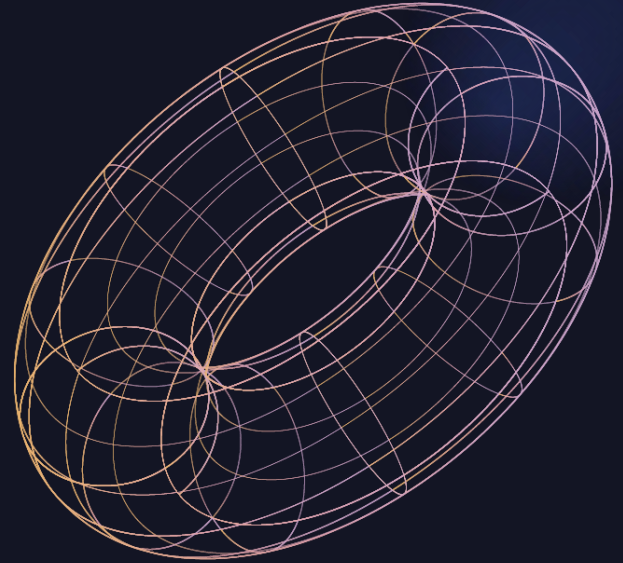# Earnings Calls Analyst App

Prepared by : Ahmed Taha Ahmed

Meet the team

Eng. Ahmed Taha Ahmed
- Computer Software Engineer
- Obsessed with AI
- Hobbies: Reading and Searching

Table of contents

01

Task Overview

Task Overview

- Fine-tune BERT on Earnings call dataset to predict whether the stock price will increase or decrease in next quarter
- Named Entity Recognition model using BERT to extract key financial metrics from earnings call transcripts
- Sentiment analysis model using BERT to determine management sentiment (positive/negative) from the transcript text.
- pipeline architecture that ingests new earnings call transcripts, passes them through the NLP models to extract key phrases, metrics and sentiment, and makes an investment recommendation based on the outputs.
- Calculate return on investment and other metrics.
- Identify any potential biases in your NLP models that could negatively impact performance.
- Report , Paper and Interactive demo

DataSet

- Earnings call transcripts scraped from <u>link</u>
- Each row consists of a date, exchange, quarter, ticker and transcript.
- Each transcript contains the information about the quarter status at this time of the call .
- Each transcript consists of different parts the transcript main content ,which contains the manger talk about quarter talk then the next part is the Q&A part which contains some questions about the quarter status.
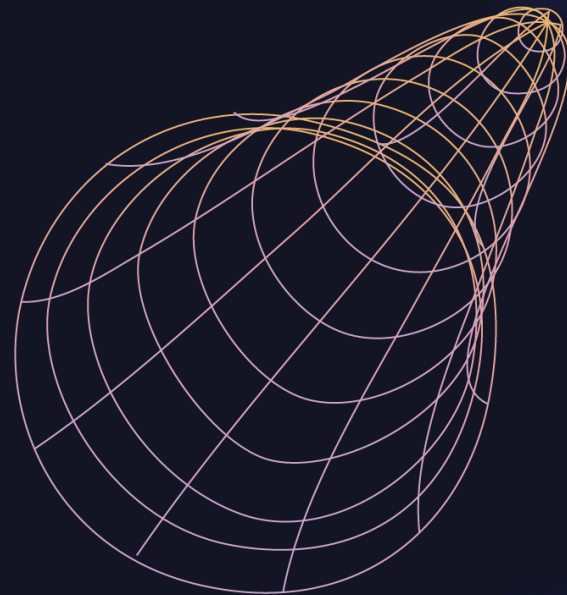
Pipeline Overview
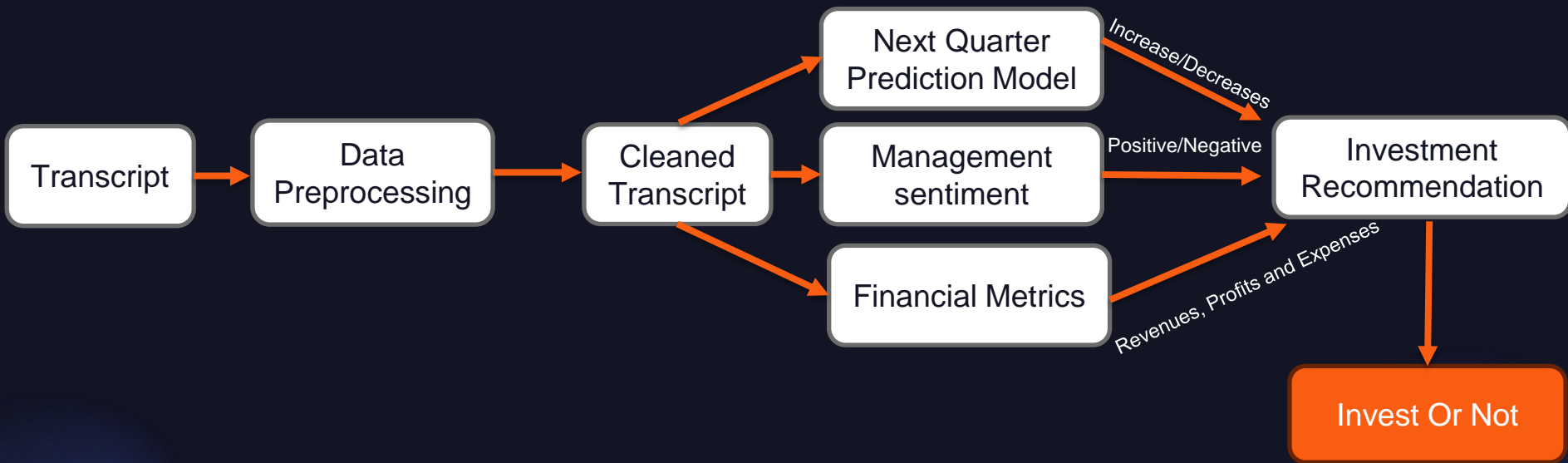
02
Data Preprocessing

