

Digital House - Data Science - Proyecto Final

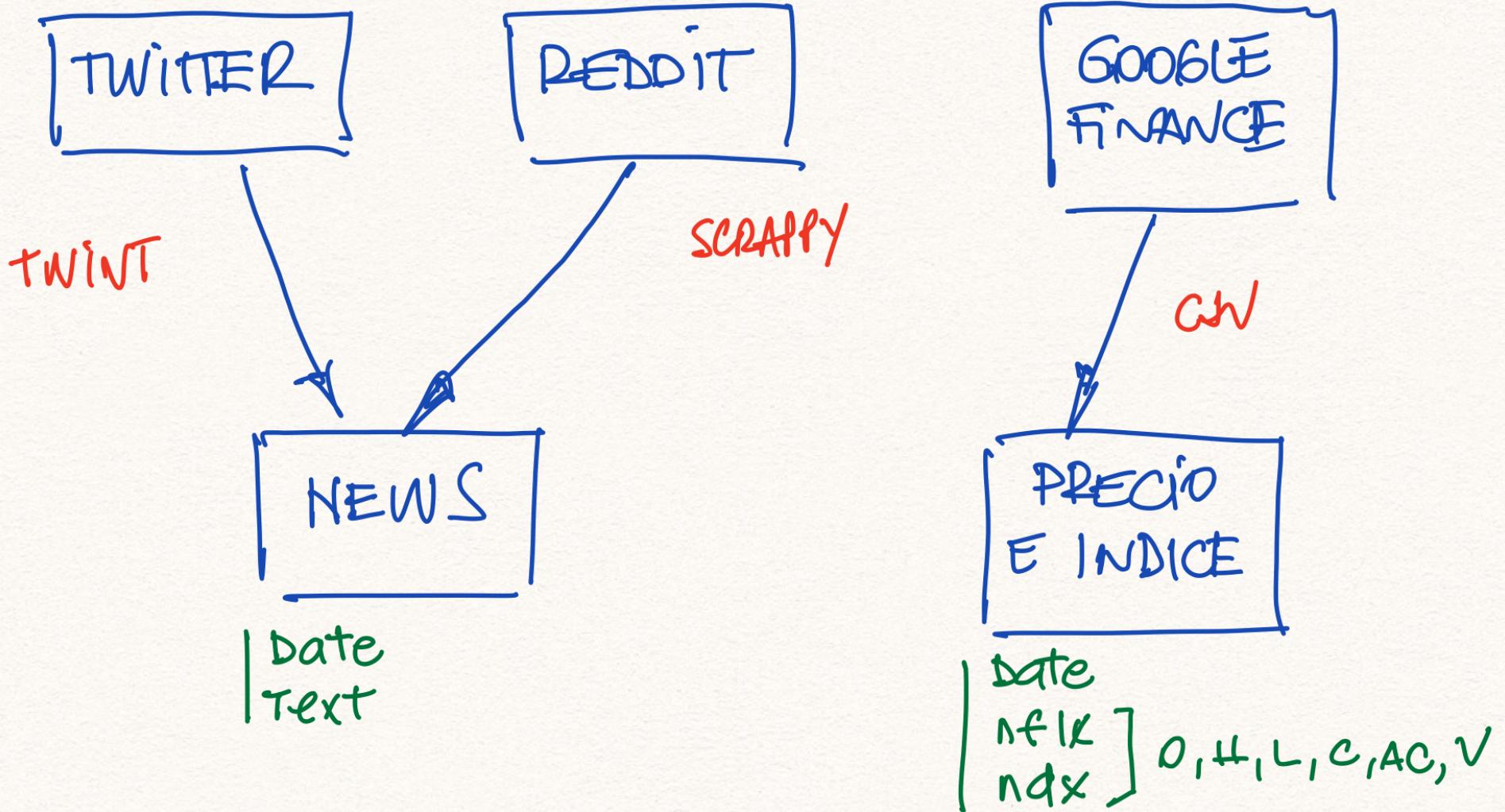
Objetivo: Predecir si el retorno excedente de la acción de Netflix (NFLX) respecto al índice NASDAQ 100 (NDX) es positivo o negativo utilizando tweets y noticias de Reddit

