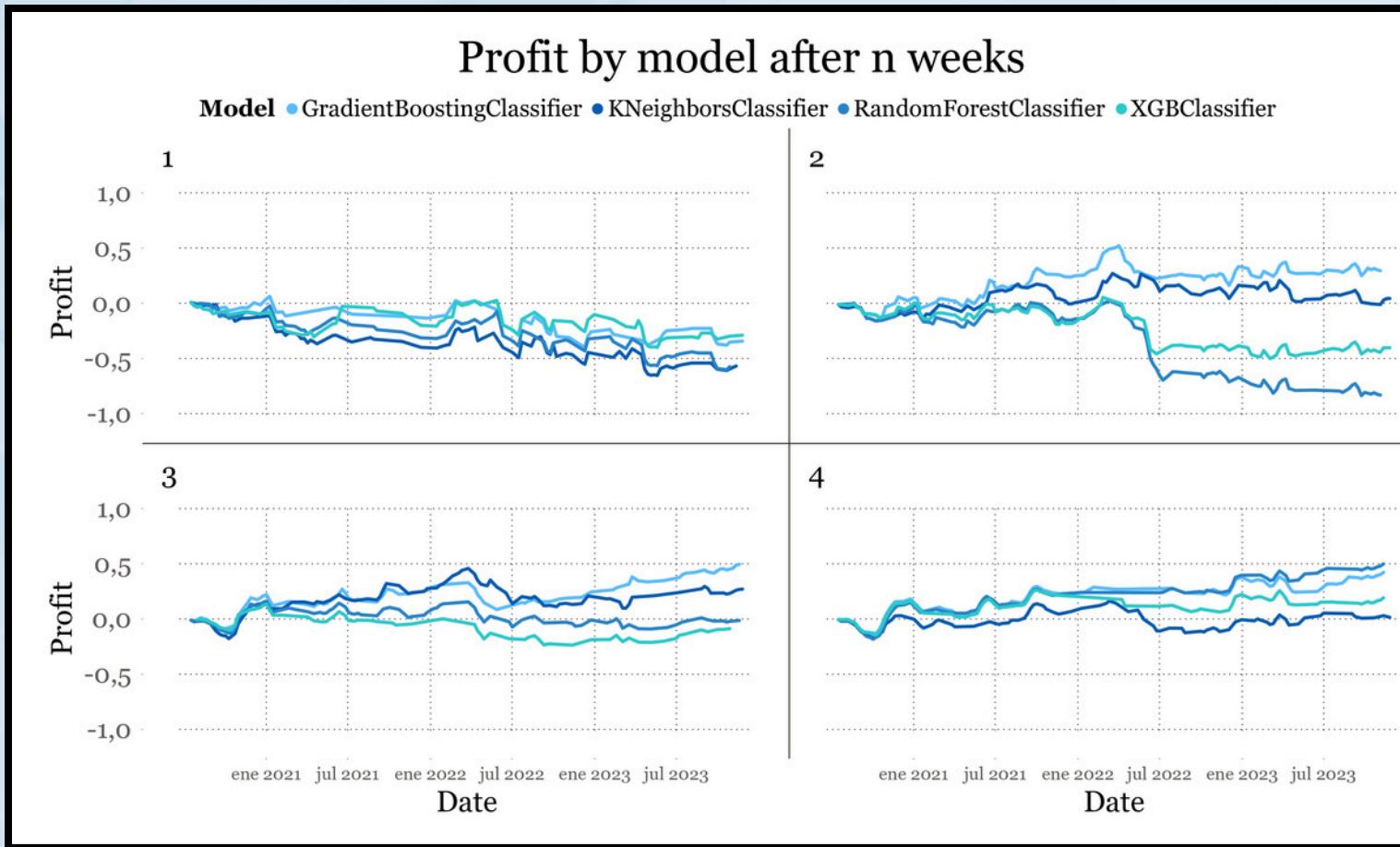
Resultados: Error absoluto

Absolute error by model after n weeks

Models ● GradientBoostingRegressor ● KNeighborsRegressor ● RandomForestRegressor ● XGBRegressor



Resultados: Rentabilidad



Resumen y conclusiones

- El análisis de sentimiento y modelado de tópicos permite generar **variables clave** como insumo para los modelos de aprendizaje automático en base a **diferentes marcos temporales**
- Los titulares de noticias son más efectivos para predecir **movimientos en periodos de 3 y 4 semanas**, con el modelo GradientBoostingClassifier logrando **49.02% de rentabilidad y 56% de precisión** en tendencias para la semana 3, superando la estrategia de **buy and hold de un 6.04%**
- Se recomienda replicar este enfoque en **otros activos financieros**, incluyendo el **cuerpo de los artículos** así como analizar más a fondo el **impacto de los tópicos en los precios de los activos.**

Predicción de acciones usando modelado de tópicos y técnicas de análisis de sentimiento: Un caso de estudio de Machine Learning en Ecopetrol

Edwar Valenzuela Cortés
evalenzuelac@unal.edu.co



Santiago Puentes Núñez
spuentesn@unal.edu.co



¡Gracias!



UNIVERSIDAD
NACIONAL
DE COLOMBIA

