Machine Learning in Finance

A talk by The Artificial Intelligence Society(ARIES)-IIT Delhi and The Mathematics Society, IIT Delhi

Speakers

Vaibhav Seth

Student Researcher at LCS2-IITD

Research Intern, Inria

Incoming Quant Intern, Quadeye

Co-Founder, **ARIES**

Vaibhav Gupta

Research Intern, HKU

Global Rank 1, IQC '23

Incoming Quant Intern, WorldQuant

Co-founder, **ARIES**

What is Machine Learning?

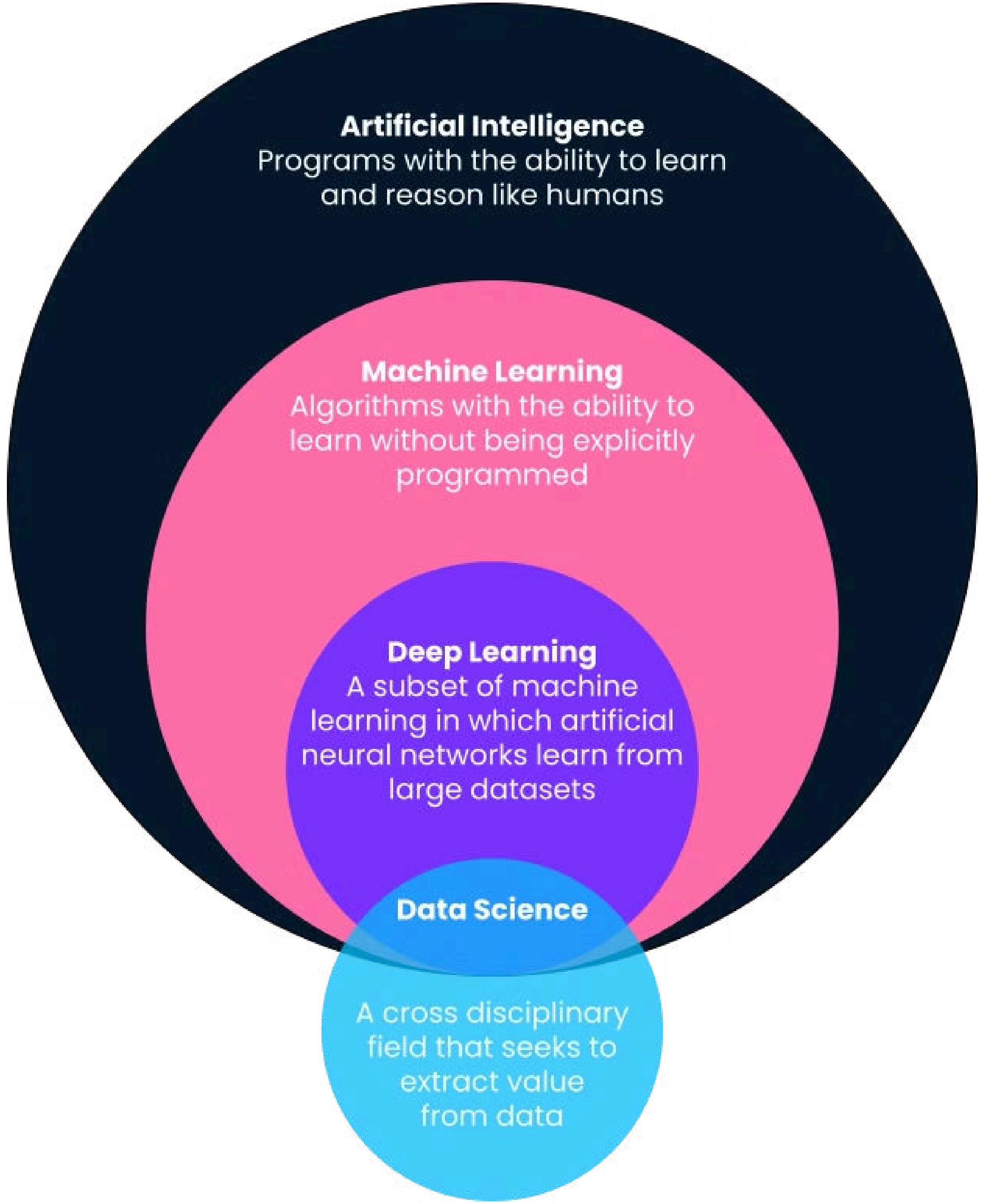
~~Giving a prompt to ChatGPT to generate wonderful results and solve assignments is Machine Learning.~~ It is much more.

