

# Stock market meets external events: can we predict the reaction?

Elizaveta Kovtun  
Sber, Skoltech

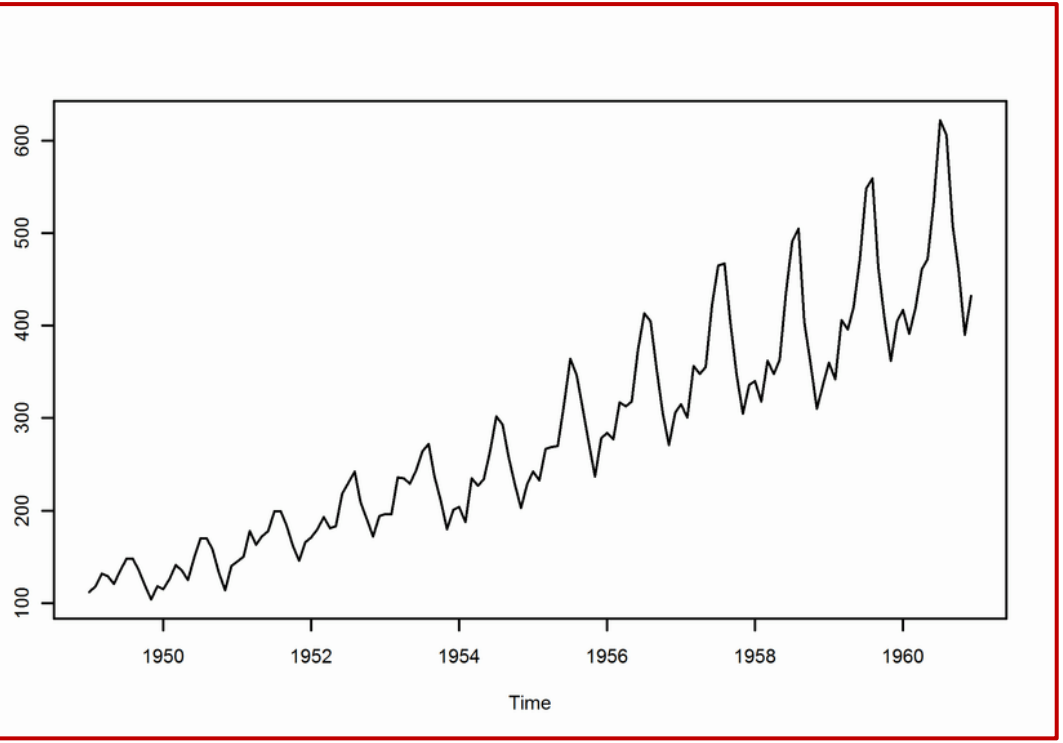
# Stock price



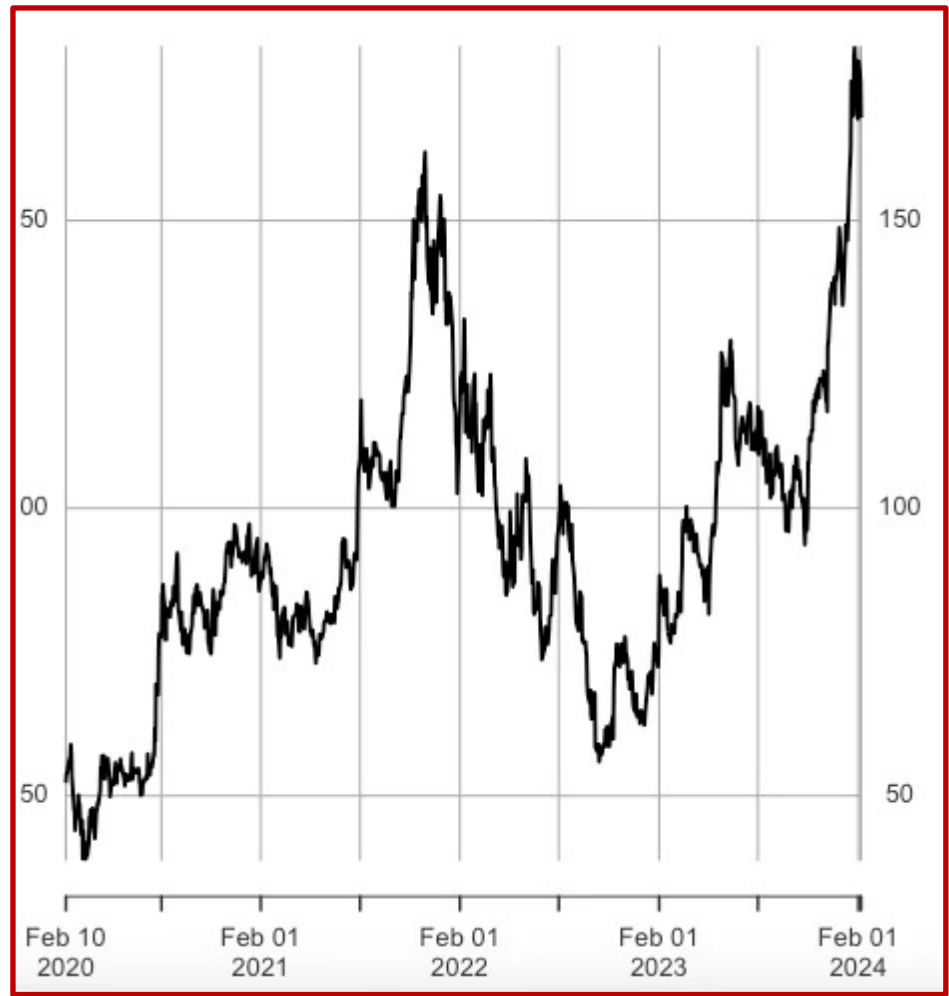
# Stock price



# Time series of different nature



**VS**



# Stock price and external events



# Problem statement



Our goal is to predict price change after influential events

# Published paper

**Scientific Reports (Nature)**

**New drugs and stock market: a machine learning framework for predicting pharma market reaction to clinical trial announcements**

Semen Budenny, Alexey Kazakov, Elizaveta Kovtun, and Leonid Zhukov

# Published paper

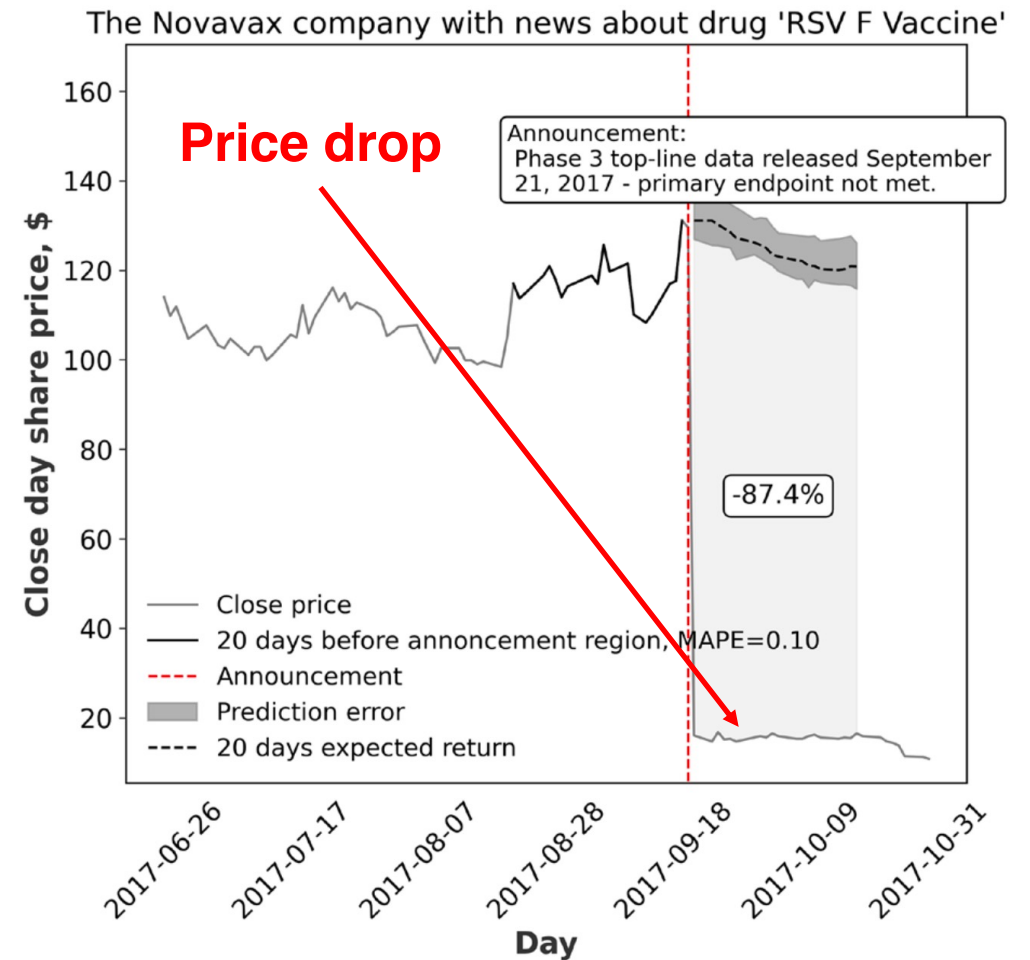
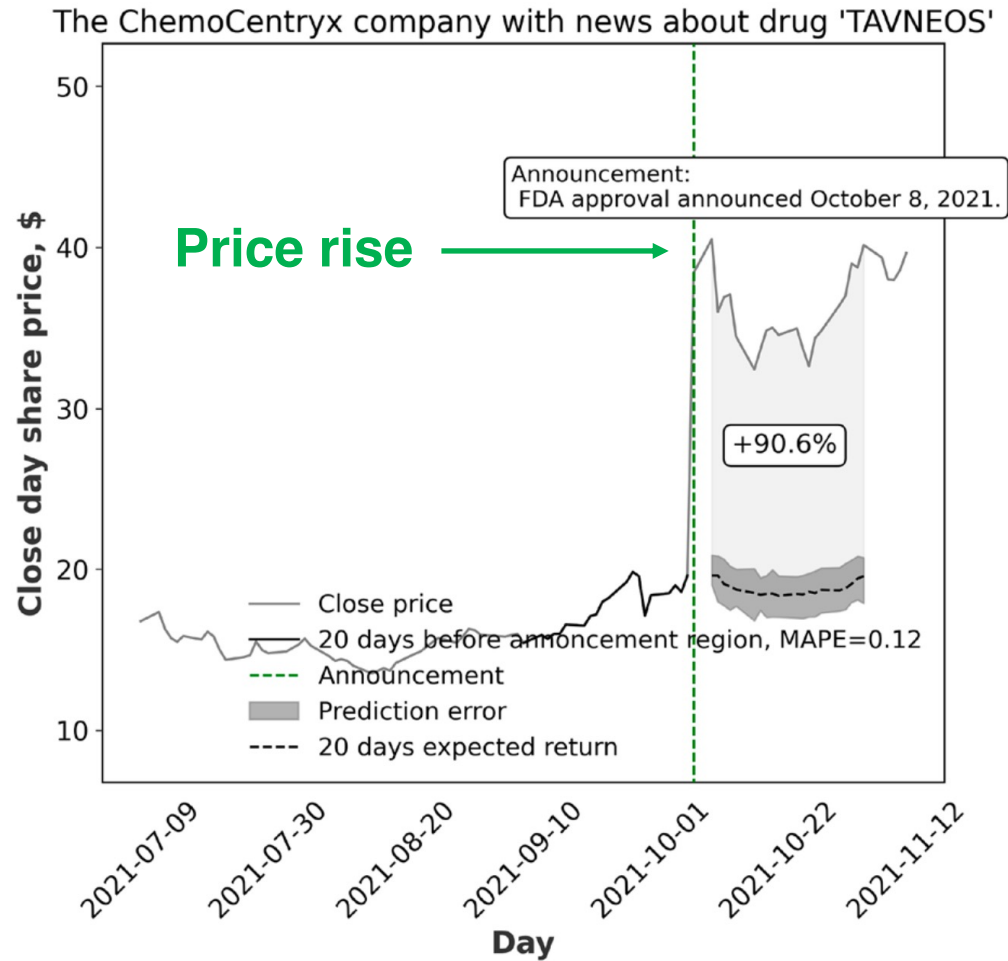
**New drugs and stock market: a machine learning framework for predicting pharma market reaction to clinical trial announcements**

