



IB QUANT CASE STUDY

Divyansh
IIT Kanpur

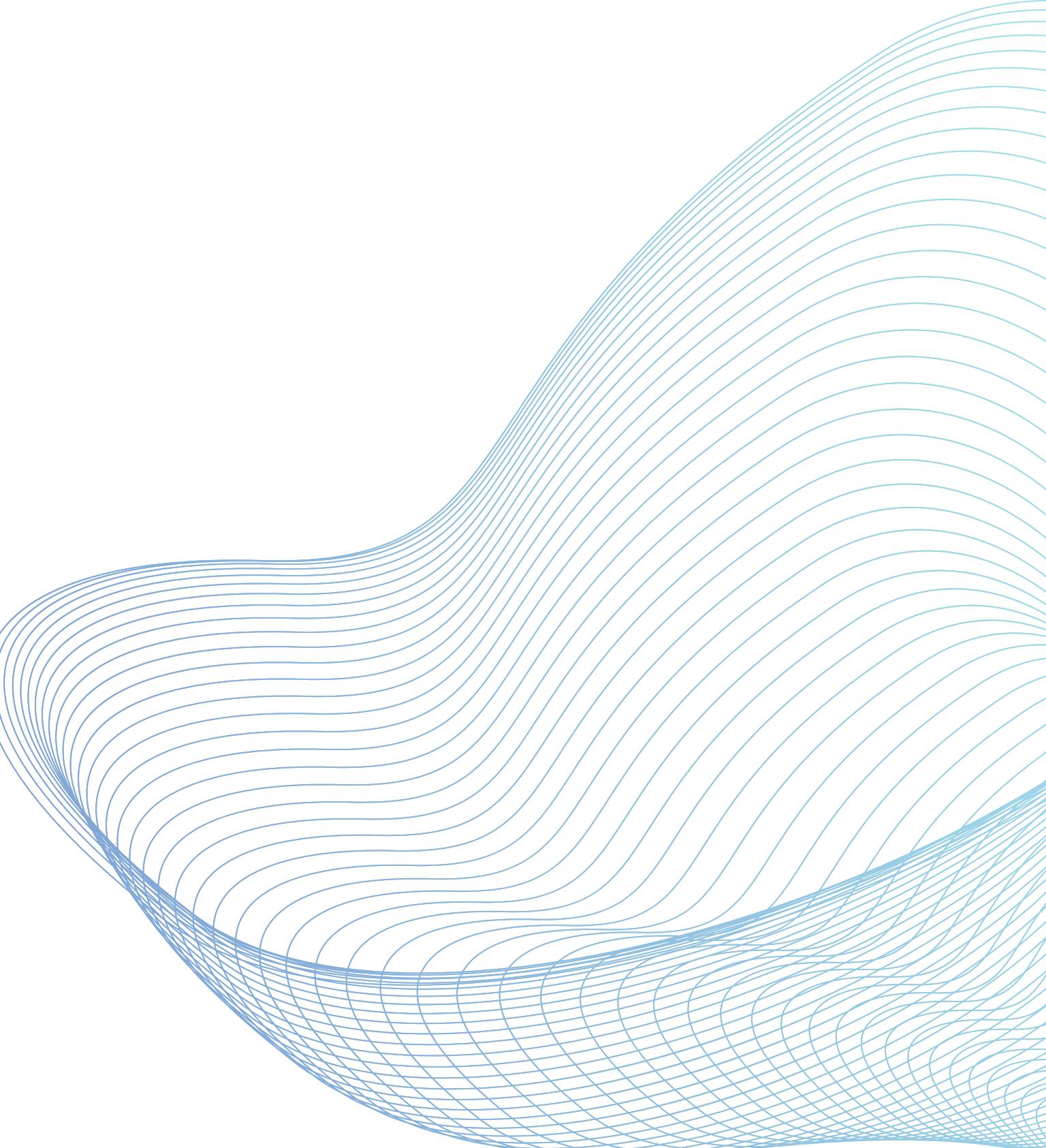


TABLE OF CONTENT

1. INTRODUCTION

2. CASE UNDERSTANDING

3. SOLUTION APPROACH

INTRODUCTION

ABOUT ME

I, Divyansh, am final year undergraduate student in the CSE department at IIT Kanpur.

Passion

I consider myself tech enthusiast and tech optimist. I am excited about how the recent advancements in technology will shape the future of Health, Finance, etc sectors

Interests

I am exploring different tech like systems, ML, etc and majorly their applications in various domains. Apart from these, I also like to paint, and read blogs



EXPERIENCE



Research Internship

May - July, 2024

- Multimodal Attribution in Large Language Models



Student Research Intern

April - July, 2023

- Short Video propaganda detection with LLMs

Data Science Leader

Sept'2022 - April'2024

- Poverty Prediction with Haryana govt.
- AI Medical Nurse Assistant
- Tele-ambulance System with UP govt



CASE UNDERSTANDING

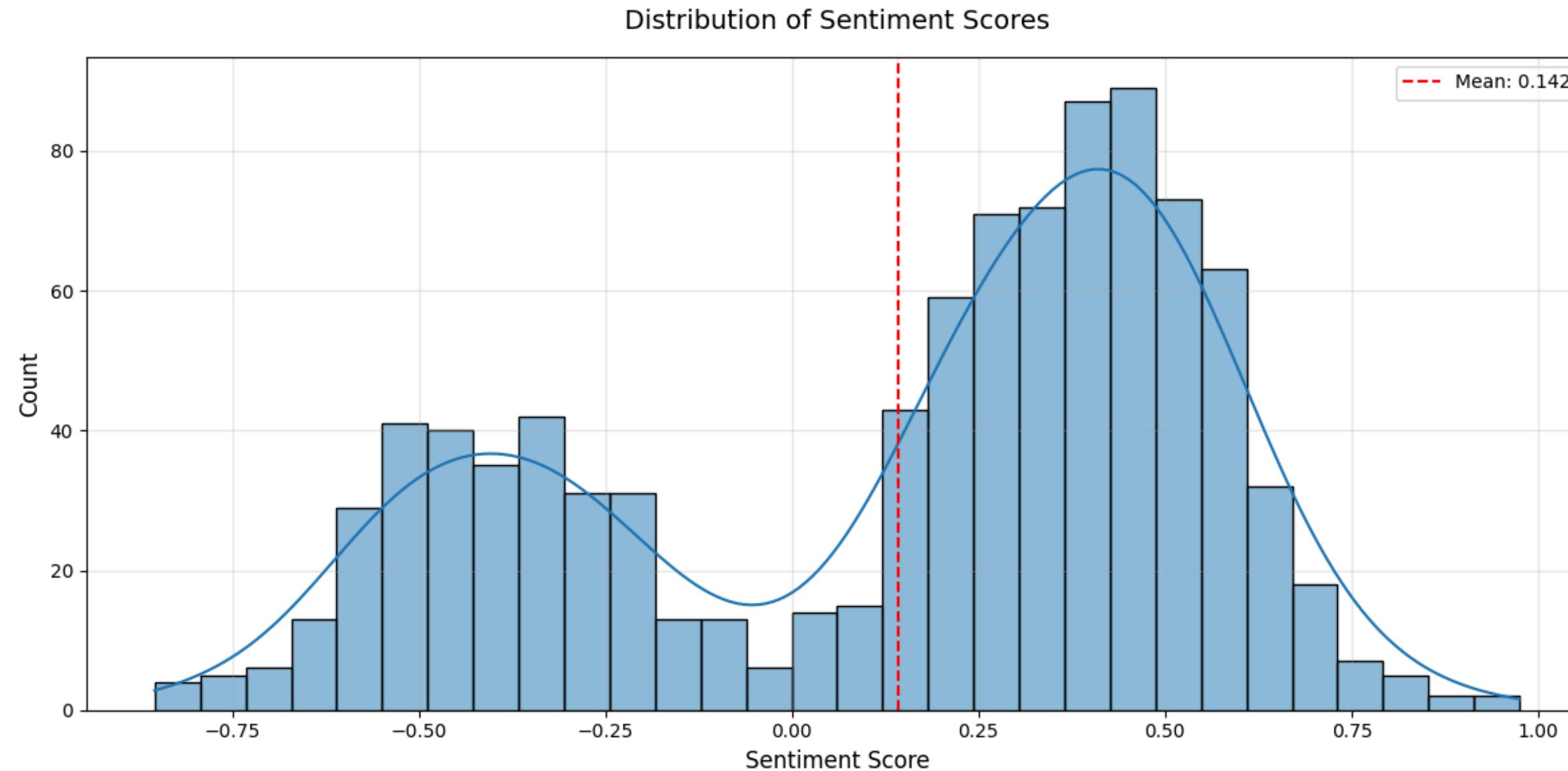
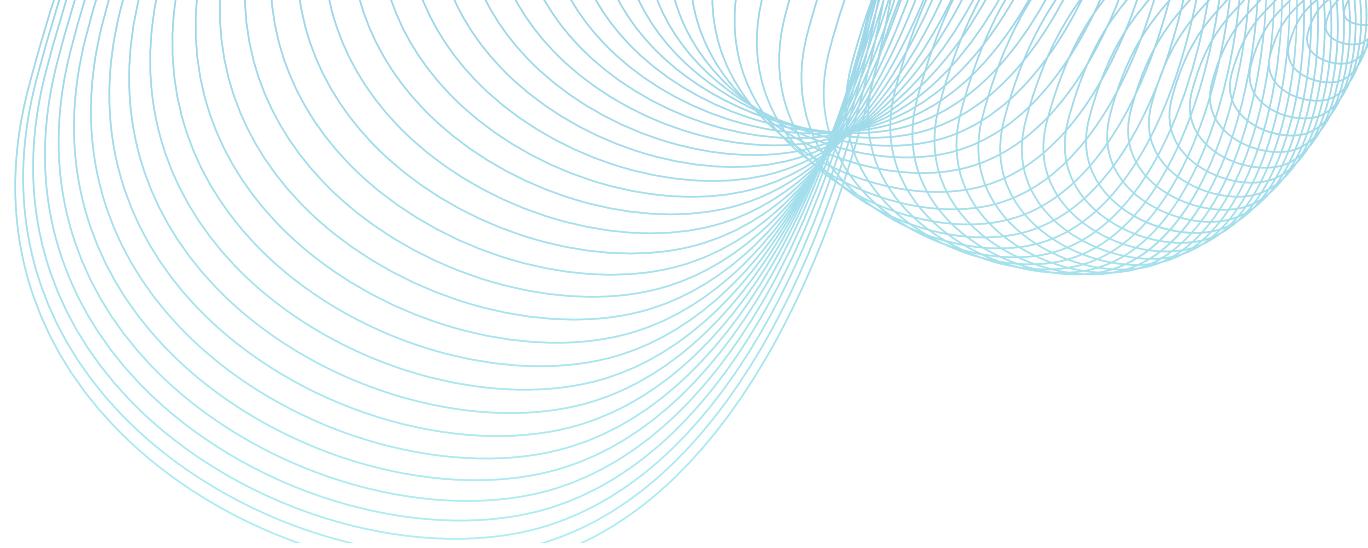
PROBLEM STATEMENT