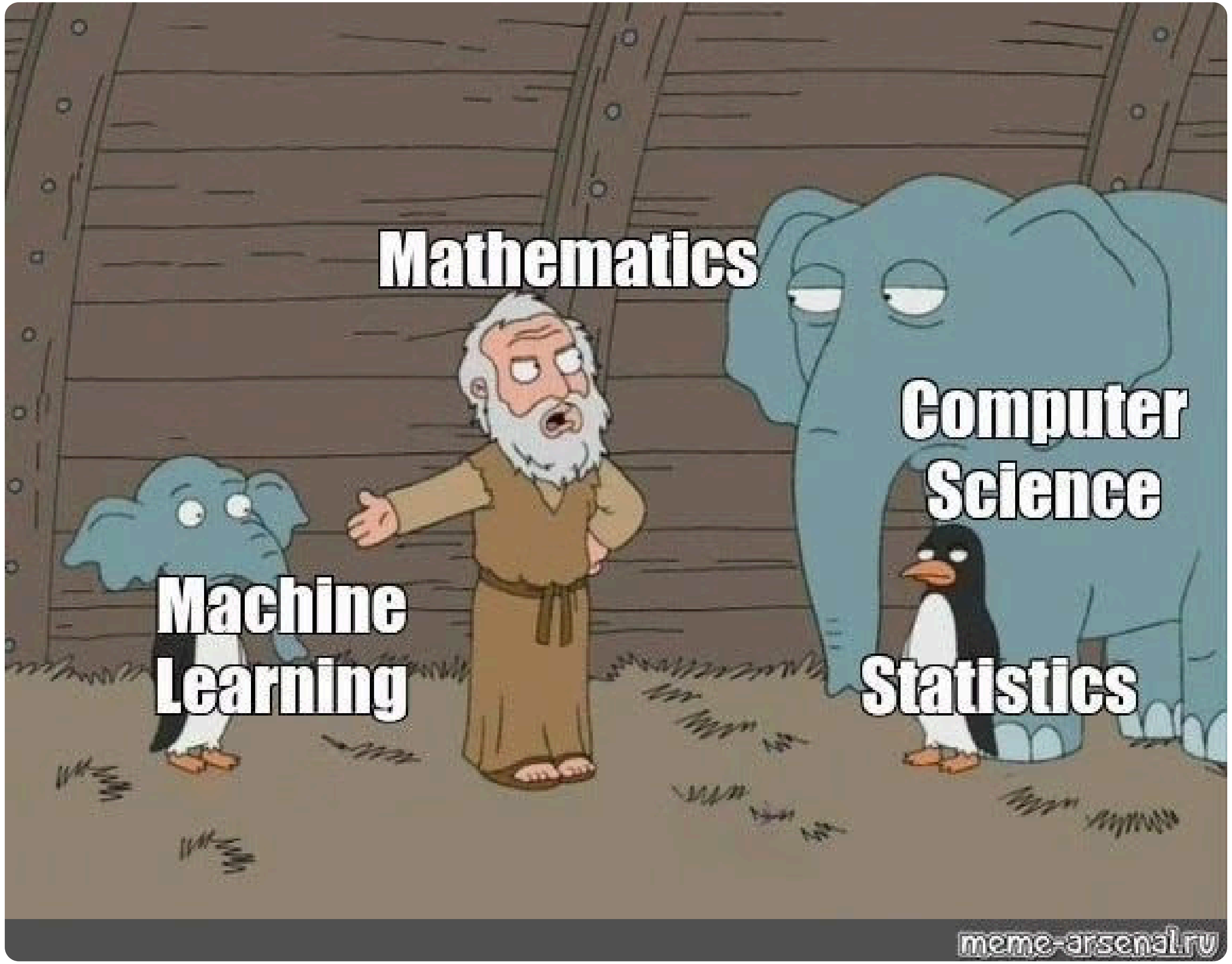
ChatGPT uses Deep Learning, which is a form of Machine Learning, to generate its results.

Machine learning is a **field of study in artificial intelligence** concerned with the **development and study of statistical algorithms** that can **learn from data** and **generalise to unseen data.**

It is a **Data-Driven** approach to solving problems.

With the advent of Deep Learning and large complex datasets, it has also become a computational task.