Step-1

Splitting the transcript into three sections:

- Transcript content section which contains talk about the quarter earnings and it's status this part is the important in our work since

- The Q&A part which contains the talk between some Analysts, Executive officers, Financial officer and so on , the questions mainly about the company markets and Status for this quarter.

- The final part of the transcript is the speakers during the call and there titles.

Transcript Content
Q&A
Speakers

Step-2

Abbreviations Replacements

	Abbreviations	Meaning
0	Q1	first quarter
1	Q2	second quarter
2	Q3	third quarter
3	Q4	fourth quarter
4	q1	first quarter
5	q2	second quarter
6	q3	third quarter
7	q4	fourth quarter
8	FY	fiscal year
9	YoY	year over year
10	MoM	month over month
11	EBITDA	earnings before interest, taxes, depreciation,...
12	ROI	return on investment
13	EPS	earnings per share
14	P/E	price-to-earnings
15	DCF	discounted cash flow
16	CAGR	compound annual growth rate
17	GDP	gross domestic product
18	CFO	chief financial officer
19	GAAP	Generally Accepted Accounting Principles
20	SEC	U.S. Securities and Exchange Commission
21	IPO	initial public offering
22	M&A	mergers and acquisitions

|	Abbreviations	Meaning
22	M&A	mergers and acquisitions
23	EBIT	earnings before interest and taxes
24	IRR	internal rate of return
25	ROA	return on assets
26	ROE	return on equity
27	NAV	net asset value
28	PE ratio	price-to-earnings ratio
29	EPS growth	earnings per share growth
30	Fiscal Year	financial year
31	CAPEX	capital expenditures
32	APR	annual percentage rate
33	P&L	profit and loss
34	NPM	net profit margin
35	EBT	earnings before taxes
36	EBITDAR	earnings before interest, taxes, depreciation,...
37	PAT	profit after tax
38	COGS	cost of goods sold
39	EBTIDA	earnings before taxes, interest, depreciation,...
40	E&Y	Ernst & Young
41	B2B	business to business
42	B2C	business to consumer
43	LIFO	last in, first out
44	FIFO	first in, first out
45	FCF	free cash flow

Step-3   Remove Stopwords

Adjusted gross margin expanded 370 basis points to 48.9%, primarily driven by favorable business mix,
and higher service and software margin.
The favorable year-on-year impact from China tariffs was offset by $11 million of incremental premium freight charges.
Adjusted operating expenses as a percentage of sales improved 280 basis points.
We have been accelerating high return investments in the business, while prudently managing discretionary costs.'