Given a financial text (news/microblog/headline), we have to identify the aspects being discussed (from predefined categories) and predict the sentiment score (-1 to +1) for each target.

	sentence	snippets	target	sentiment_score	aspects	format	label
0	Still short \$LNG from \$11.70 area...next stop ...	['Still short \$LNG from \$11.70 area...next sto...']	LNG	-0.543	['Stock/Price Action/Volatility/Short Selling']	post	2
1	\$PLUG bear raid	['bear raid']	PLUG	-0.480	['Stock/Price Action/Bearish']	post	2
2	How Kraft-Heinz Merger Came Together in Speedy...	['Merger Came Together in Speedy 10 Weeks']	Kraft	0.214	['Corporate/M&A/M&A']	headline	0
3	Slump in Weir leads FTSE down from record high	['down from record high']	Weir	-0.827	['Market/Volatility/Volatility']	headline	2
4	\$AAPL bounces off support, it seems	['bounces off support']	AAPL	0.443	['Stock/Price Action/Bullish/Bullish Behaviour']	post	0

DATA ANALYSIS

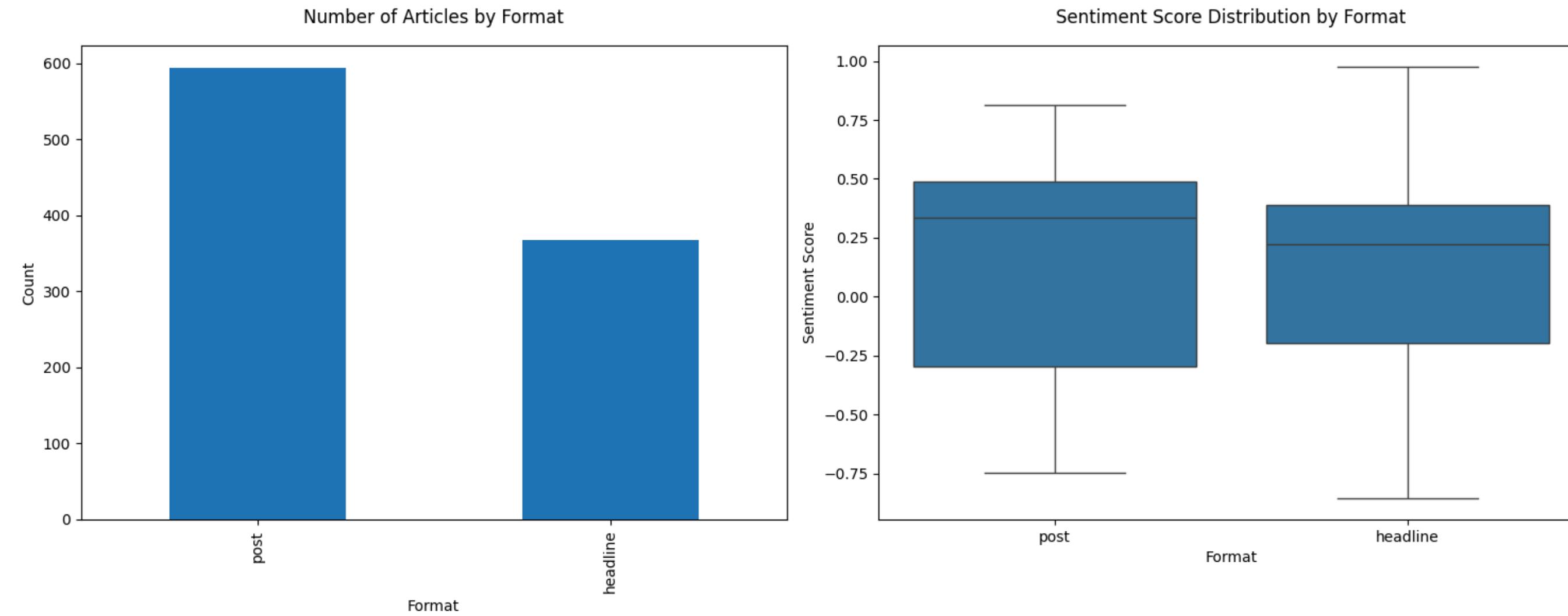
Studying the distribution of sentiment scores over all data points.



Majorly, we have positive and negative sentiment scores in the data with few data points with neutral sentiment score as well

DATA ANALYSIS

How format of news affects the sentiments?

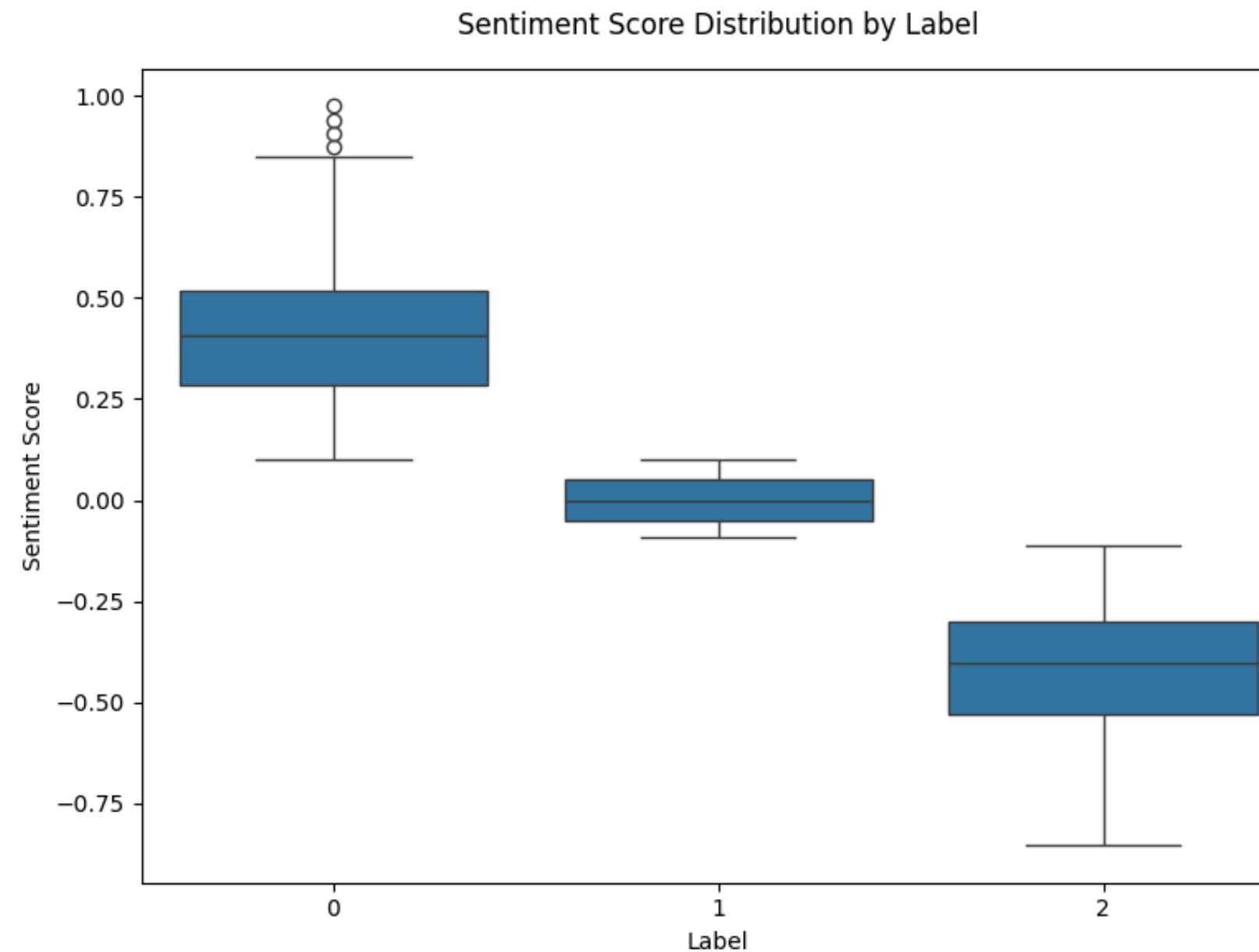


We can see that, whether the news is in post or headline format it doesn't really impact the sentiment of the sentence in it.

DATA ANALYSIS



What is the label for?