Artificial Intelligence
Programs with the ability to learn and reason like humans
Machine Learning
Algorithms with the ability to learn without being explicitly programmed
Deep Learning
A subset of machine learning in which artificial neural networks learn from large datasets
Data Science
A cross disciplinary field that seeks to extract value from data


Mathematics
Computer Science
Machine Learning
Statistics
meme-arsenal.ru

A question to the audience

According to you, what kinds of data exists in finance?

Open, Close, High, Low, Volume ?

Financial Datasets are much more Diverse!

The Rise of ML in Finance

- **Tons of data available** and ML is **Data-Driven.** We have multiple data points for each asset, company, exchanges and much more.
- Changes in the **market microstructure**: electronic trading and the integration of global markets
- Development of strategies that focus on mathematical quantities like risk, as opposed to asses classes
- The revolutions in **computing power**, **data generation and management**, **statistical methods,** and the rise of **Deep Learning**
- The **outperformance of** algorithmic trading relative to human, discretionary investors.

- Sufficiently large datasets allow for training of large models without **overfitting.**
- Machine Learning can be used to make various **predictions**, like **prices of assets**, **market movement**, **opening and closing prices** and much more.
- Various parameters for a strategy can be predicted by Analysing Historical data using ML
- **Autonomous Trading Agents** using **Reinforcement Learning**
- **Natural Language Processing** can be used to extract **Market Sentiment** and use that as a trading signal.

Time Series Analysis using RNNs

A study by Two Sigma

Parts of the following slides are adapted from Two Sigma's presentation on ML in Finance

Limit Order Book

- Stocks are mainly traded via electronic exchanges
- Electronic markets seek to match buyers and sellers via **limit order book.**
- Investors submit **limit orders** at the prices they are willing to buy or sell

Price ($)	9.98	9.99	10.00	10.01	10.02
Number of Orders	100	200	0	800	500

Table: Example of buy and sell limit orders in an order book.

Limit Order Book: Transactions

- The lowest price for a sell order is the **best ask price.**
- The highest price for a buy order is the **best bid price.**
- An investor can immediately buy/sell a stock via a **market order**, which is executed at the best ask/bid price

Price ($)	9.98	9.99	10.00	10.01	10.02
No. of Orders	100	200−100	0	800	500

↓

Price ($)	9.98	9.99	10.00	10.01	10.02
No. of Orders	100	100	0	800	500

Limit Order Book: New Submission

- The order book changes over time as new orders are submitted, cancelled, and executed.
- An investor can submit a new limit order at any time.

Price ($)	9.98	9.99	10.00	10.01	10.02
No. of Orders	100	200	0 +100	800	500

↓

Price ($)	9.98	9.99	10.00	10.01	10.02
No. of Orders	100	200	100	800	500

Machine Learning using Order Book Data

- The order book data represents the visible **supply and demand** for a stock, and it can be used to model future price changes.
- The **order flow** generates a high-dimensional data sequence which can be used to train Models.

$$X_t = (....., O_t^{B_t - nc}, ...O_t^{B_t}, O_t^{A_t}, ...O_t^{A_t + nc}, ...)$$

- B represents the bid price and Bt is the best bid price at time t. Similarly A is defined for ask. And O is used to represent the number of orders.
- We get data points for each time step. This type of data is know as Time Series Data.

A caveat

- Even an accuracy of greater than 50% does not guarantee a profitable trading strategy exists.
- The buy-sell spread of the asset can result in our strategy failing.

Price ($)	9.98	9.99	10.00	10.01	10.02
No. of Orders	100	1,000	0	200 −100	100

↓

Price ($)	9.98	9.99	10.00	10.01	10.02
No. of Orders	100	1,000	100 −100	100	100

- Prediction: Price will rise, so buy at 10.01. Price does rise but the best bid price is still less than the original selling price. Total profit = -1