Len:453

Remove stopwords

Adjusted gross margin expanded 370 basis points 48.9 % ,
primarily driven favorable business mix , higher service software margin .
The favorable year-on-year impact China tariffs offset $ 11 million incremental premium freight charges .
Adjusted operating expenses percentage sales improved 280 basis points .
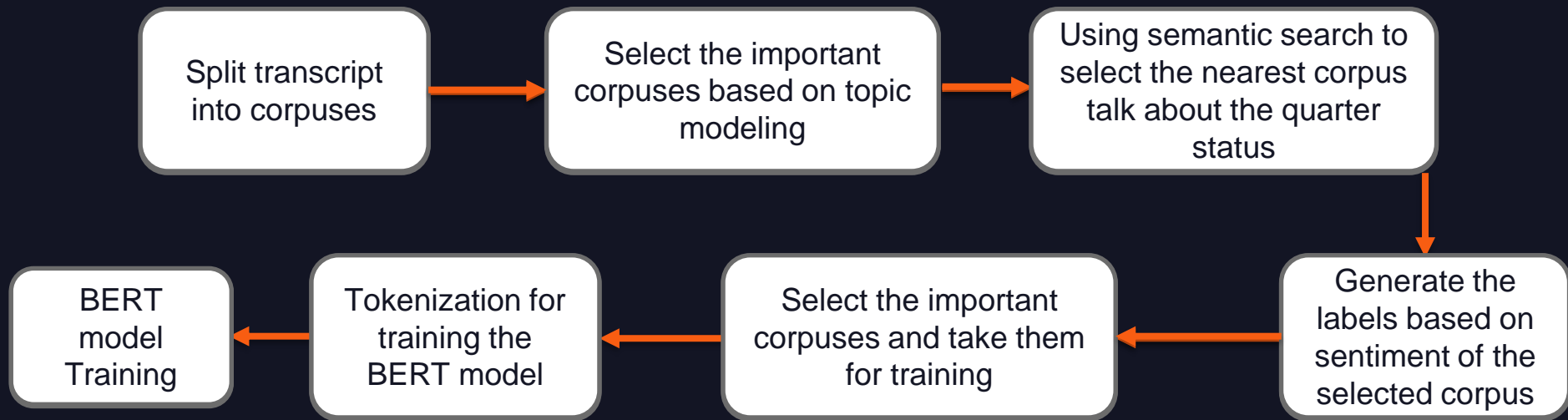We accelerating high return investments business , prudently managing discretionary costs

Len:402
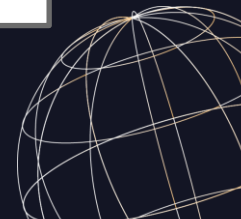
03

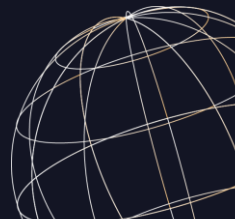# Next Quarter Status
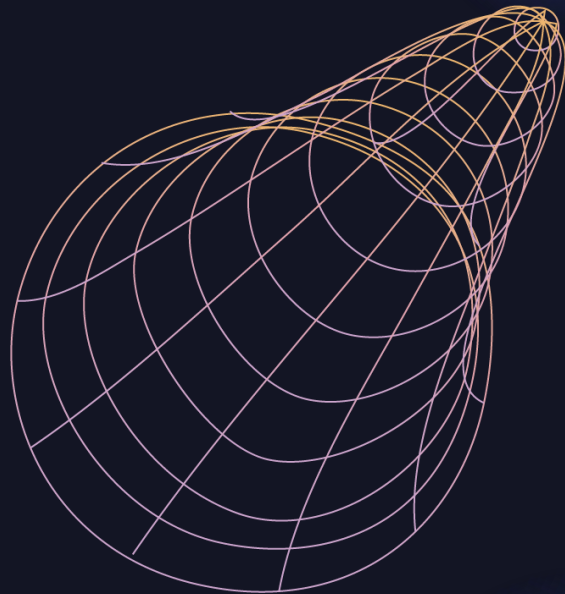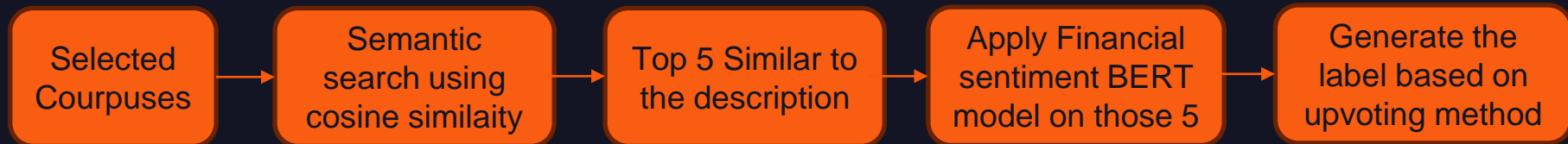
Model preparation Steps