Clearly the label seems to be extra redundant information for the sentiment score and we will not use this for our further task since it will make the sentiment score prediction task too trivial

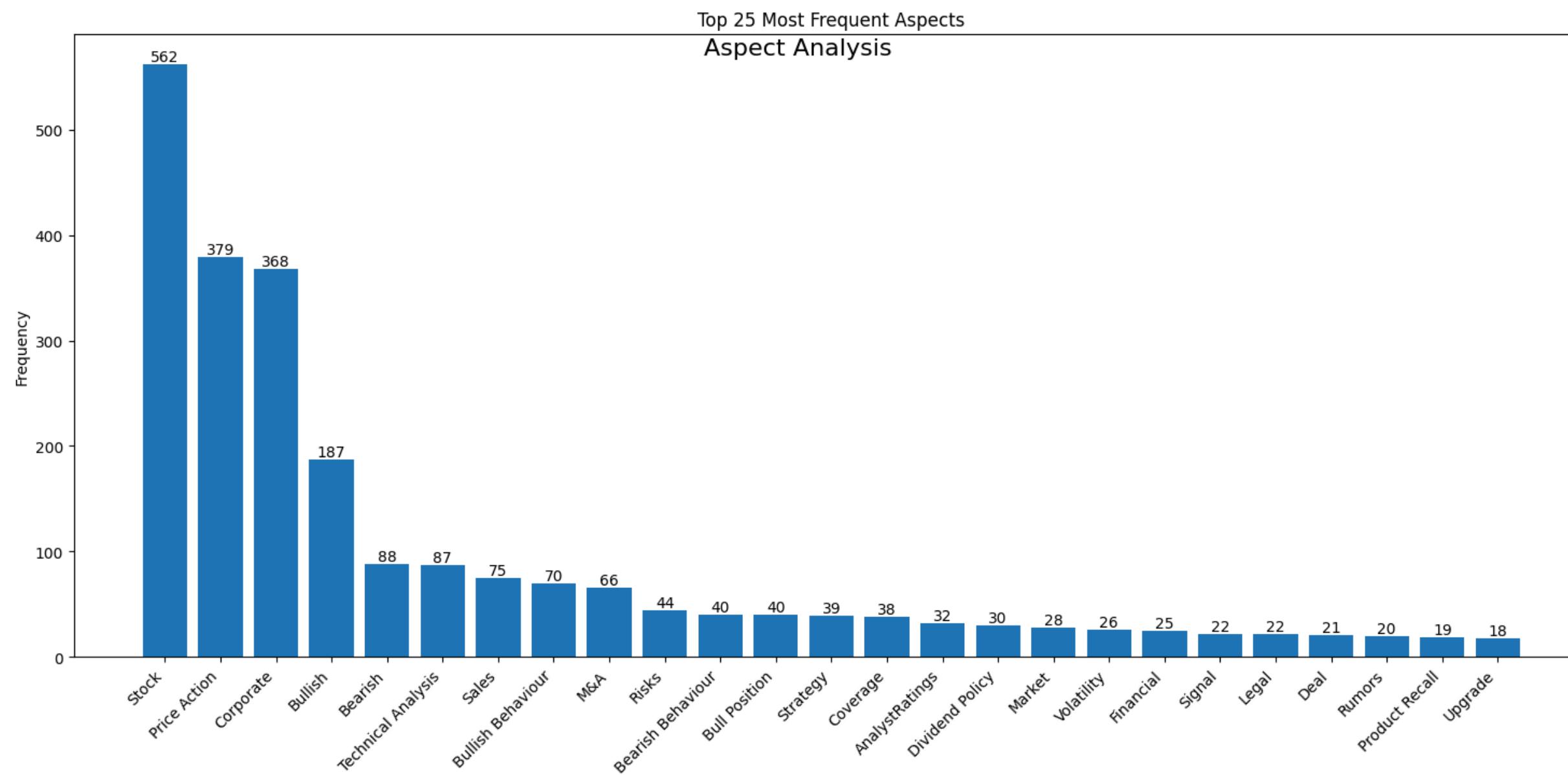
ASPECTS ANALYSIS

We first preprocess the aspects of each of input data points.

['Stock/Price Action/Bullish/
/Stock/Bullish Behaviour']



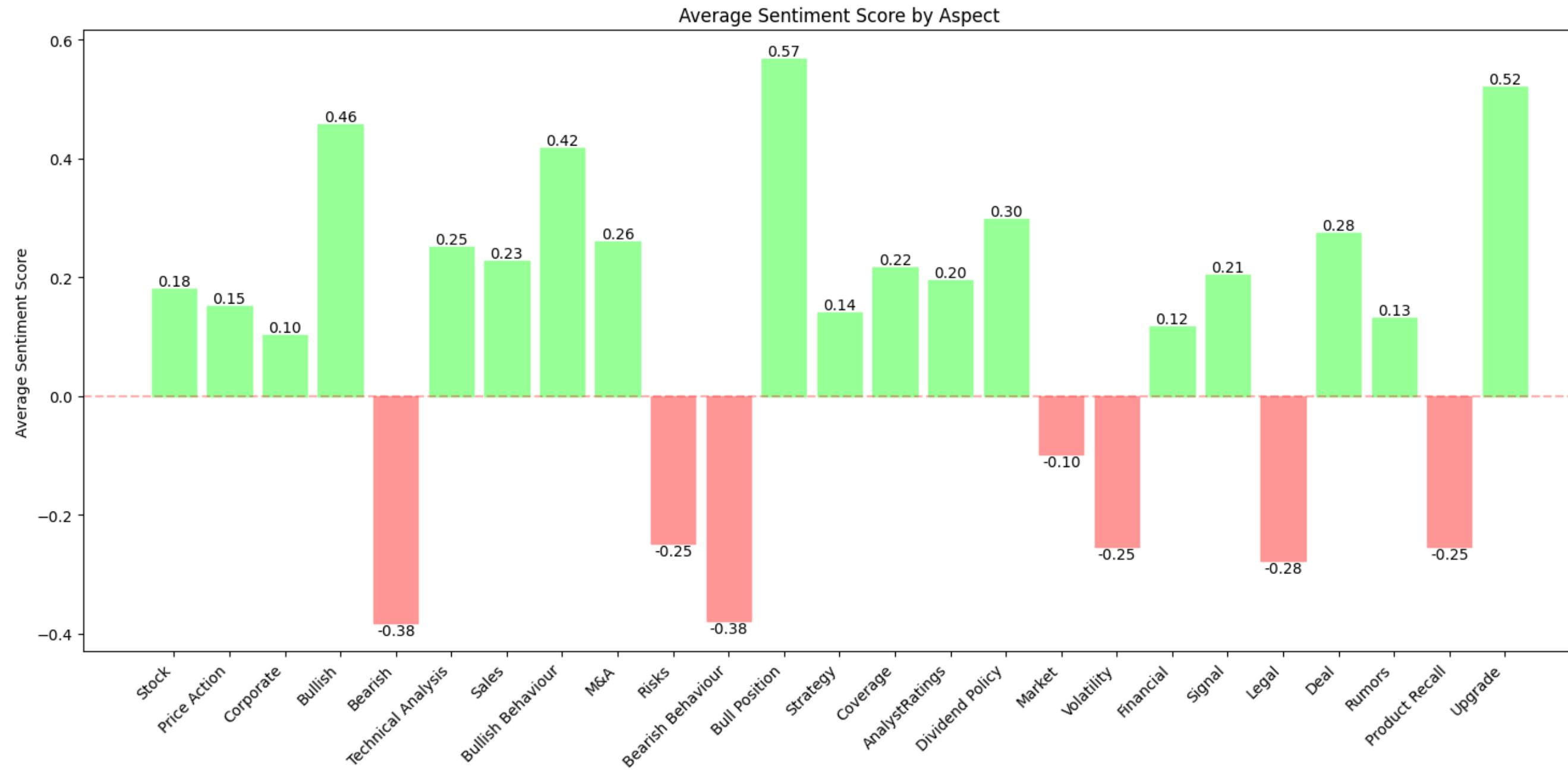
["Stock", "Price Action", "Bullish",
"Bullish Behaviour"]



There is a clear data imbalance with the frequency of different aspects.