Time Series

↑  
Events



# Why pharma market and clinical trial announcements?



# 1. Sentiment polarity extraction from clinical announcements

**Positive  
Announcements**

Keywords: “approve”,  
“meets”, “show”

**Negative  
Announcements**

Keywords: “failed”,  
“halted”, “did not reach”

**Neutral  
Announcements**

1. Compose dictionaries with keywords that reflect the announcement polarity
2. Train BERT on announcement texts and a rule-based markup
3. Complement dictionaries with keyword from mistakenly classified texts

# 1. Sentiment polarity extraction from clinical announcements

## Positive Announcements

Keywords: “approve”,  
“meets”, “show”

+

“demonstrate”,  
“potential”, “accepted”,  
“encouraging”

## Negative Announcements

Keywords: “failed”,  
“halted”, “did not reach”

+

“terminated”,  
“discontinued”,  
“insufficient”, “paused”

## Neutral Announcements

# 1. Sentiment polarity extraction from clinical announcements

**Positive  
Announcements**

**Negative  
Announcements**

**Neutral  
Announcements**

**Rule-based markup is reasonable since a message of announcement is straightforward**

## 2. Construction of feature space

### Market features

- NASDAQ biotechnology index
- Mean number of trading volume peaks per year
- Stock price trend for the last 30 days before the event

### Company features

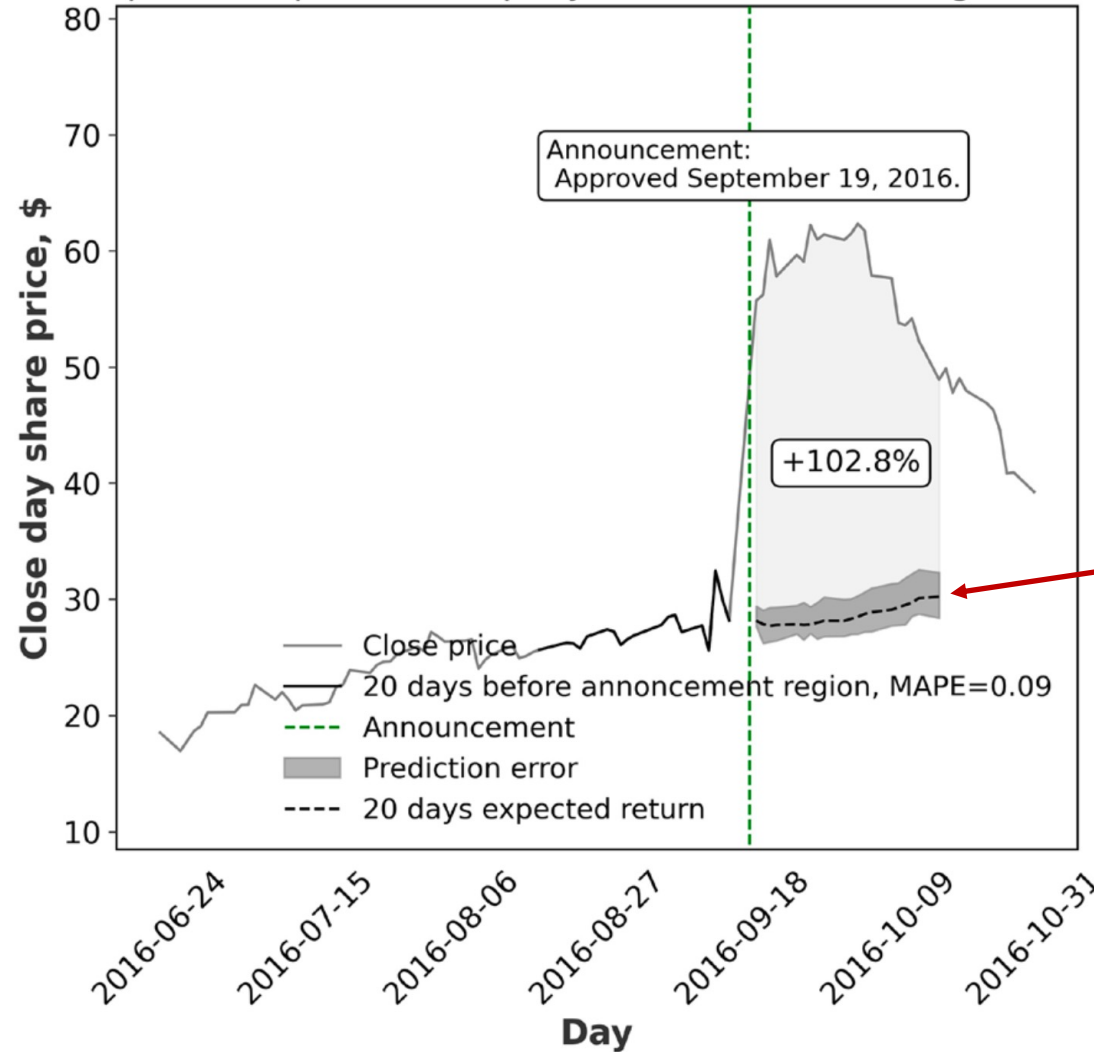
- Income statement
- Full-Time Employees
- Cash flow

### Announcement features

- Announcement sentiment polarity
- ICD-10 codes

# 3. Evaluation of expected return

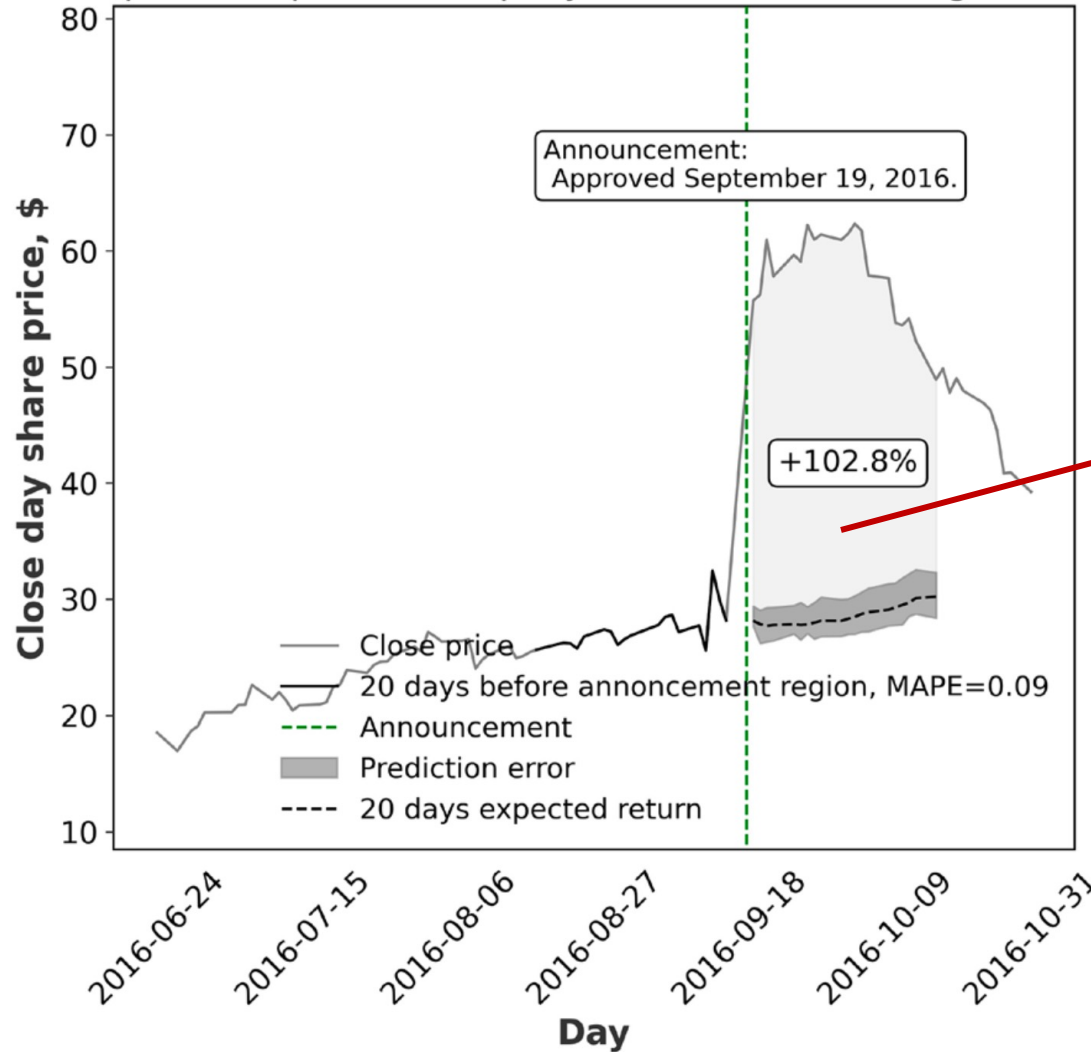
The Sarepta Therapeutics company with news about drug 'EXONDYS 51