Selecting Important corpus

- The process done In two steps the first one Is splitting the transcript into chunks of 256 word size this leads as to not overcome the BERT max token length.
- Then we have choose topic modeling which is based on BERT which is fine tuned on the Twitter Financial news topic this model can detect the following classes I have take some of them to be the corpus will continue training with them.
- **Selected Topics are** :

      Stock Movement ,Earnings,IPO,Stock Commentary,Currencies,M&A |
        Investments,Financials,Macro,Analyst Update,Company | Product News

- **Unselected Topics are**:

   Legal | Regulation,Personnel Change,Treasuries | Corporate Debt,Dividend,Gold |
Metals | Materials,Markets,Fed | Central Banks,Politics,Energy | Oil,General News |
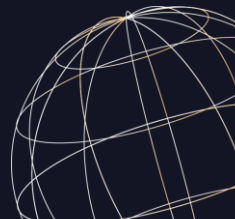   Opinion
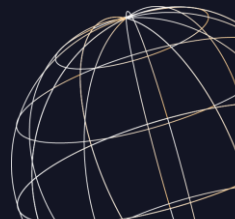
Labeling

- The process done In two steps applying the semantic search using this description
Description: "the second quarter results of the 2020 year"
This description is compared by the corpus that generated from topic modeling then compared with them using cosine Similarity after the embedding have been generated.

- From that process I have choose the 5 top similar corpuses for this description then pass it to the sentiment model

- Then apply sentiment model based on BERT that is Fine tuned on Financial data which tell us if this talk is BULLISH or BERISH or neutral then we can got the label based on the Up voting

| Selected Courpuses | → | Semantic search using cosine similaity | → | Top 5 Similar to the description | → | Apply Financial sentiment BERT model on those 5 | → | Generate the label based on upvoting method |

Select corpus for training

Here I apply the same semantic search method to get the top 3 corpus as the training data since passing to BERT this large text will not working since the BERT max tokens length is 512 that is leads an error when , try to training .using the same description I have used above to catch the similar corpuses so can train on small amount of text that related to the label .
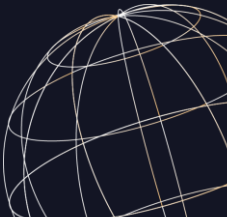
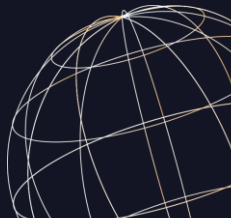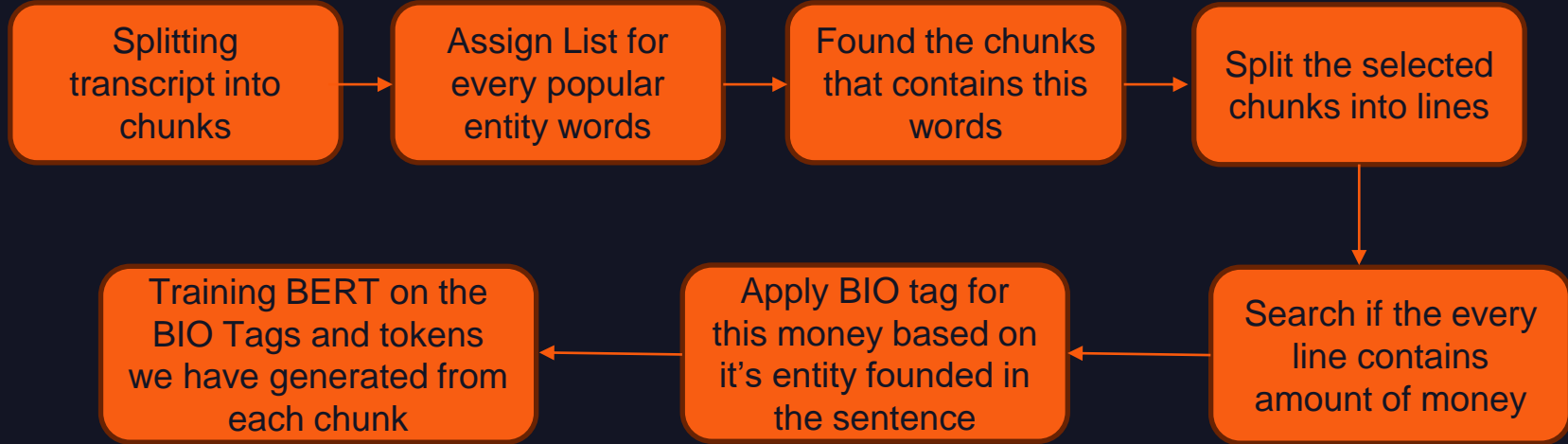# Finance NER BERT
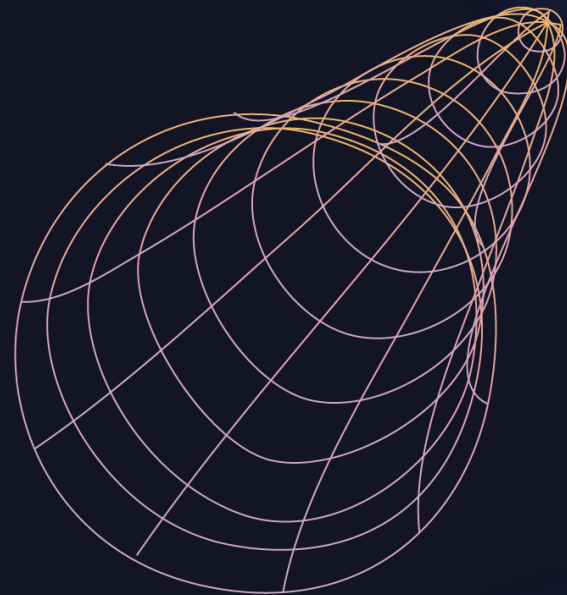
Labeling