ASPECTS ANALYSIS

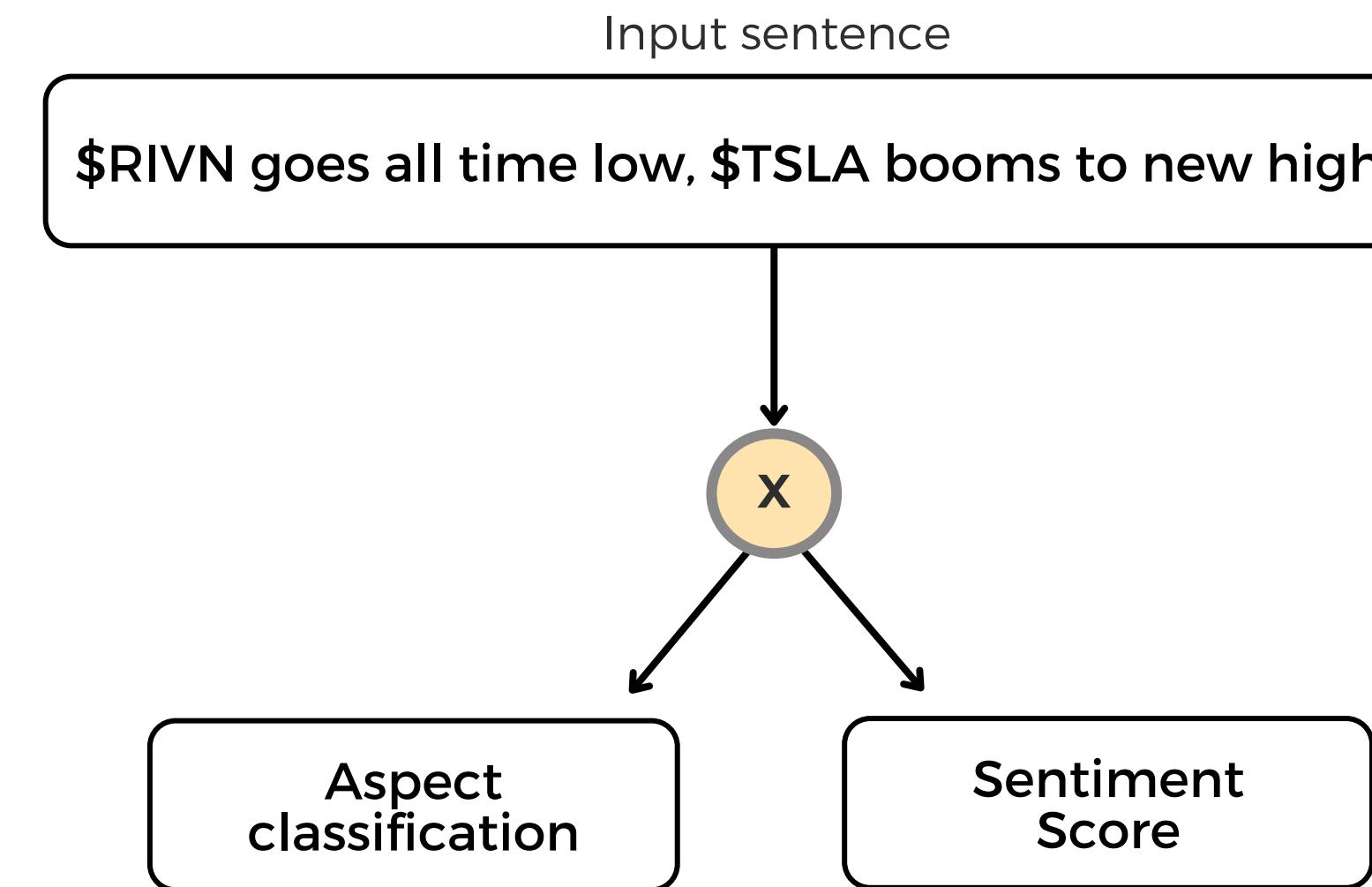
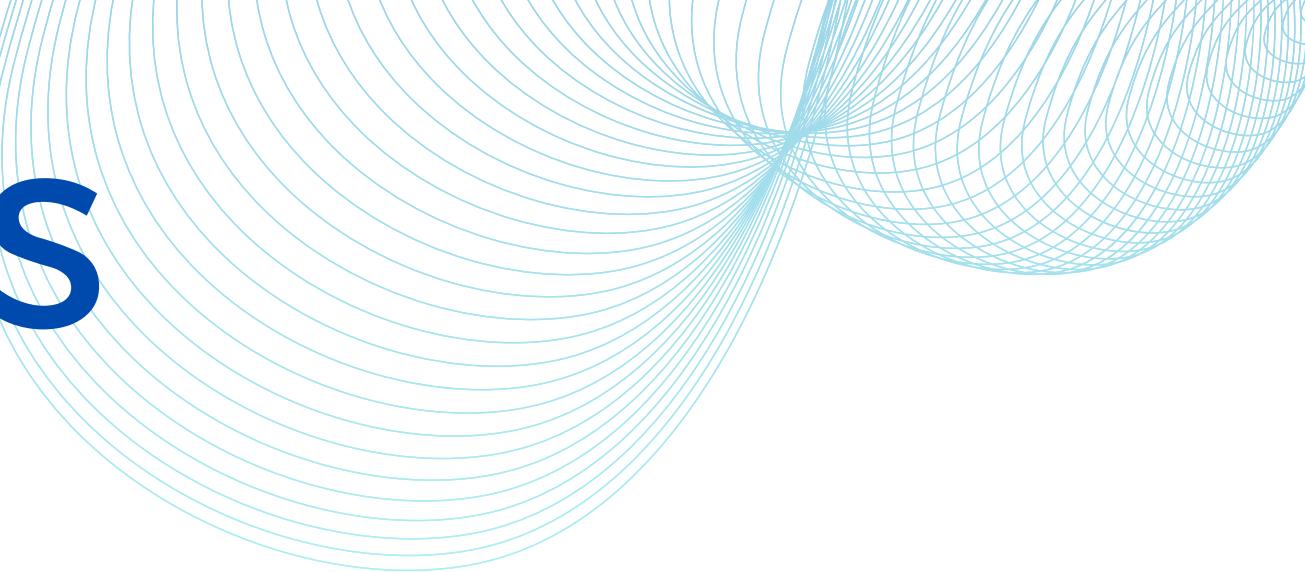
Are aspects and sentiments related?



We can notice here that sentiment score prediction and aspect classification are not completely independent task we can try to benefit from one another