El trabajo se inspira en el paper:

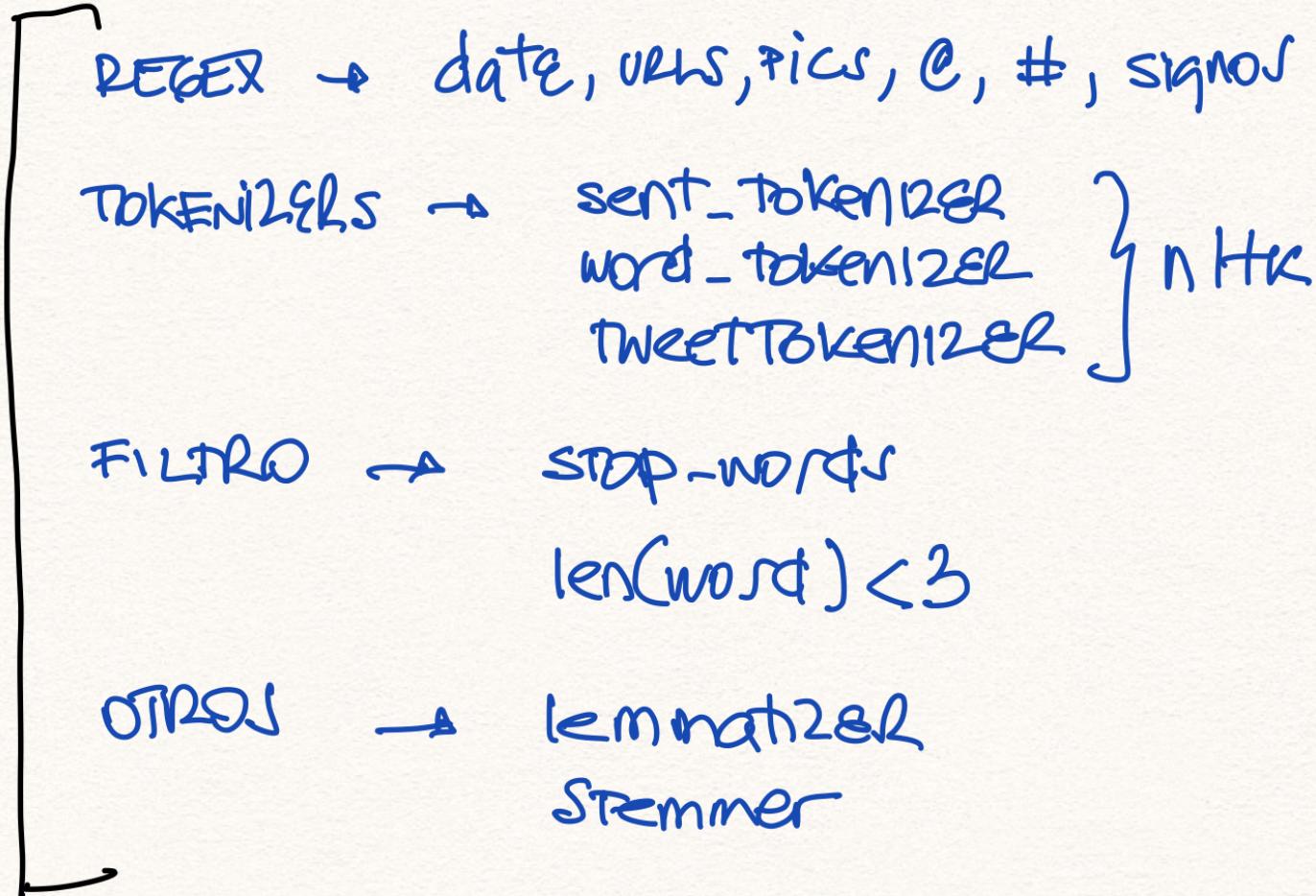
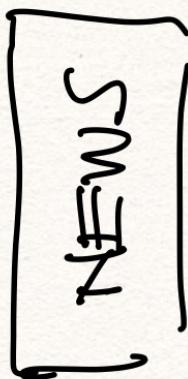
['Sentiment analysis of Twitter data for predicting stock market movements'](#)

de Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda y Babita Majhi

CAPTURA DE DATOS



PREPROCESAMIENTO



PRECIO



CÁLCULO DEL RETORNO DIARIO

NFLX
NDX

CÁLCULO DEL RETORNO EXCEDENTE ($\text{NFLX} - \text{NDX}$)

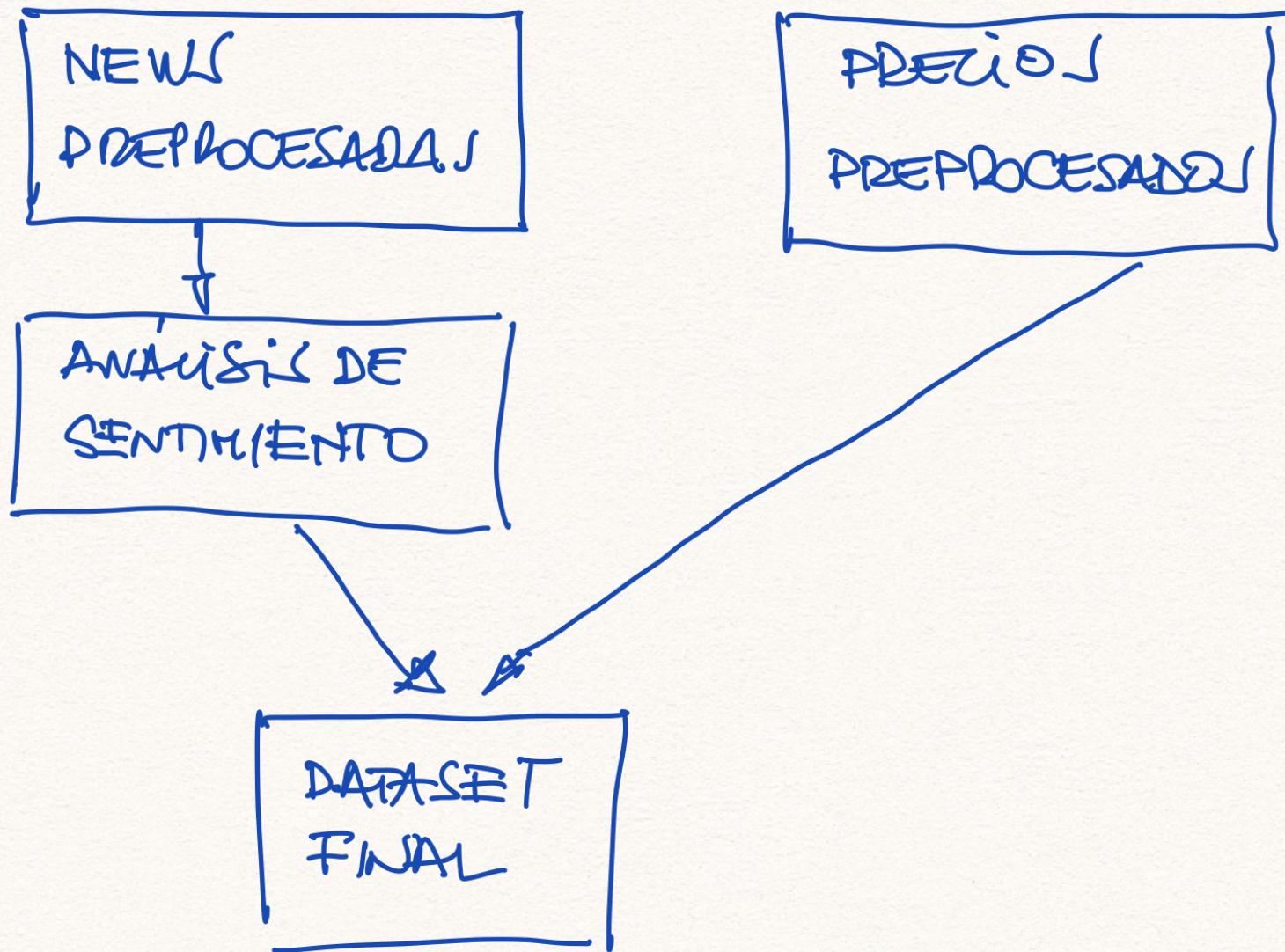
TRANSFORMACIÓN EN VARIABLE BINARIA

↳ MANUAL (T-F) → LABEL binARIER (L~O)

ANÁLISIS DE SENTIMENTO

- TEXTBLOB (wrapper de NLTK)
- VADER Sentiment Intensity Analyzer (NLTK)
 - ↳ POLARIDAD (SCORE)
 - ↳ POSITIVA
 - ↳ NEGATIVA
 - ↳ NEUTRA
- RESULTADOS AGRUPADOS POR DÍA
 - promedio de scores
 - shift de retornos

DATASET FINAL



date	SCOPE			RET	RET%