Problem:
The problem how to found the amount of money that related to which entity in this large text

Solution:
What if we searching for words related to our entity like revenue as example we need to find words like :"revenue", "sales", "income", "earnings", "gross profit", "net profit","top-line growth" , I have used ChatGPT to generate for me an list of words that may be appears in the text then I can catch the sentences that contains this words and search for the any money in them and assign them for the entity class
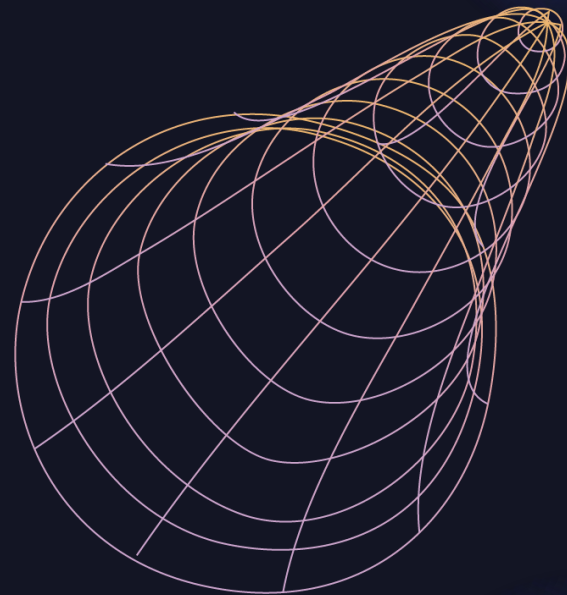
Labeling

Splitting transcript into chunks → Assign List for every popular entity words → Found the chunks that contains this words → Split the selected chunks into lines

Training BERT on the BIO Tags and tokens we have generated from each chunk ← Apply BIO tag for this money based on it's entity founded in the sentence ← Search if the every line contains amount of money
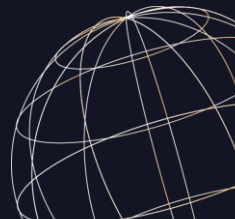
05

Sentiment Analysis

Processing and Labeling

- Splitting the transcript into small chunks for training bert
- Using TEXTBLOB for sentiment labeling for every chunk
- Training BERT for based on this data
- Inference: splitting the transcript into chunks and generate label for each chunk and get the up voting as the sentiment for the hole transcript.
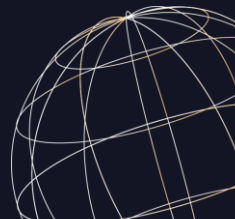
06