KEY CHALLENGES

Role of target



KEY CHALLENGES

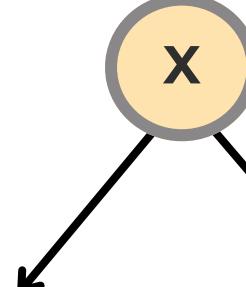
Role of target

Input sentence

\$RIVN goes all time low, \$TSLA booms to new high

+

RIVN
target



Aspect
classification

Sentiment
Score

KEY CHALLENGES

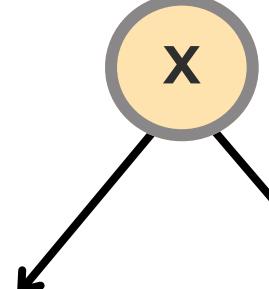
Role of target

Input sentence

\$RIVN goes all time low, \$TSLA booms to new high

+

RIVN
target



Aspect
classification

Sentiment
Score

KEY CHALLENGES

Role of target

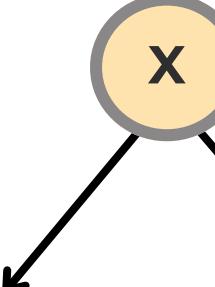
Input sentence

\$RIVN goes all time low, \$TSLA booms to new high

+

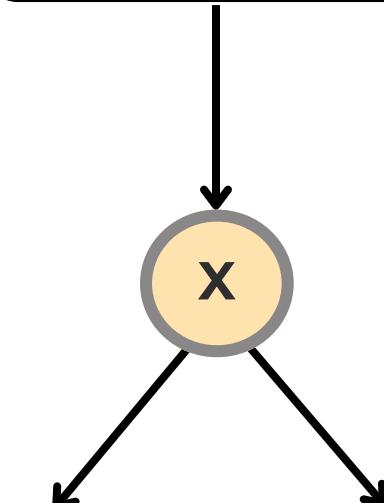
TSLA

target



Aspect
classification

Sentiment
Score



KEY CHALLENGES

Role of target

Input sentence

\$RIVN goes all time low, \$TSLA booms to new high

+

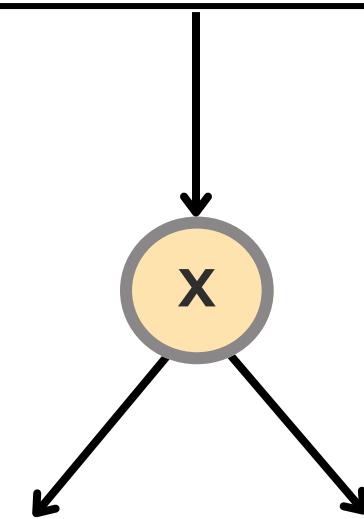
target

TSLA

X

Aspect
classification

Sentiment
Score



SOLUTION APPROACH

SEMANTIC MEANING