**Expected return**

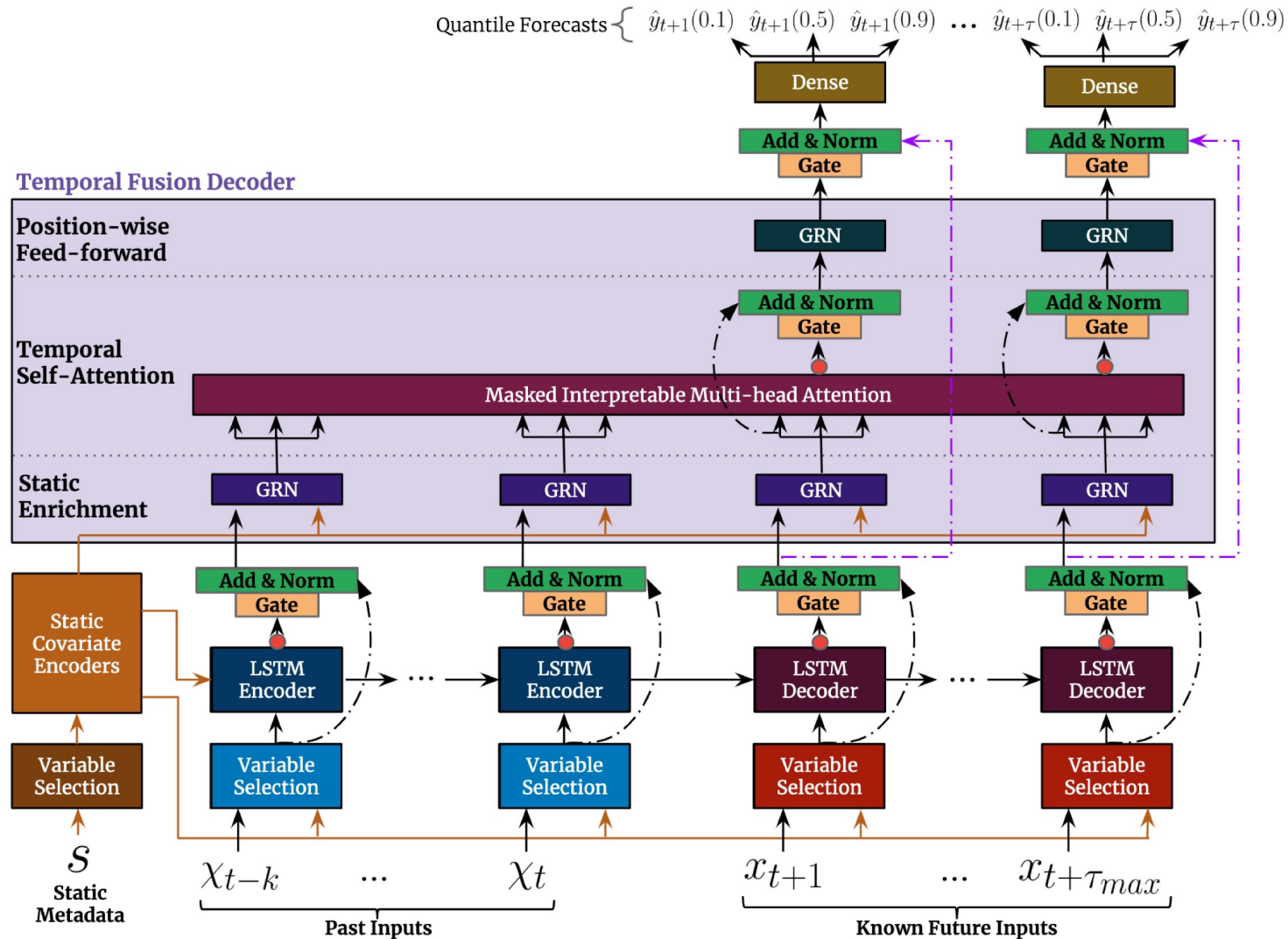
### 3. Evaluation of expected return

The Sarepta Therapeutics company with news about drug 'EXONDYS 51



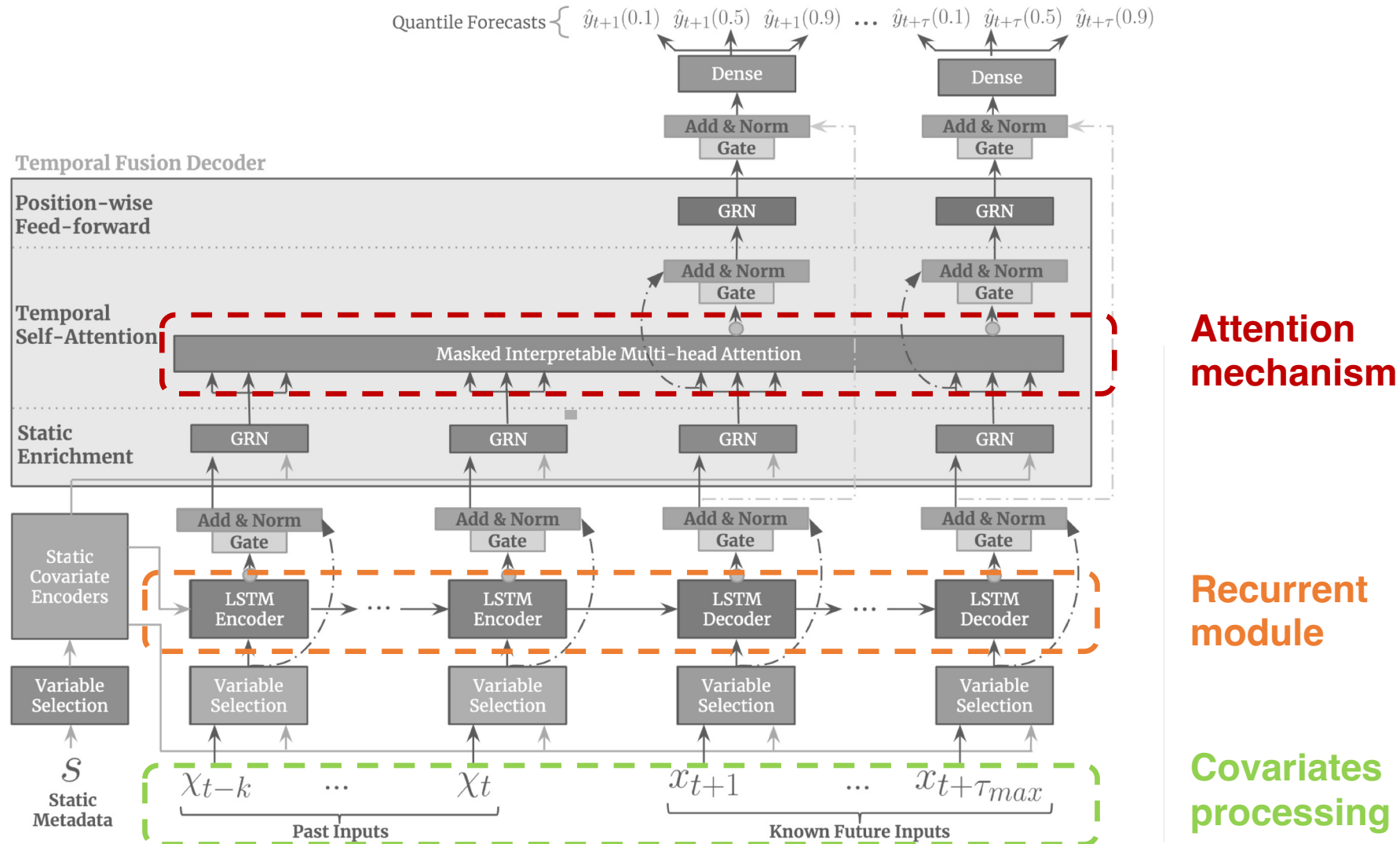
**Target measure:  
NCAR\_20**

# 3. Evaluation of expected return





# 3. Evaluation of expected return



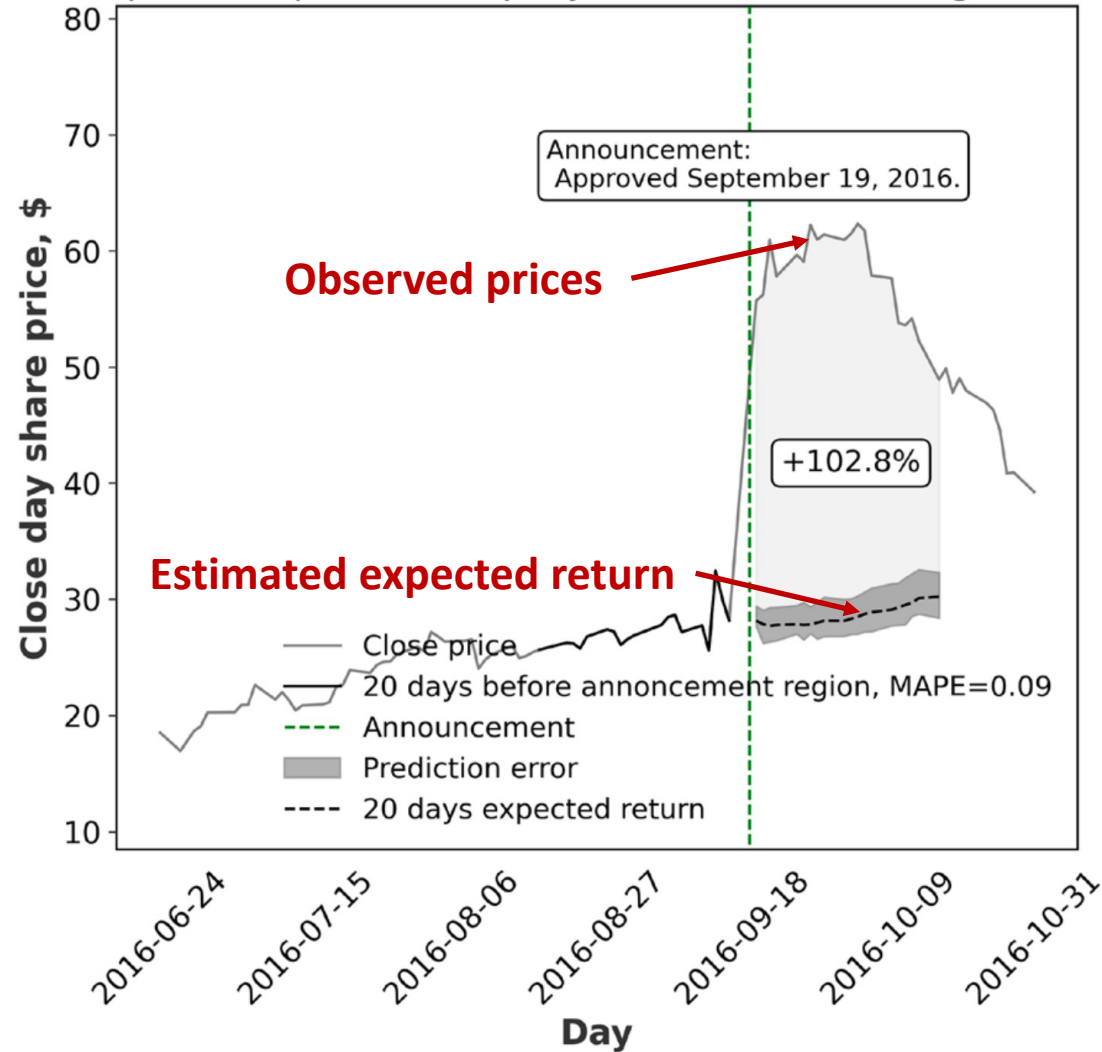
### 3. Evaluation of expected return

#### Estimation of expected return:



### 3. Evaluation of expected return

The Sarepta Therapeutics company with news about drug 'EXONDYS 51



Expected return estimation allows calculating a target value, NCAR\_20

# From regression to classification

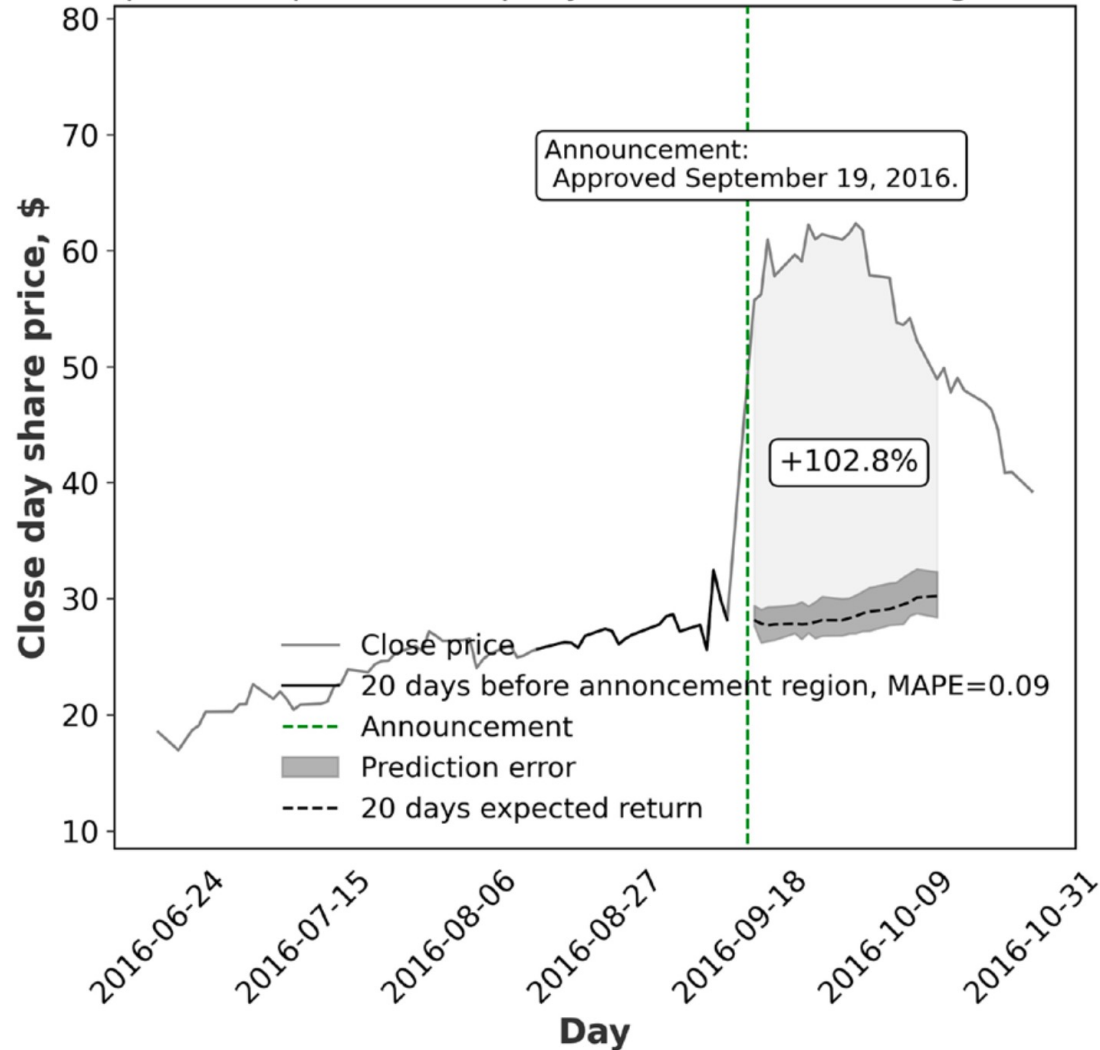
**Regression setting:**  
Prediction of NCAR\_20