	POS	NEG	NEU		
.					

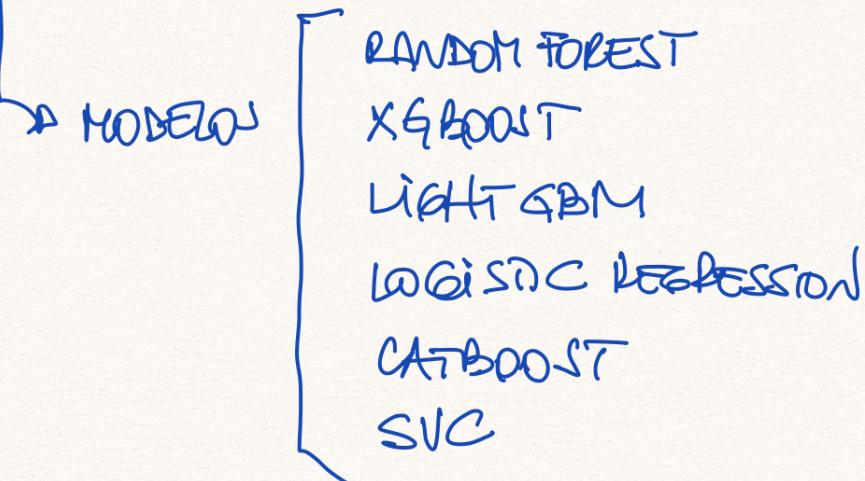
CLASIFICADORES

- SPLIT TRAIN | TEST

TIME SERIES SPLIT

[TRAIN N 320 d
TEST N 15 d]

- RANDOMIZED SEARCH
- HYPERPARAMETROS
MANUALES
 - # DIAZ (1-5)
 - SCORE NEUTRAL (±1) NO



- METRIC: ROC_AUC SCORE

RESULTADOS

- MUY MALOS RESULTADOS
 - TRAIN : 55 - 59%
TEST : 54 - 79% → + INESTABLE
 - MEJOR MODELO:
RF W=3 N=TWEET
TWEETTOKENIZER
STOP WORDS
- TRAIN: 58.7%
TEST: 79.2%

- MEJORAS:

Ⓐ AMPLIAR DATOS HISTÓRICOS

↳ ubicar + ciclos de mercado

Ⓐ DESARROLLAR UN CLASIFICADOR
DE POLARIDAD PROPIO

Ⓐ AGREGAR FEATURES

↳ EVOLUCIÓN INTEGRACIÓN FINANCIERA