Vectorsize the natural text to give it numerical semantic meaning

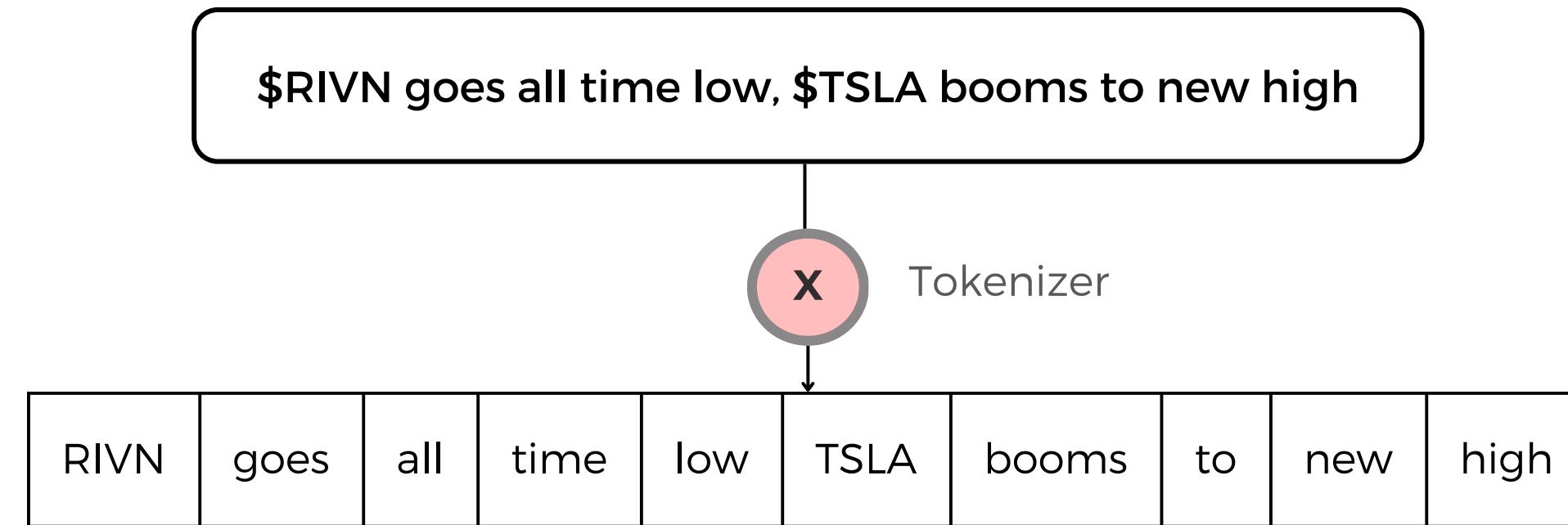
Input sentence

\$RIVN goes all time low, \$TSLA booms to new high

SEMANTIC MEANING

Vectorsize the natural text to give it numerical semantic meaning

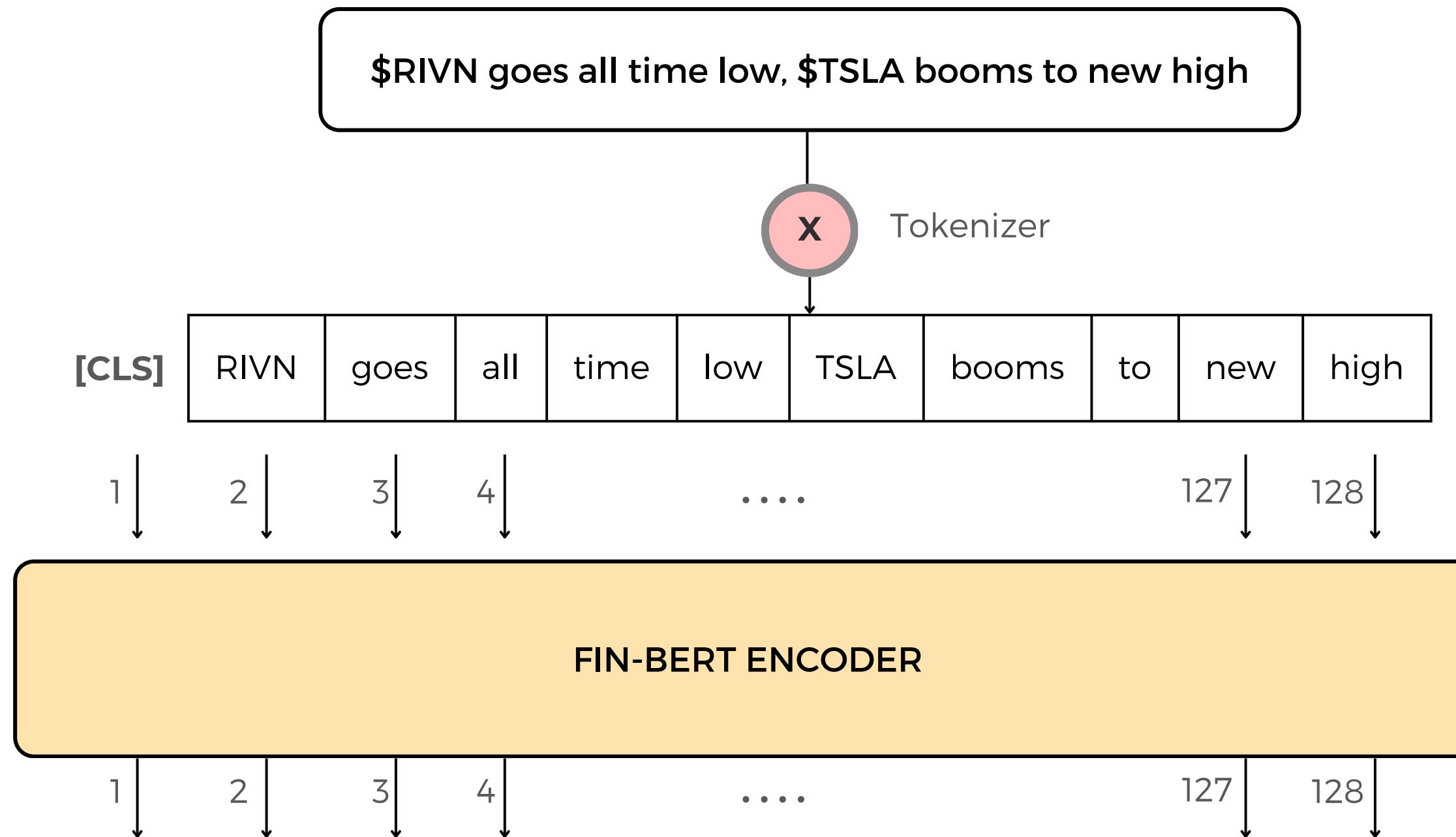
Input sentence



SEMANTIC MEANING

Vectorsize the natural text to give it numerical semantic meaning

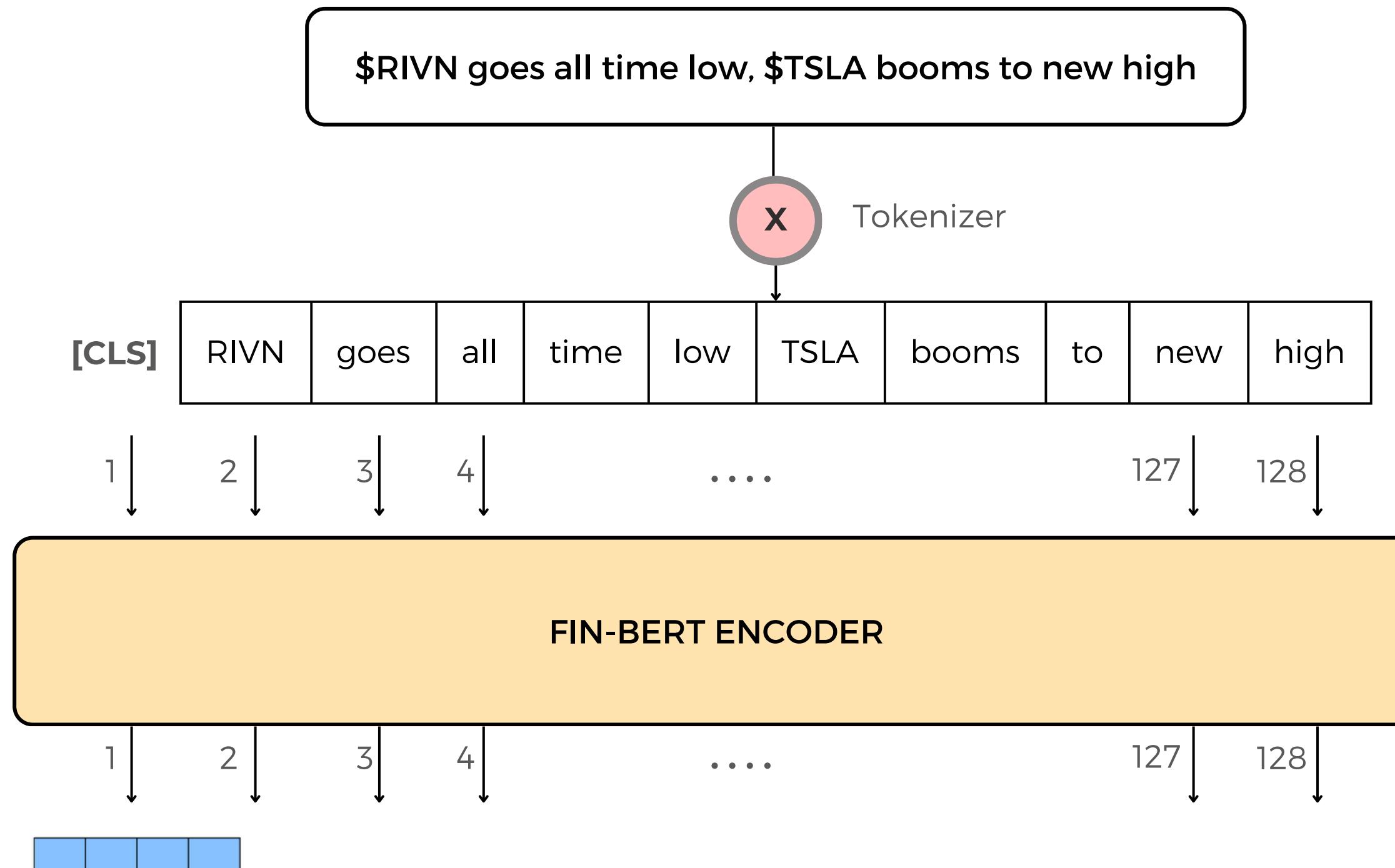
Input sentence



SEMANTIC MEANING

Vectorsize the natural text to give it numerical semantic meaning

Input sentence

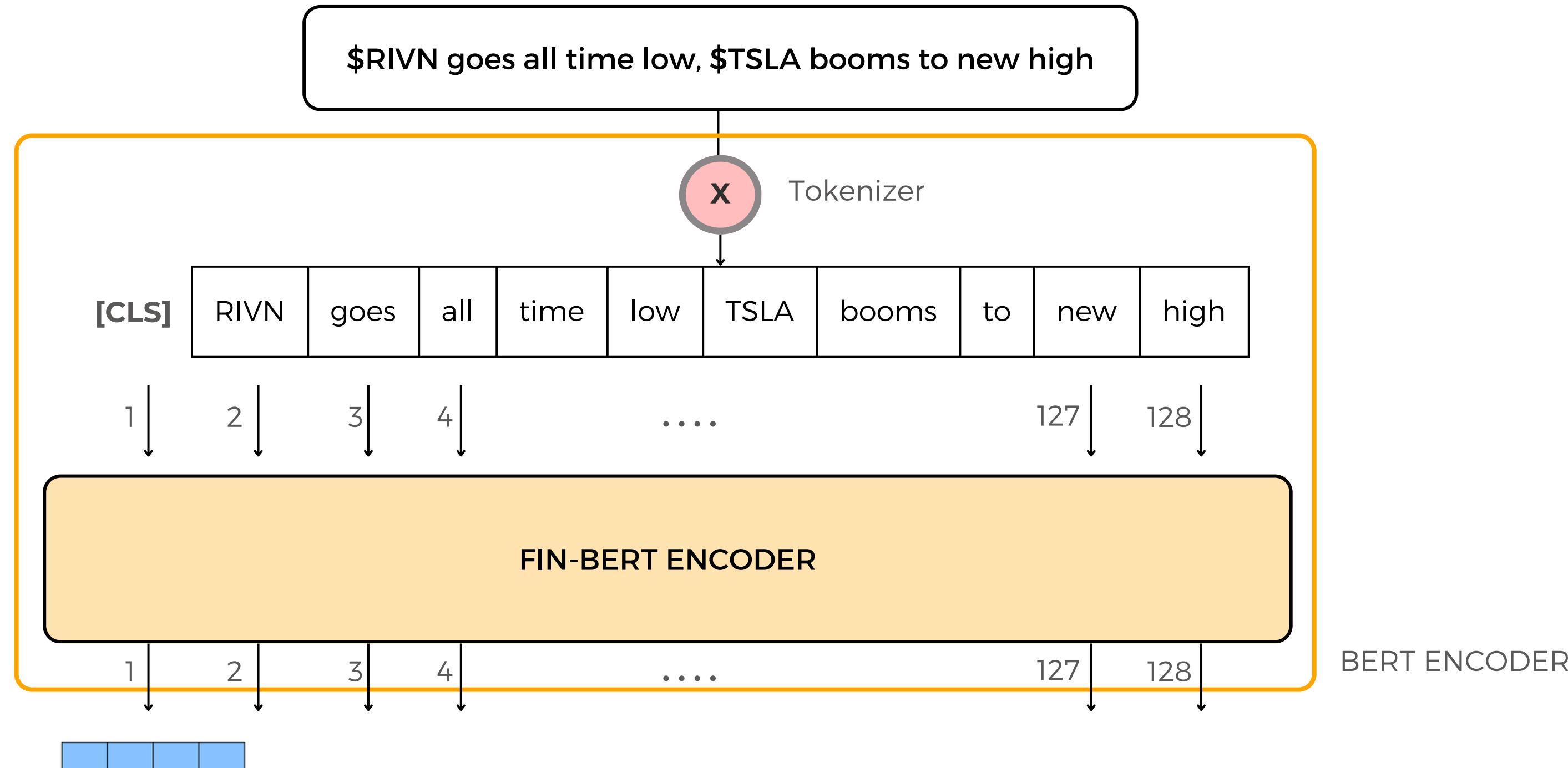


[CLS] embedding: vector representation of input sentence

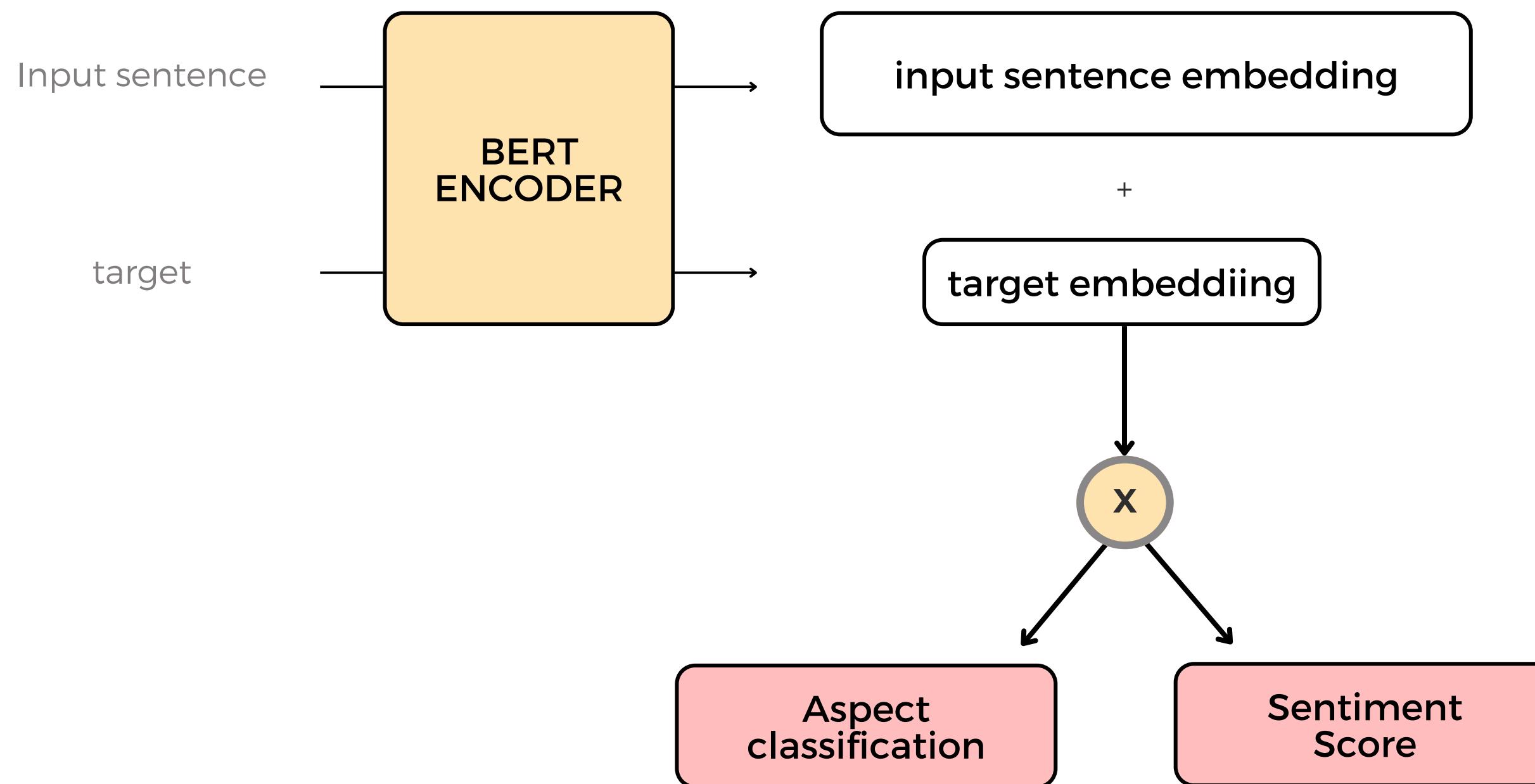
SEMANTIC MEANING

Vectorsize the natural text to give it numerical semantic meaning

Input sentence

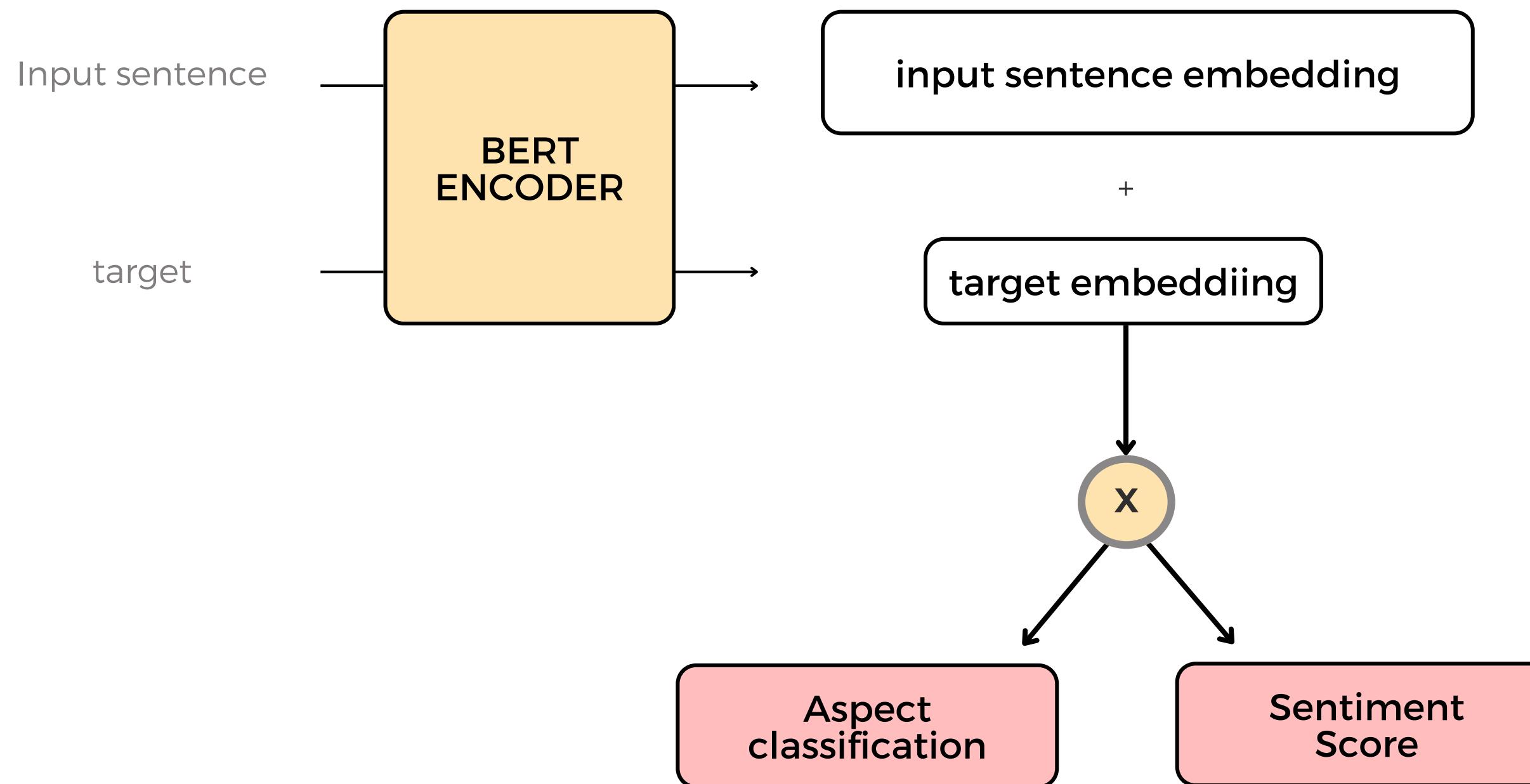


TARGET NEEDS?



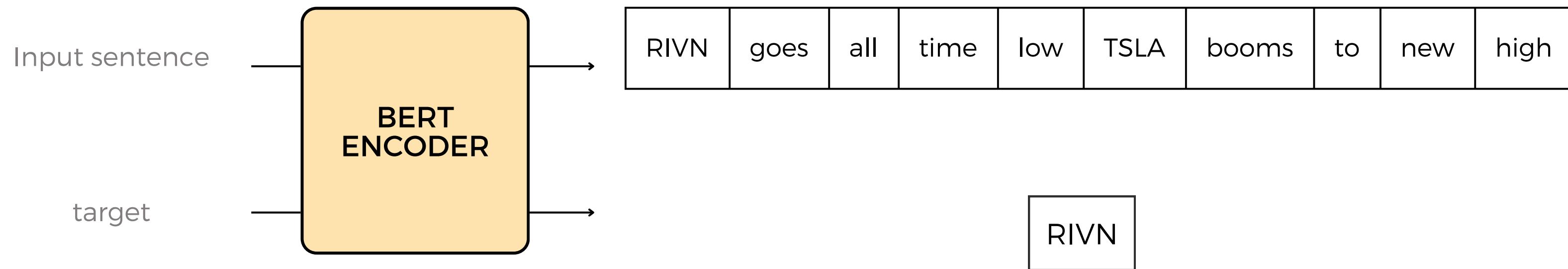
ATTENTION

Target needs attention and that's all we need.