**Multi-class  
classification setting:**  
Prediction of NCAR\_20  
change **range**

# General pipeline

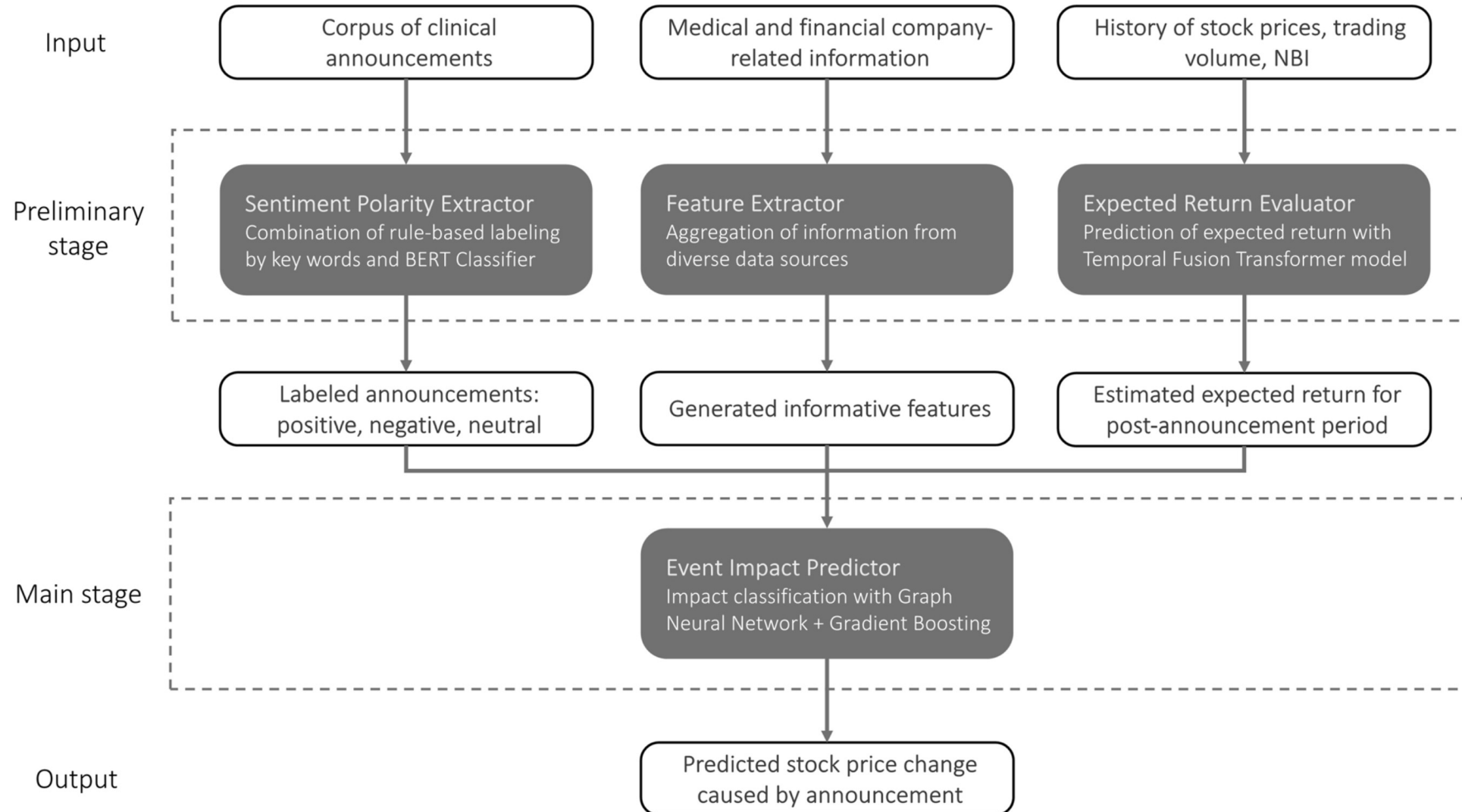
The Sarepta Therapeutics company with news about drug 'EXONDYS 51