Pipeline & Deployment
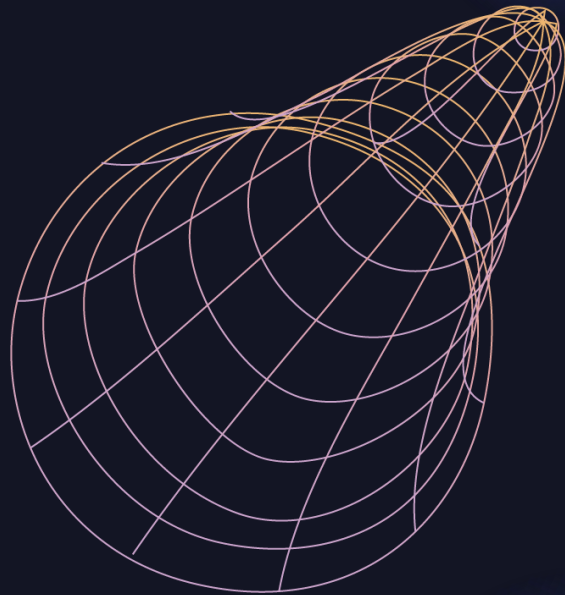
Pipeline

- The transcript pass to the three models after cleaning process
- The Next Quarter model say If the next quarter stock price will Increase or decrease based on the sentiment for the part that talk about transcript quarter.
- By using Upvoting method like I mentioned before we can generate the sentiment of the management.
- Last the NER model generate the list of values that founded in the transcript for the four entities profits , expenses and revenues .
- For how we select the choice of investment based on the amount of profit and revenues and the two models outputs
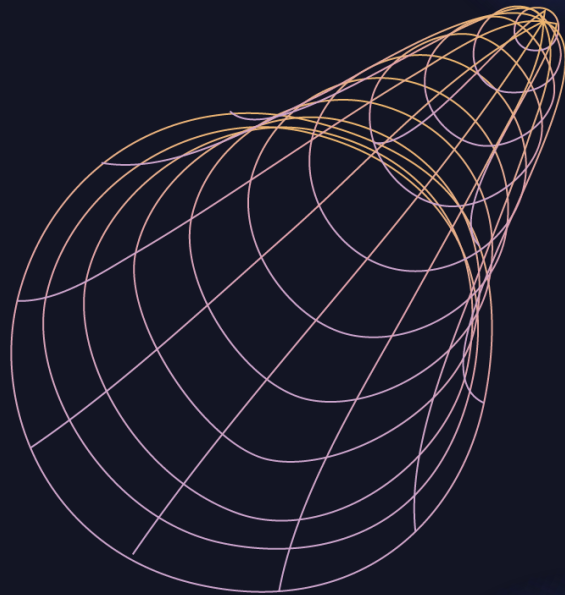
Deployment

- I have using hugging face spaces for deployment which you got an free CPU with 2 cores and 16GB memory that is so enough for our task as DEMO.

07

Models Performances

Models Performances

Next Quarter Model:

This model result not so good since there is no talks about the next quarter enough in our data to can label this text part from the transcript the best way to handle it is good human labeling for this task which takes to much time.

NER Model:

The model performance is looks so good but since there is a lot of variation for every metric name so catches the amount of money seems difficult but we can working on labeling to enhance the performance more and training on large data to got good result.

Management Sentiment model:

The model performance but if we try an larger model like long former I think may lead to better results also for our mode we need to create balanced classed , I have trained on 151325 chunks positive vs 9882 chunks negative
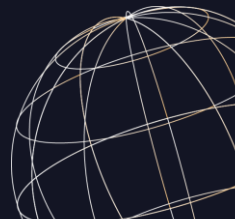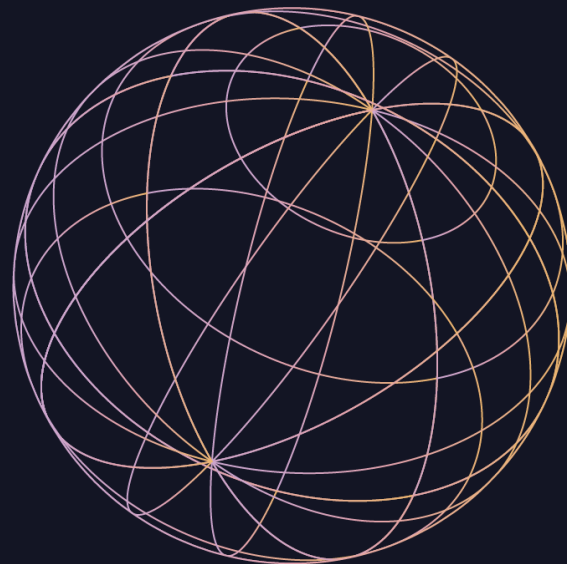
Thanks!

Do you have any questions?
ahmed.taha.7725@gmail.com
+201140069642