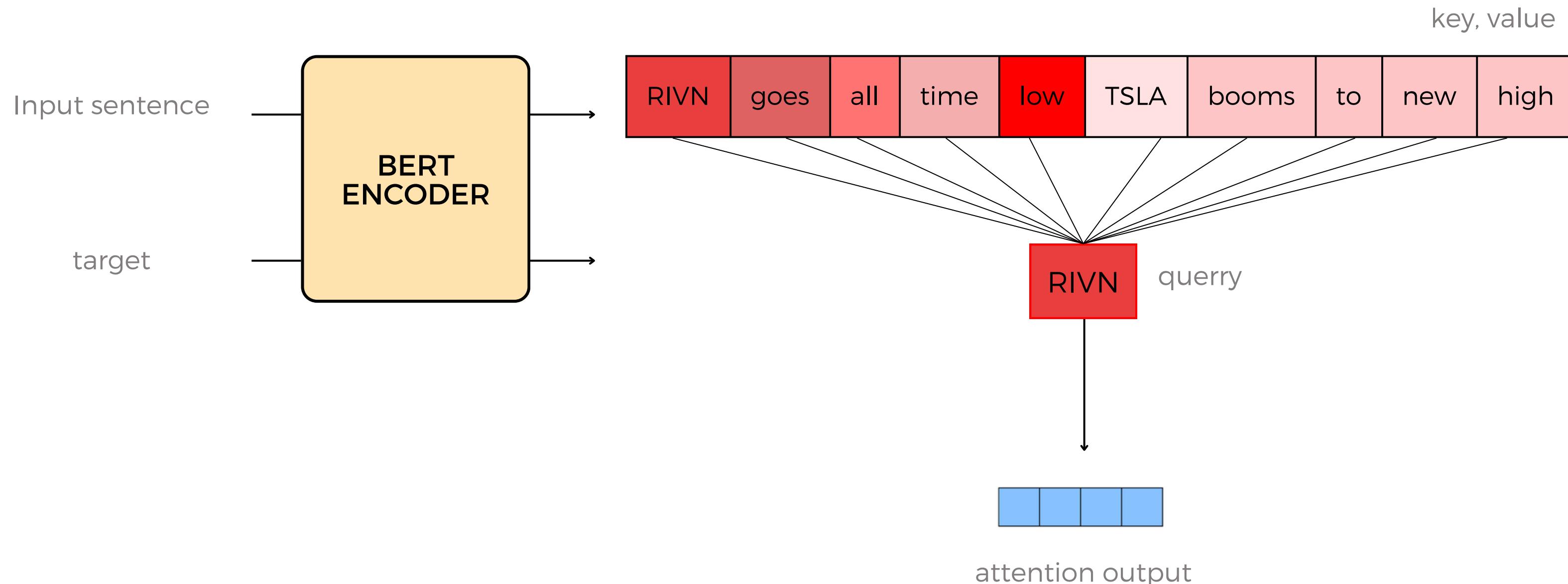
ATTENTION

Target needs attention and that's all we need.



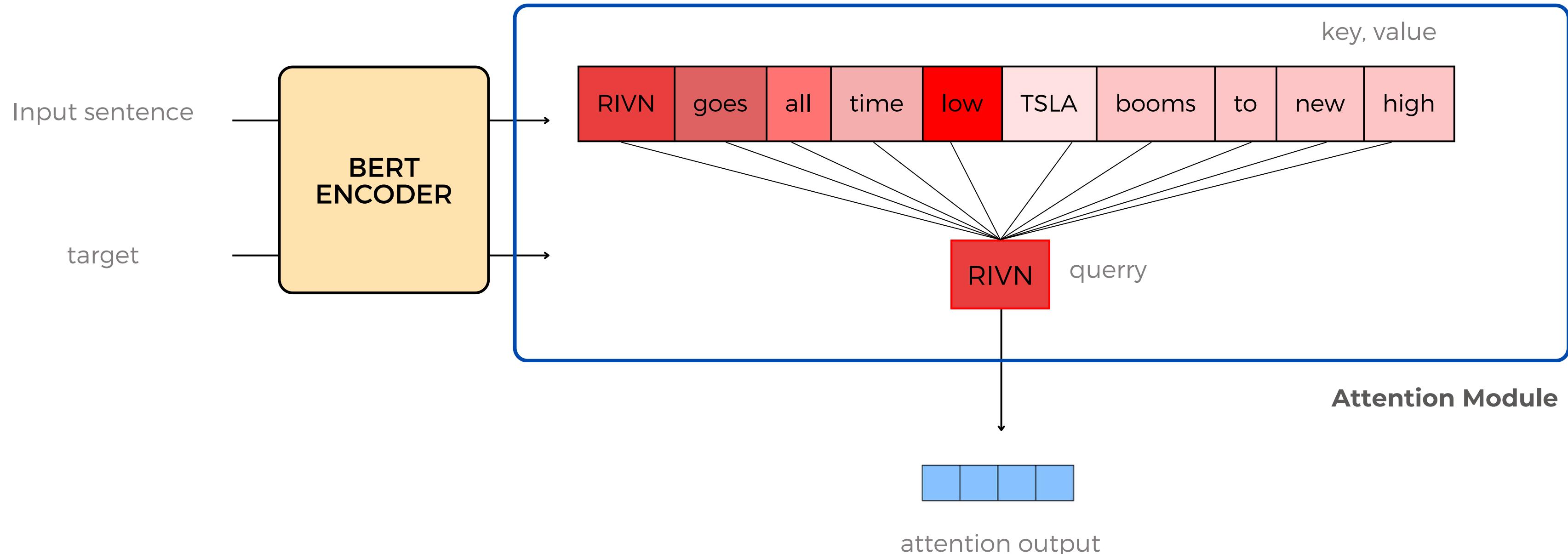
ATTENTION

Target needs attention and that's all we need.



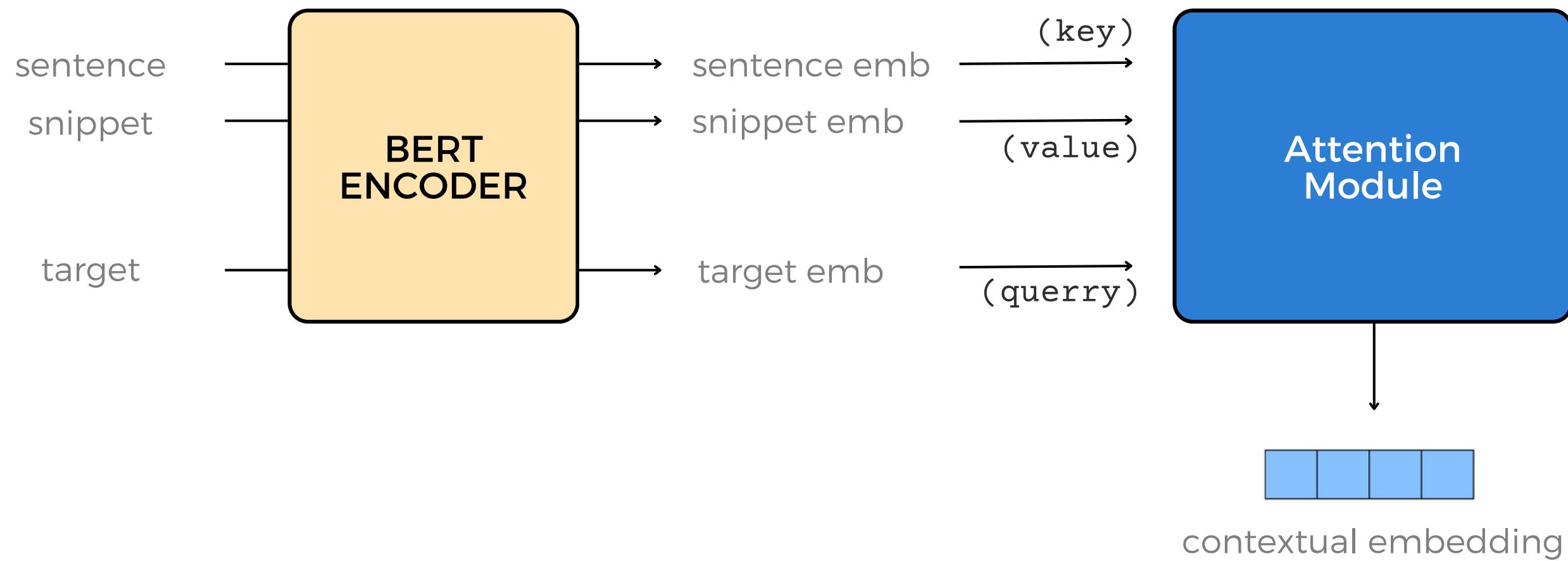
ATTENTION

Target needs attention and that's all we need.