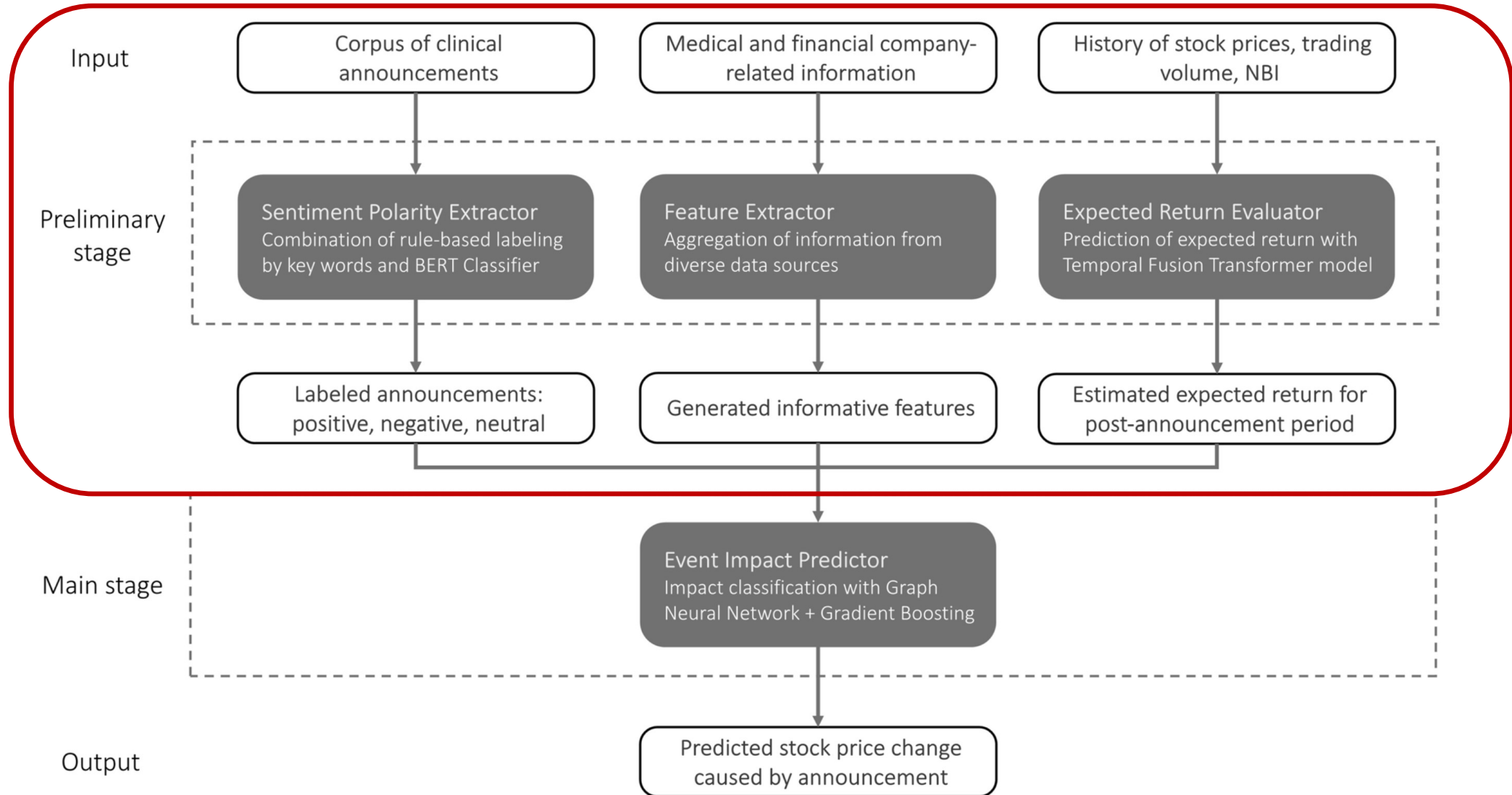
Input: time series + event

Output: prediction of event influence

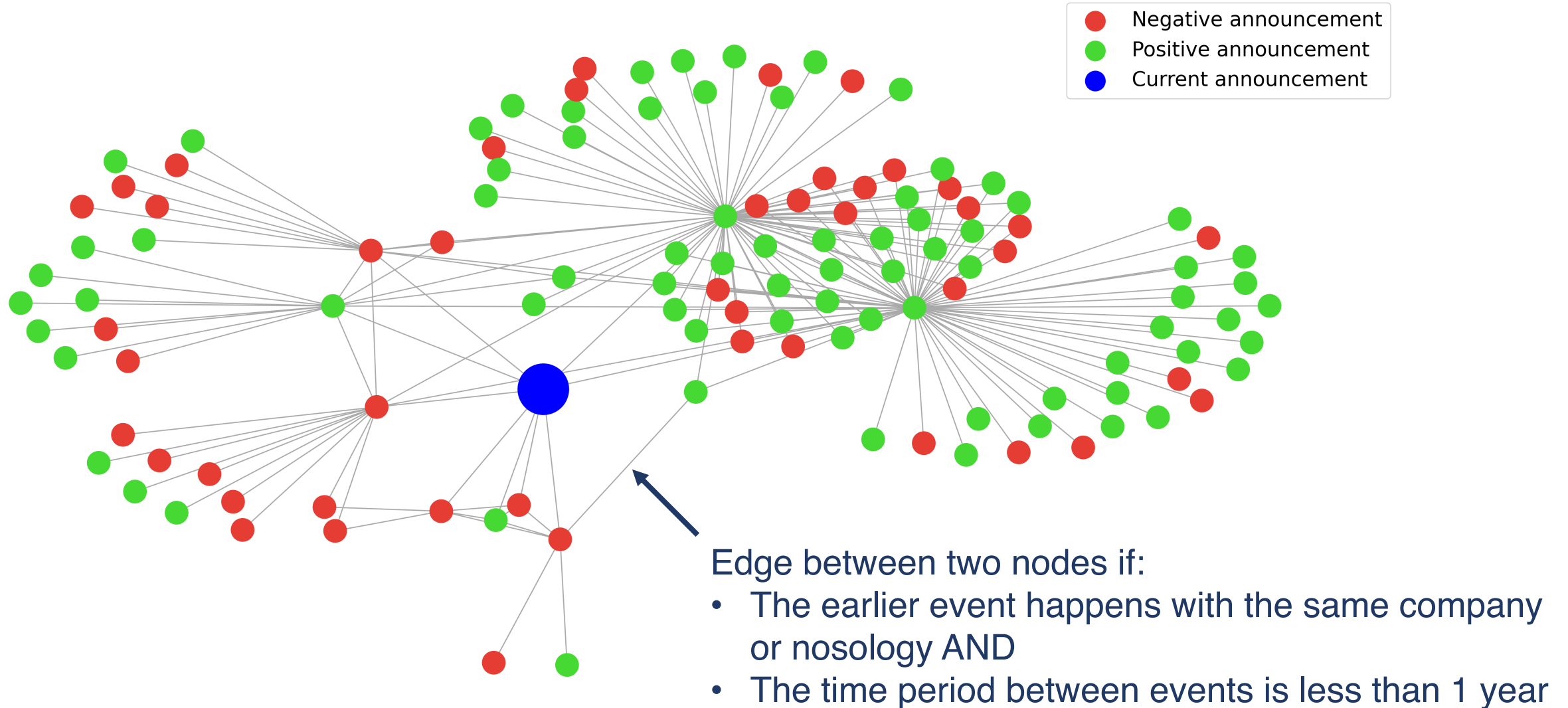
# General pipeline



# General pipeline



# Adoption of GCN





# Adoption of GCN

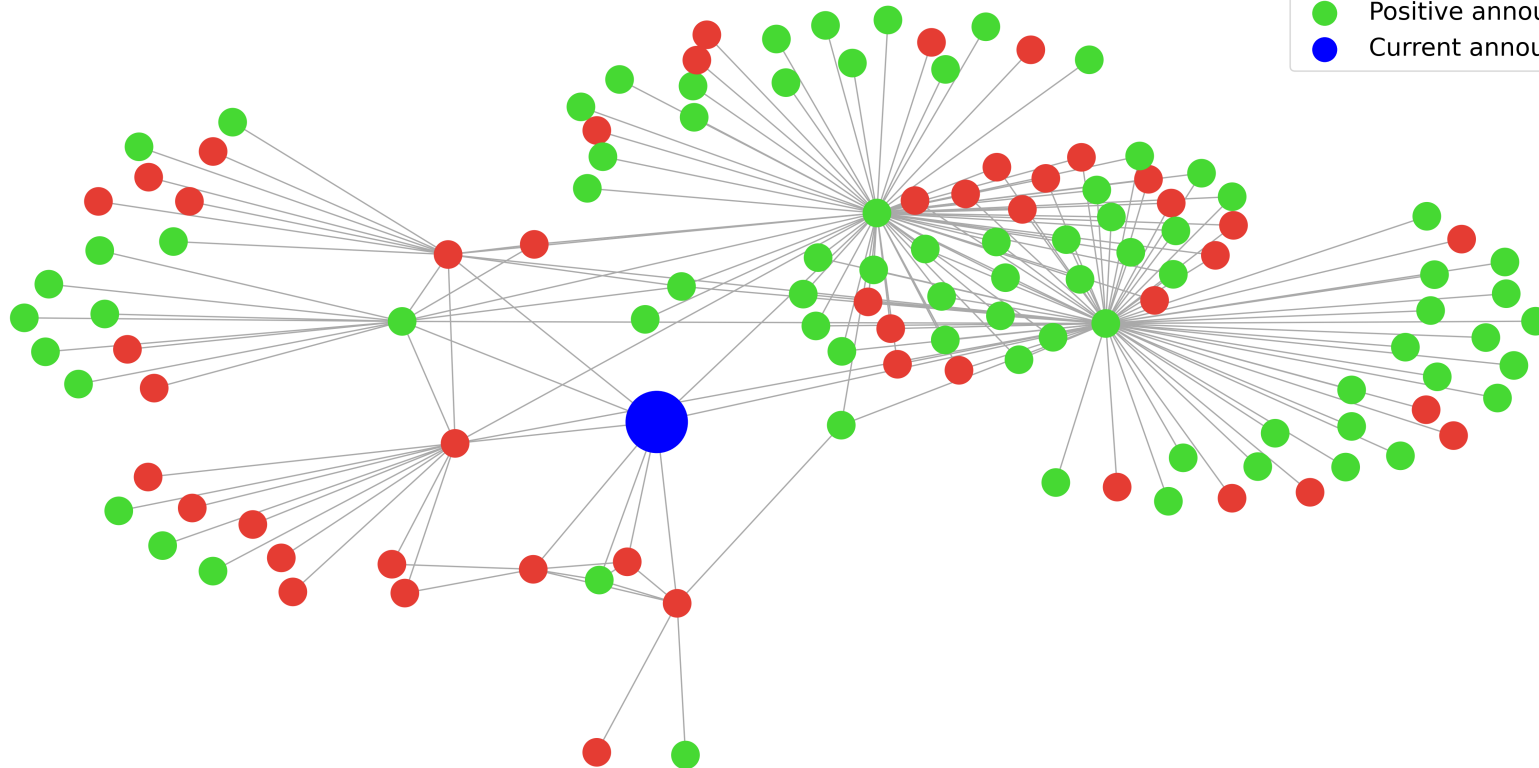
- **Announcement information**
- Graph



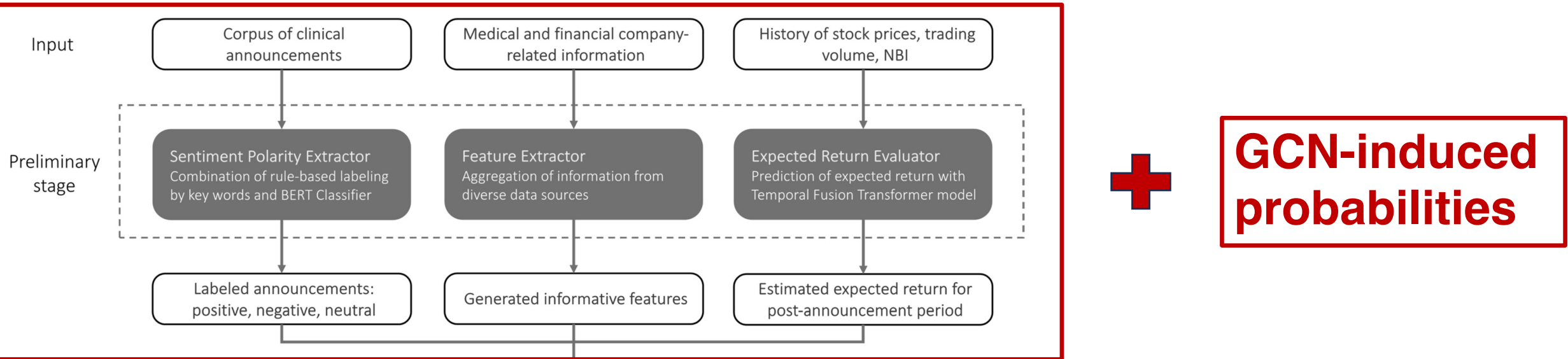
**GCN**



Class probabilities of price change range

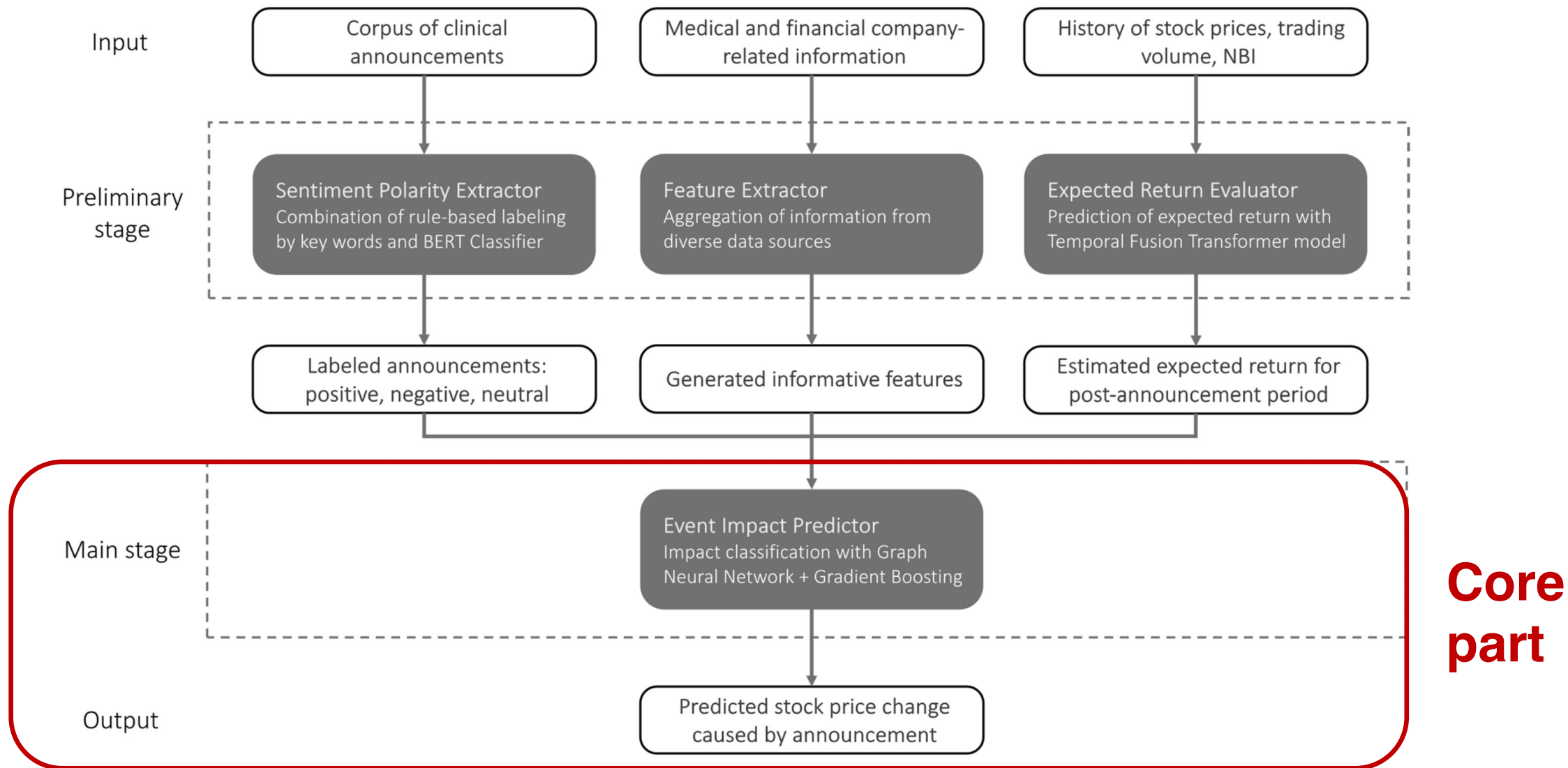


# Core classifier



**Input to Gradient Boosting**

# Core classifier





# Results

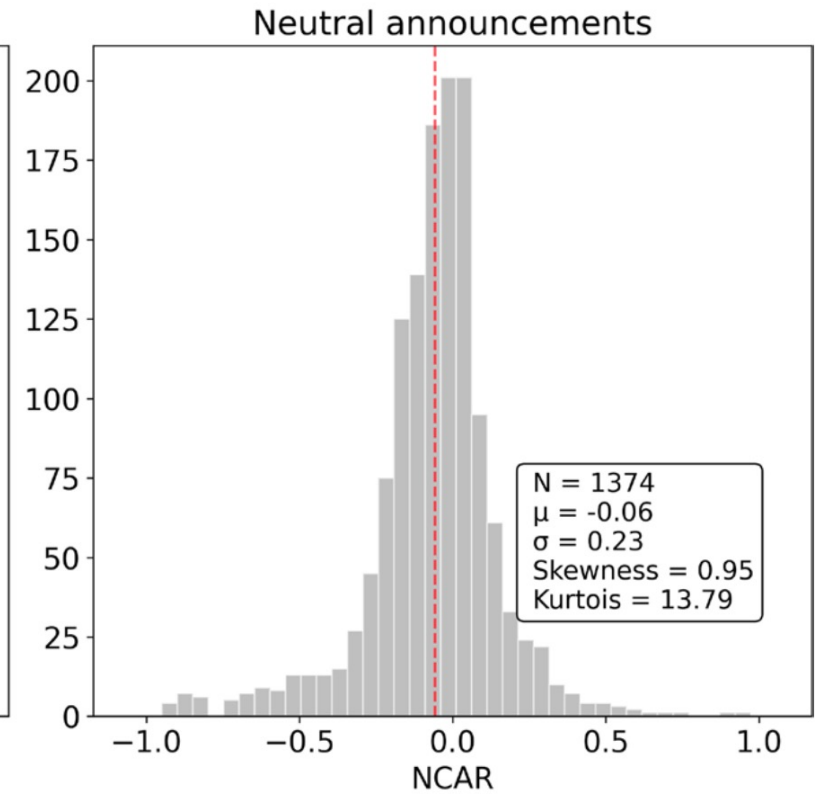
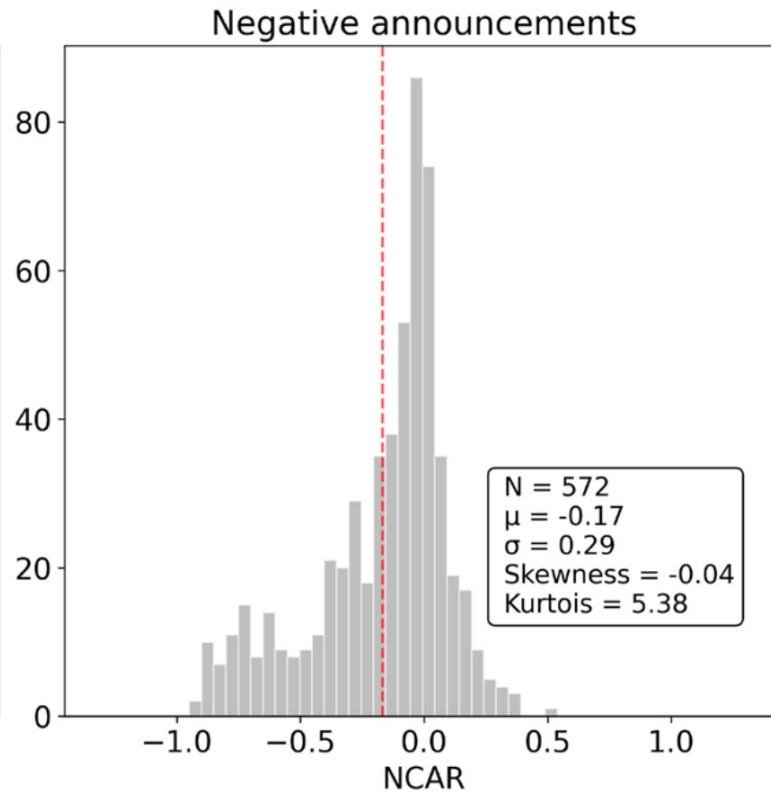
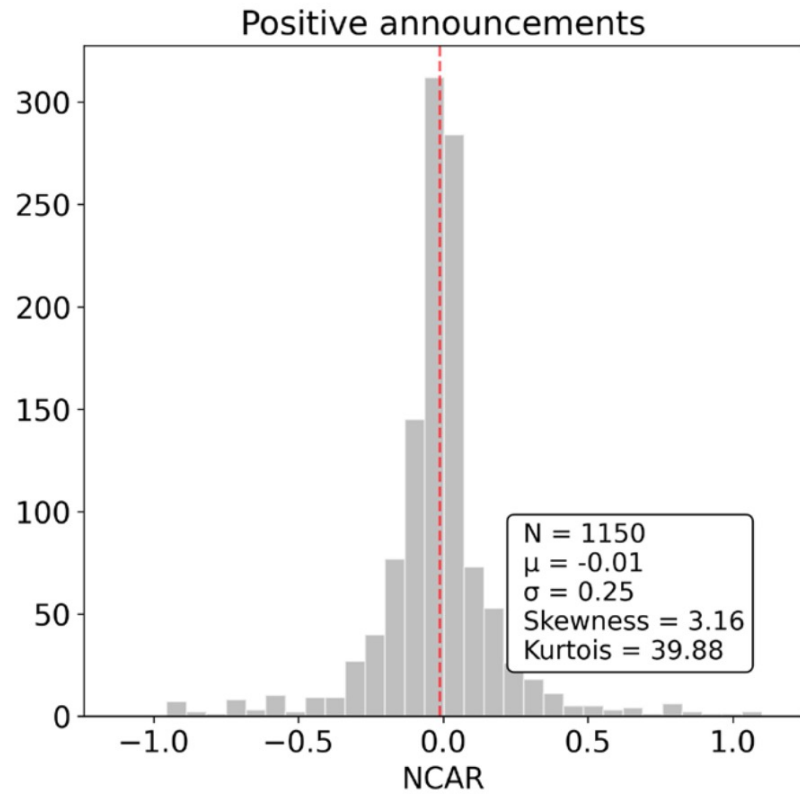
# Data