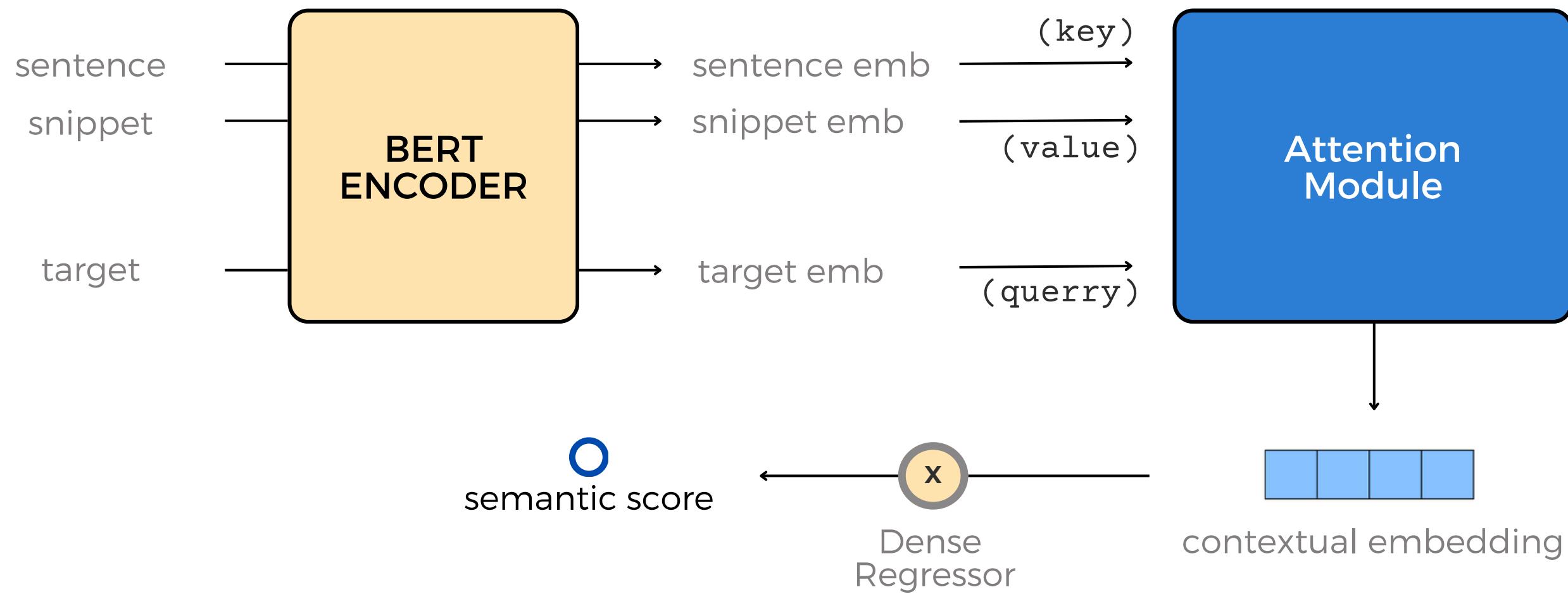
FINAL PIPELINE

Classifier for aspects and regressor for semantic score



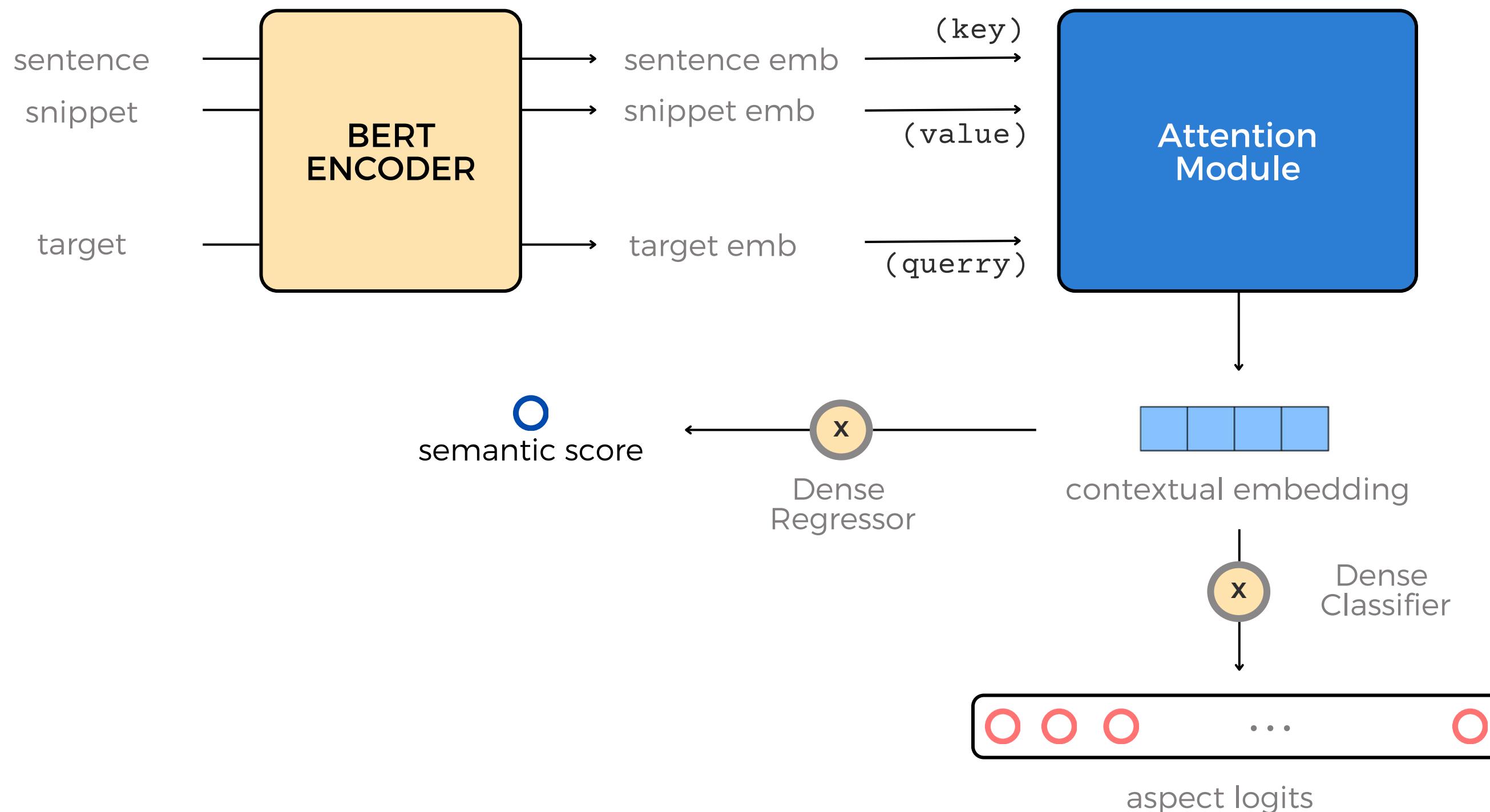
FINAL PIPELINE

Classifier for aspects and regressor for semantic score



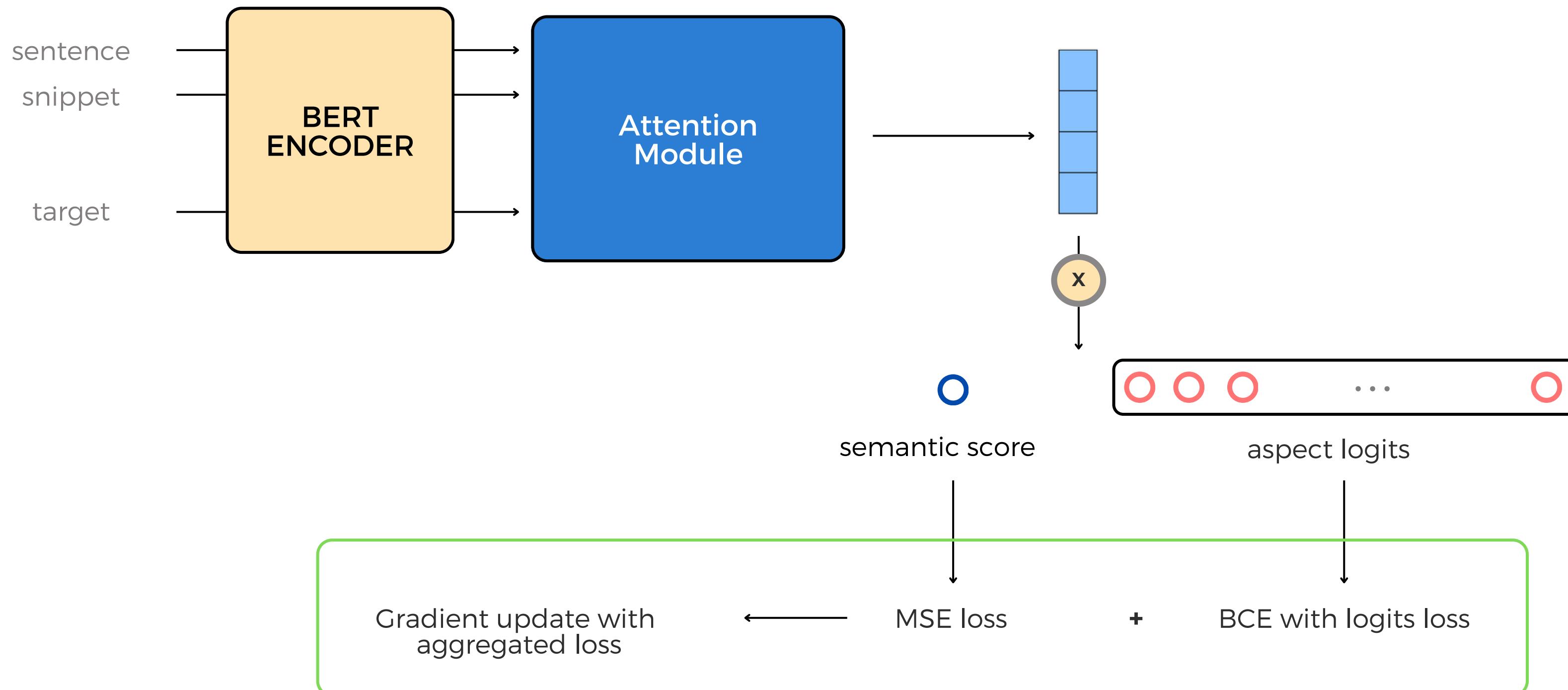
FINAL PIPELINE

Classifier for aspects and regressor for semantic score



BACKPROPAGATION

Classifier for aspects and regressor for semantic score



EVALUATION

- Results on Training Data:
 - Aspect Classification Metrics(Micro-averaged):
 - Precision: 0.4584
 - Recall: 0.5059
 - F1-score: 0.4810
 - Sentiment Regression Metrics:
 - MSE: 0.1544
- Results on Validation Data:
 - Aspect Classification Metrics (Micro-averaged):
 - Precision: 0.4582
 - Recall: 0.4275
 - F1-score: 0.4423
 - Sentiment Regression Metrics:
 - MSE: 0.1175



THANK YOU