- 5436 clinical trial announcements
- 681 companies
- years 2018-2022

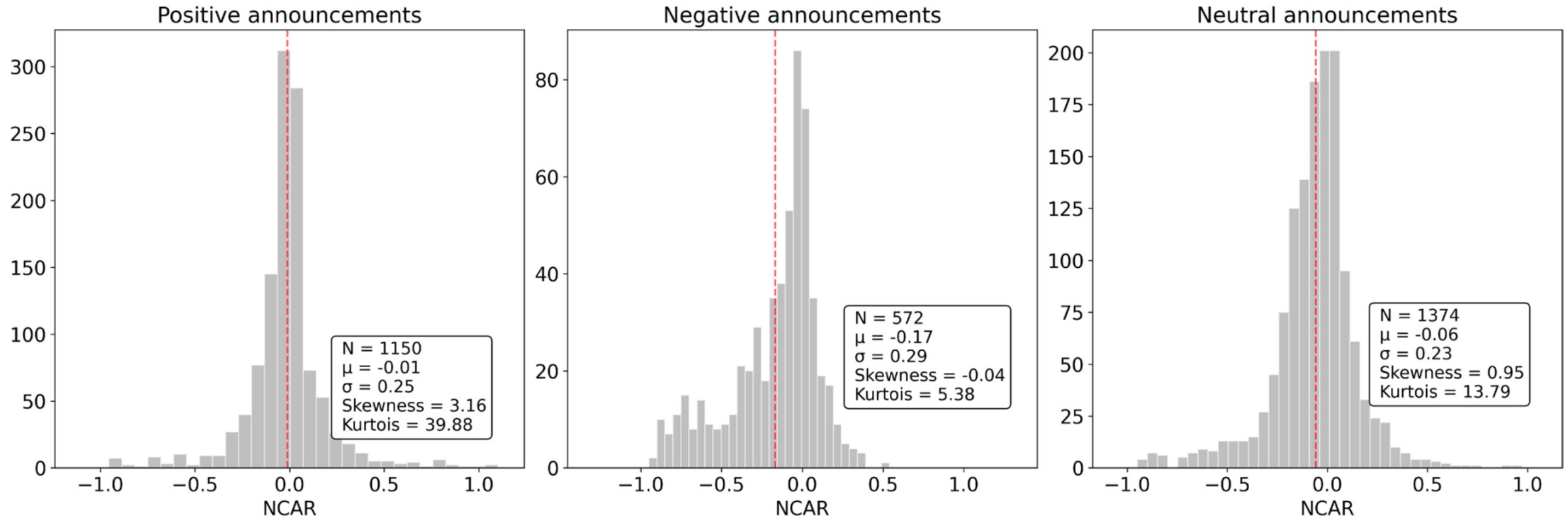
# Sentiment polarity evaluation

	# Divergences	# Coinciding positives	# Coinciding negatives	# Coinciding neutrals
With the initial keywords	207	1447	445	337
With the updated keywords	66 	1562	765	304 

# Announcement impact analysis



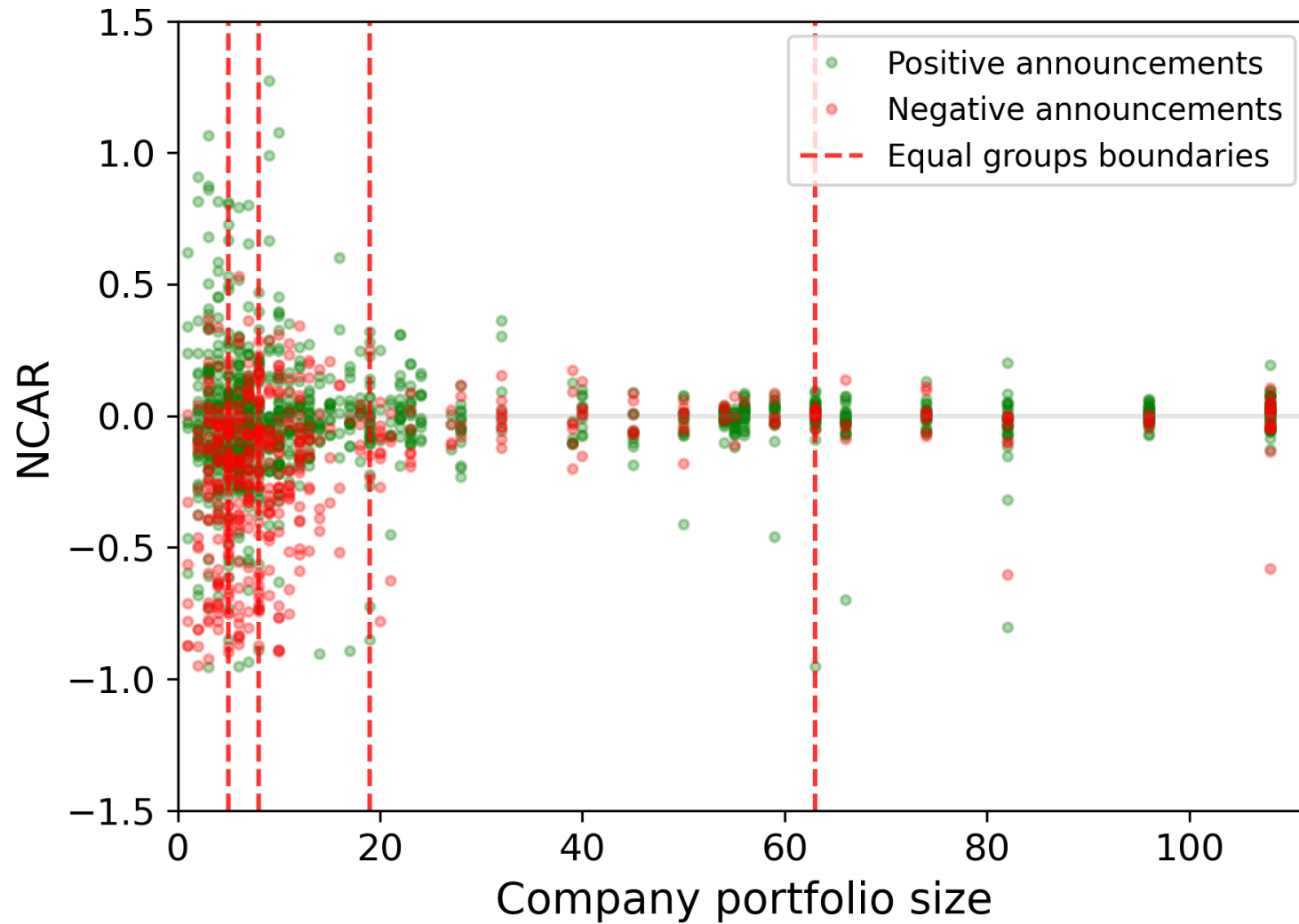
# Announcement impact analysis



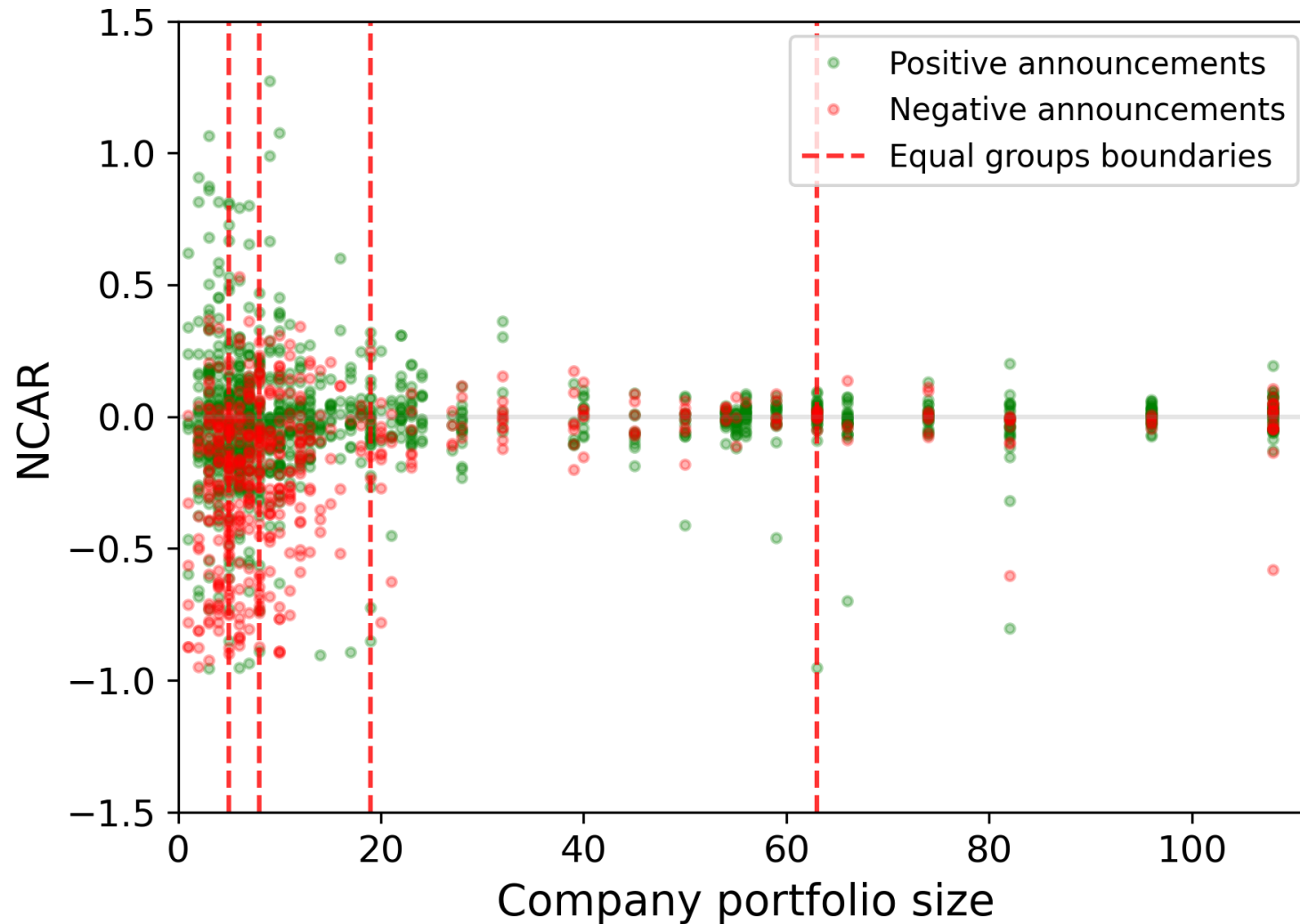
**Mann-Whitney U test: negative and non-announcement distributions are different**



# Impact of company background on stock prices

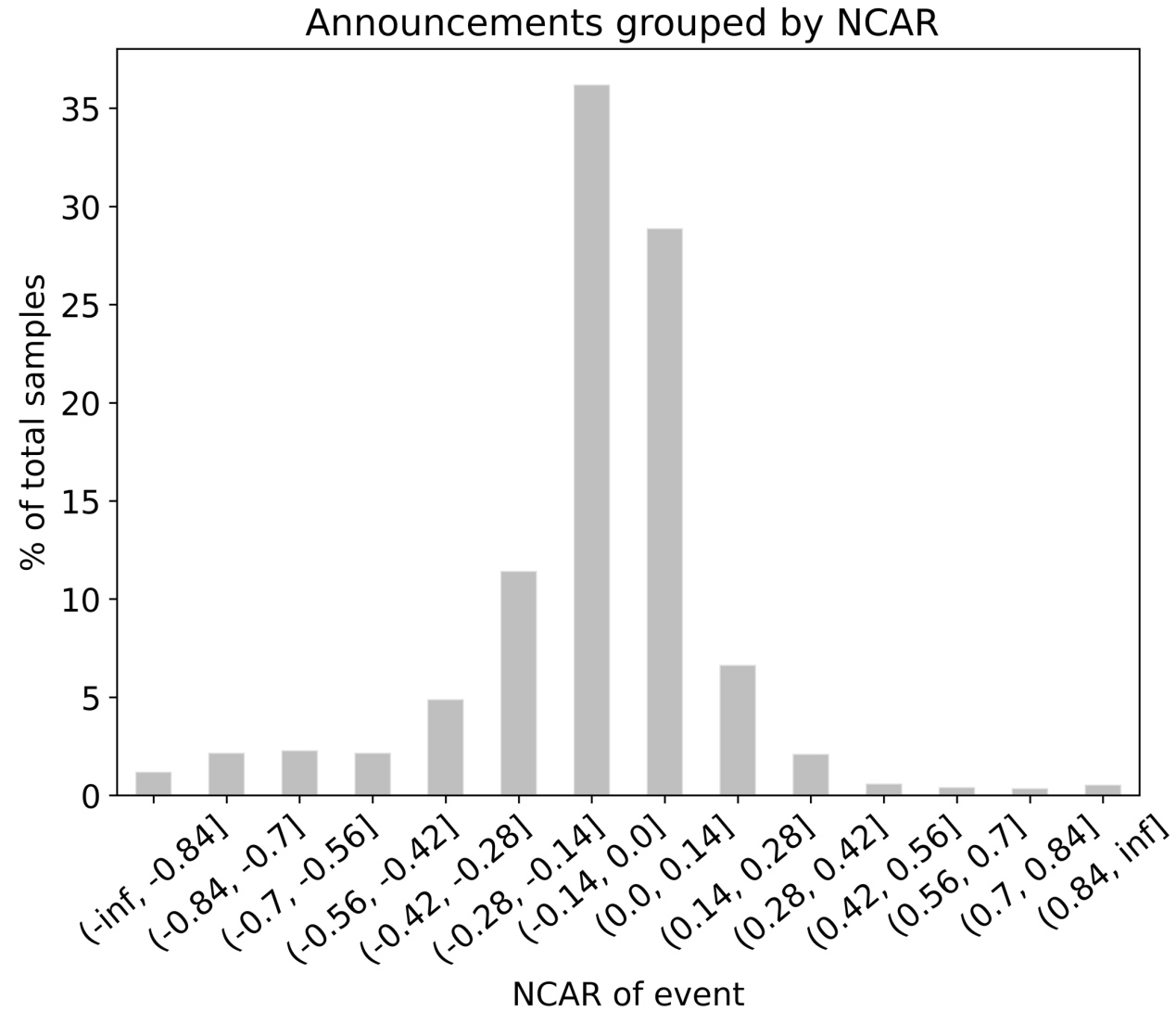


# Impact of company background on stock prices

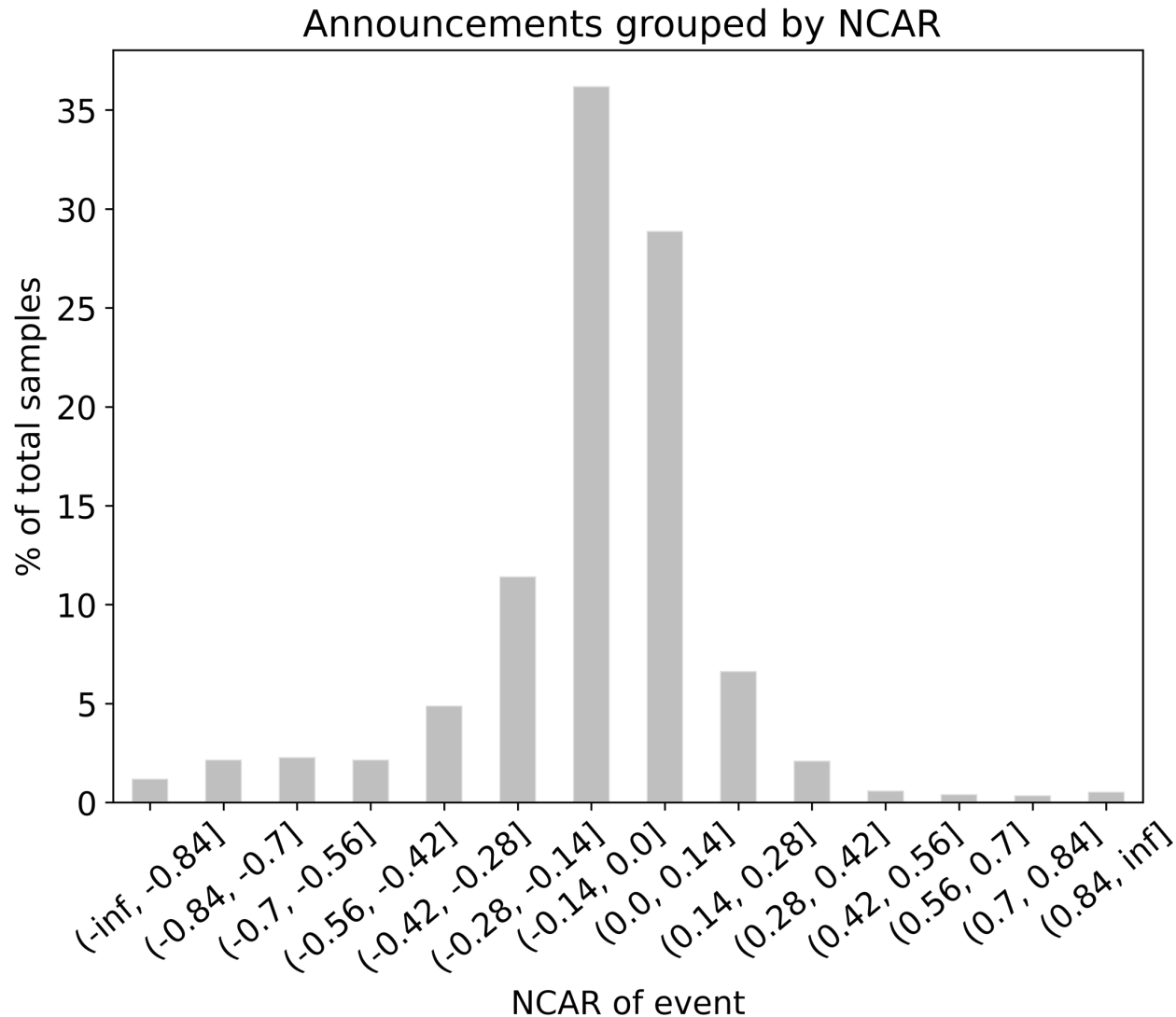


Generation of diverse feature space is important

# Class distribution



# Class distribution

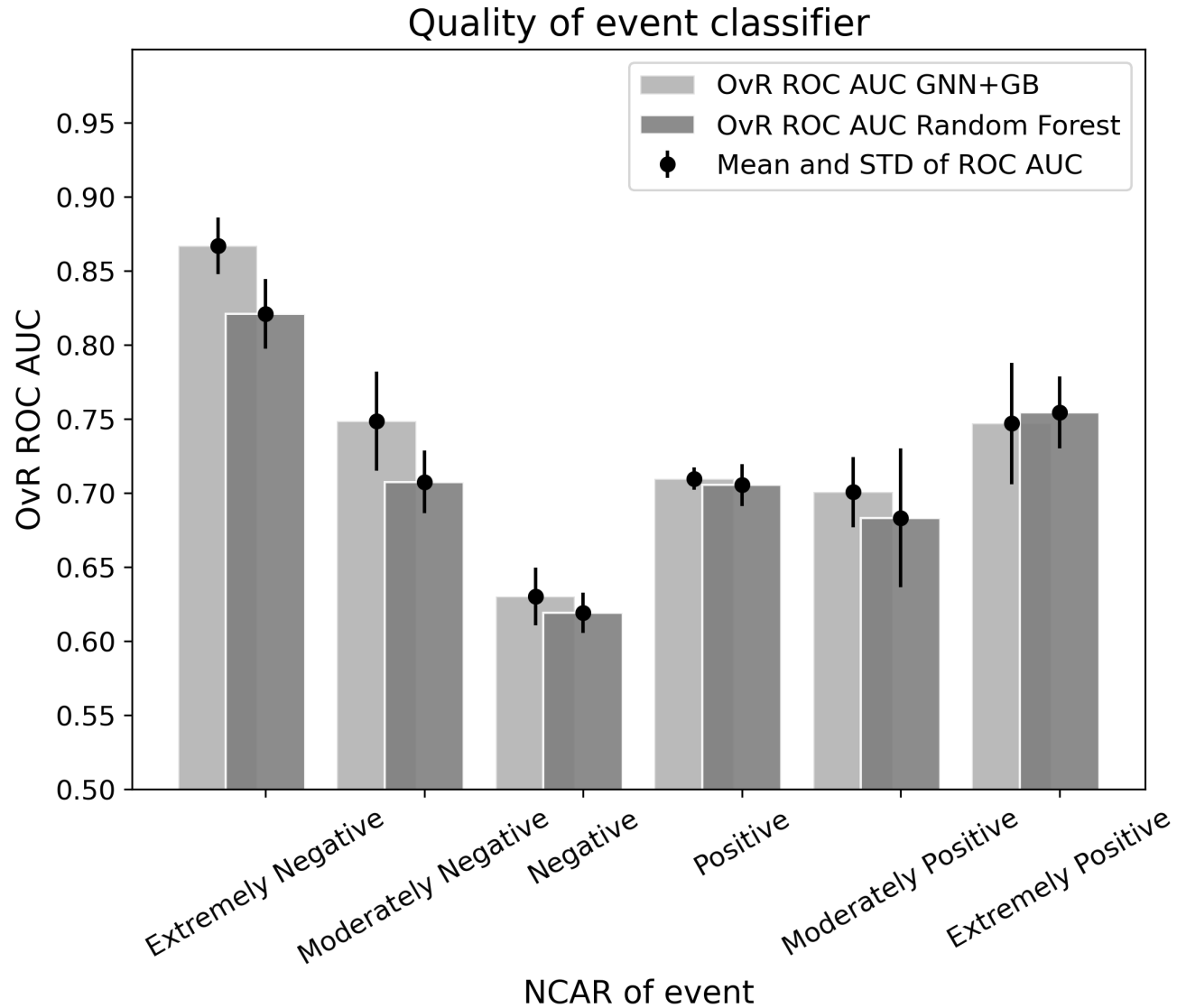


- Each class must be representative
- MAPE for expected return evaluation is 7%



**Categorize stock price change in 6 classes**

# Classification results



# Classification results

Class name*	Extremely	Moderately			Moderately	Extremely
	Negative	Negative	Negative	Positive	Positive	Positive
Stock price change range	$(-\infty, -0.28]$	$(-0.28, -0.14]$	$(-0.14, 0]$	$(0, 0.14]$	$(0.14, 0.28]$	$(0.28, +\infty)$
Number of events	211	189	599	478	110	67
Positive events**	72	106	421	366	83	57
Negative events**	139	83	178	112	27	10
OvR ROC AUC for GCN+GB	$0.87 \pm 0.02$	$0.77 \pm 0.03$	$0.63 \pm 0.02$	$0.71 \pm 0.01$	$0.70 \pm 0.02$	$0.75 \pm 0.04$
OvR ROC AUC for GB	$0.85 \pm 0.02$	$0.72 \pm 0.02$	$0.60 \pm 0.02$	$0.67 \pm 0.02$	$0.66 \pm 0.04$	$0.74 \pm 0.05$
Welch's t-test p-value***	0.09	0.05	0.002	$5.4 \times 10^{-5}$	0.02	0.65

# Classification results

Class name*	Extremely	Moderately			Moderately	Extremely
	Negative	Negative	Negative	Positive	Positive	Positive
Stock price change range	$(-\infty, -0.28]$	$(-0.28, -0.14]$	$(-0.14, 0]$	$(0, 0.14]$	$(0.14, 0.28]$	$(0.28, +\infty)$
Number of events	211	189	599	478	110	67
Positive events**	72	106	421	366	83	57
Negative events**	139	83	178	112	27	10
OvR ROC AUC for GCN+GB	$0.87 \pm 0.02$	$0.77 \pm 0.03$	$0.63 \pm 0.02$	$0.71 \pm 0.01$	$0.70 \pm 0.02$	$0.75 \pm 0.04$
OvR ROC AUC for GB	$0.85 \pm 0.02$	$0.72 \pm 0.02$	$0.60 \pm 0.02$	$0.67 \pm 0.02$	$0.66 \pm 0.04$	$0.74 \pm 0.05$
Welch's t-test p-value***	0.09	0.05	0.002	$5.4 \times 10^{-5}$	0.02	0.65

**GCN-based features improve quality**

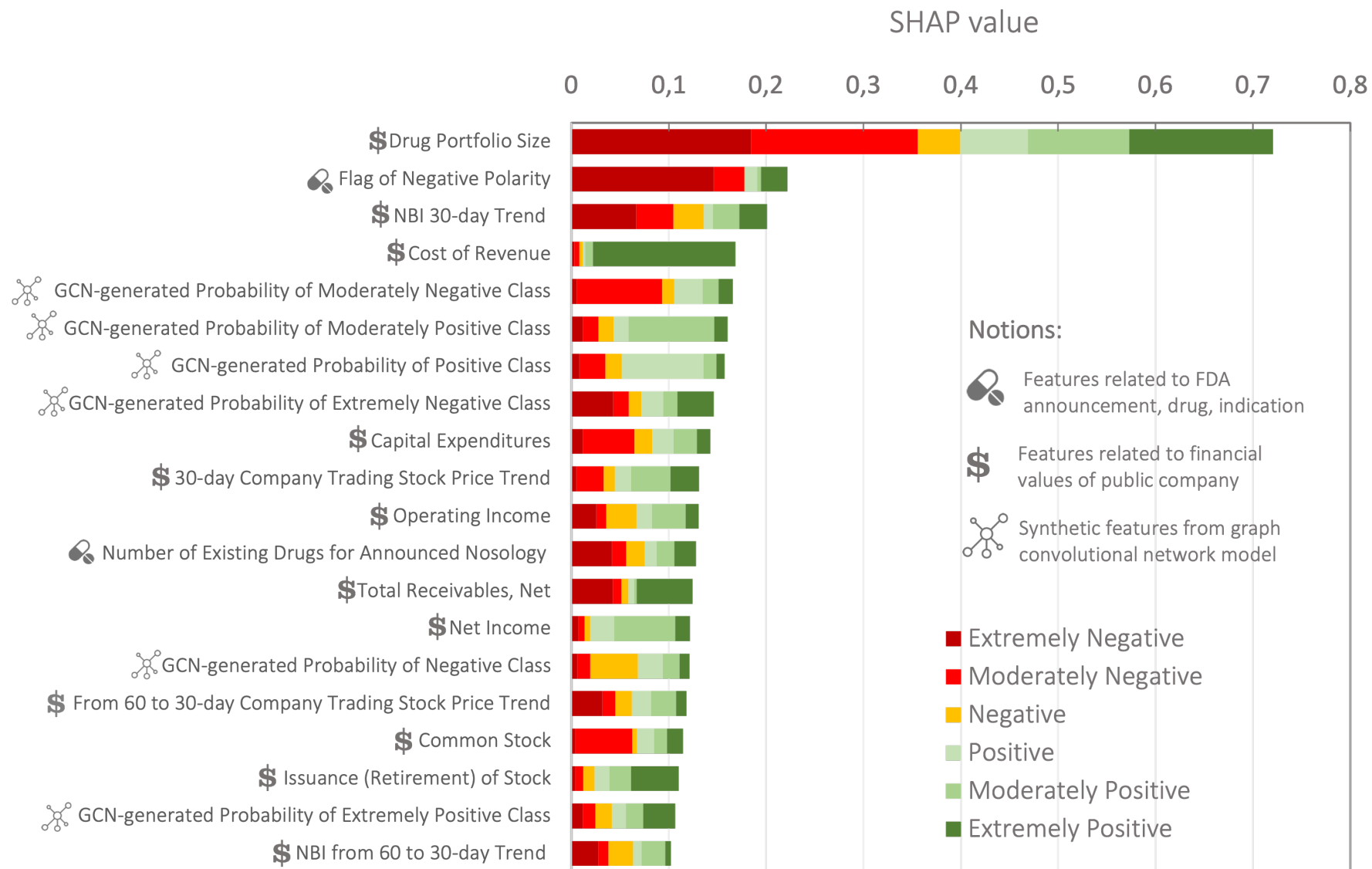
# Classification results

Class name*	Extremely	Moderately			Moderately	Extremely
	Negative	Negative	Negative	Positive	Positive	Positive
Stock price change range	$(-\infty, -0.28]$	$(-0.28, -0.14]$	$(-0.14, 0]$	$(0, 0.14]$	$(0.14, 0.28]$	$(0.28, +\infty)$
Number of events	211	189	599	478	110	67
Positive events**	72	106	421	366	83	57
Negative events**	139	83	178	112	27	10
OvR ROC AUC for GCN+GB	<b><math>0.87 \pm 0.02</math></b>	$0.77 \pm 0.03$	<b><math>0.63 \pm 0.02</math></b>	$0.71 \pm 0.01$	$0.70 \pm 0.02$	$0.75 \pm 0.04$
OvR ROC AUC for GB	$0.85 \pm 0.02$	$0.72 \pm 0.02$	$0.60 \pm 0.02$	$0.67 \pm 0.02$	$0.66 \pm 0.04$	$0.74 \pm 0.05$
Welch's t-test p-value***	0.09	0.05	0.002	$5.4 \times 10^{-5}$	0.02	0.65

**Extremely Negative class is the easiest to predict, while Negative class is the hardest one**



# Feature importance analysis

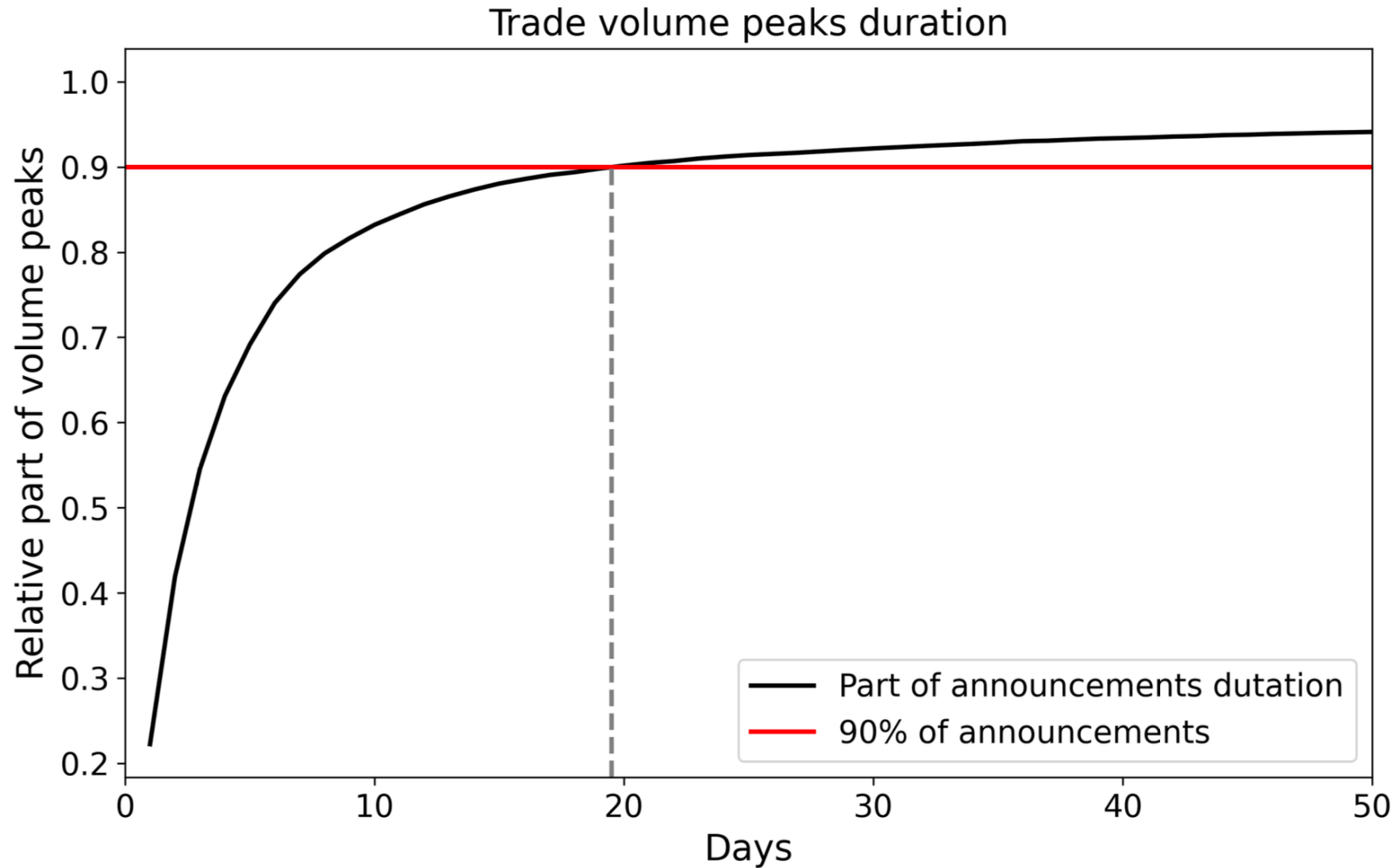


# Take away

- Prove the existence of event impact on time series before solving the prediction task
- Look at a relationship between different characteristics and a target variable
- Generate comprehensive feature space

# Appendix

# Post-announcement period duration



# Impact dependence on company age