The Dataset Used

- 3 years of event-by-event data for 1000 NASDAQ stocks
- t1, t2, t3,…. are the timestamps at which prices change
- A model is build to predict the change in price between two time periods given past data

$$P[Price_{t_{k+1}} - Price_{t_k} > 0 | OrderBook_{T<t_k}]$$

- Training happens across multiple GPUs

Recurrent Neural Networks

- They are a type of neural networks and are used for processing sequential data like natural language and time series data.
- They differ from conventional neural networks by having a feature of Memory built in them
- The memory state is a mathematical representation of the data that the neural network has seen

$$M_t = f(X_0, X_1, ...X_t)$$

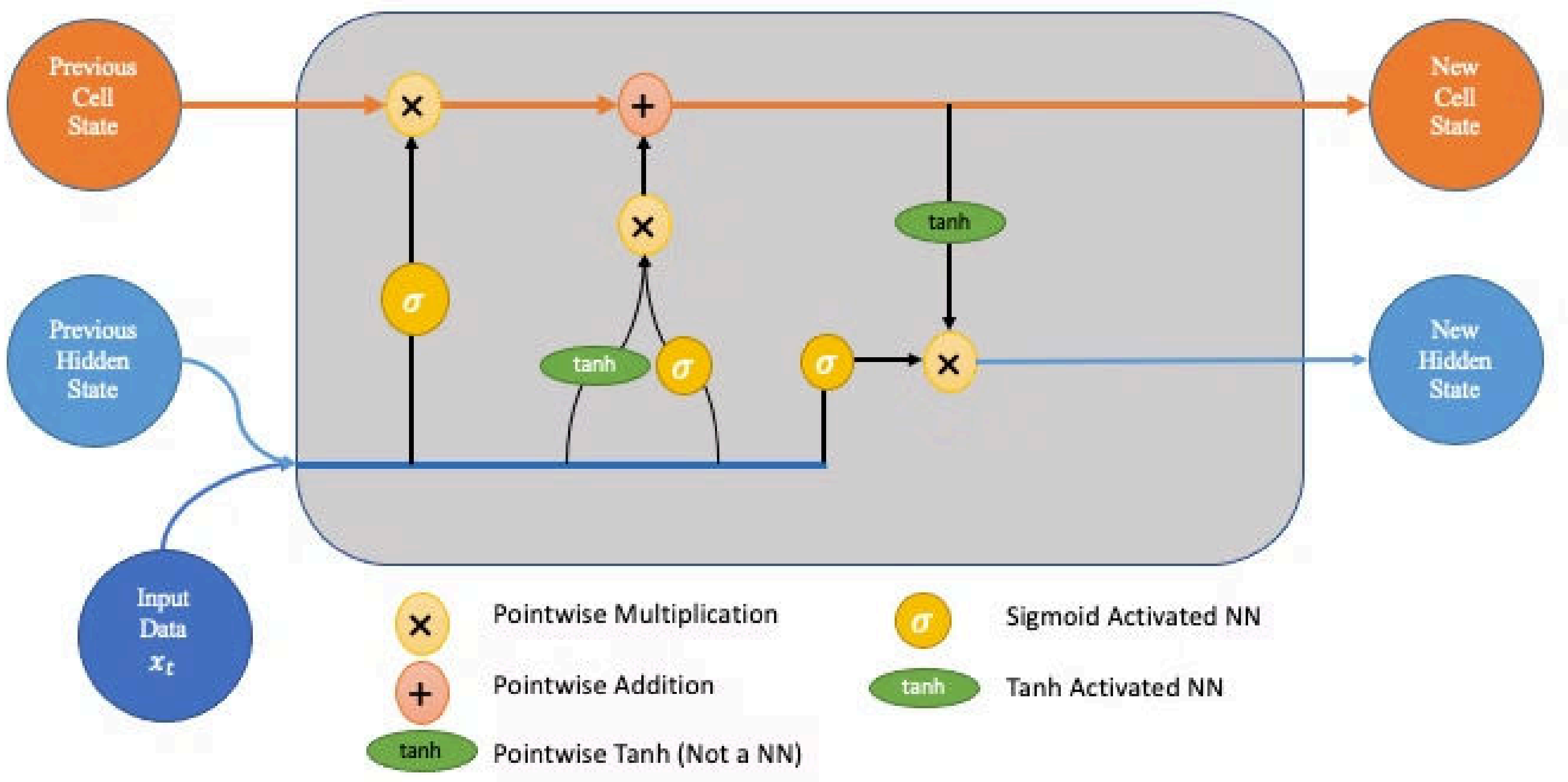
- The processing of input is done in a sequential manner. At every time step we see the input data (Xt) at that time, and the memory state Mt.
- The model now makes a prediction based on the memory state and model parameters.

$$Y_t = g(M_t; \theta)$$

- The model is now trained to minimise the difference between predictions and the actual known outcomes

LSTMs

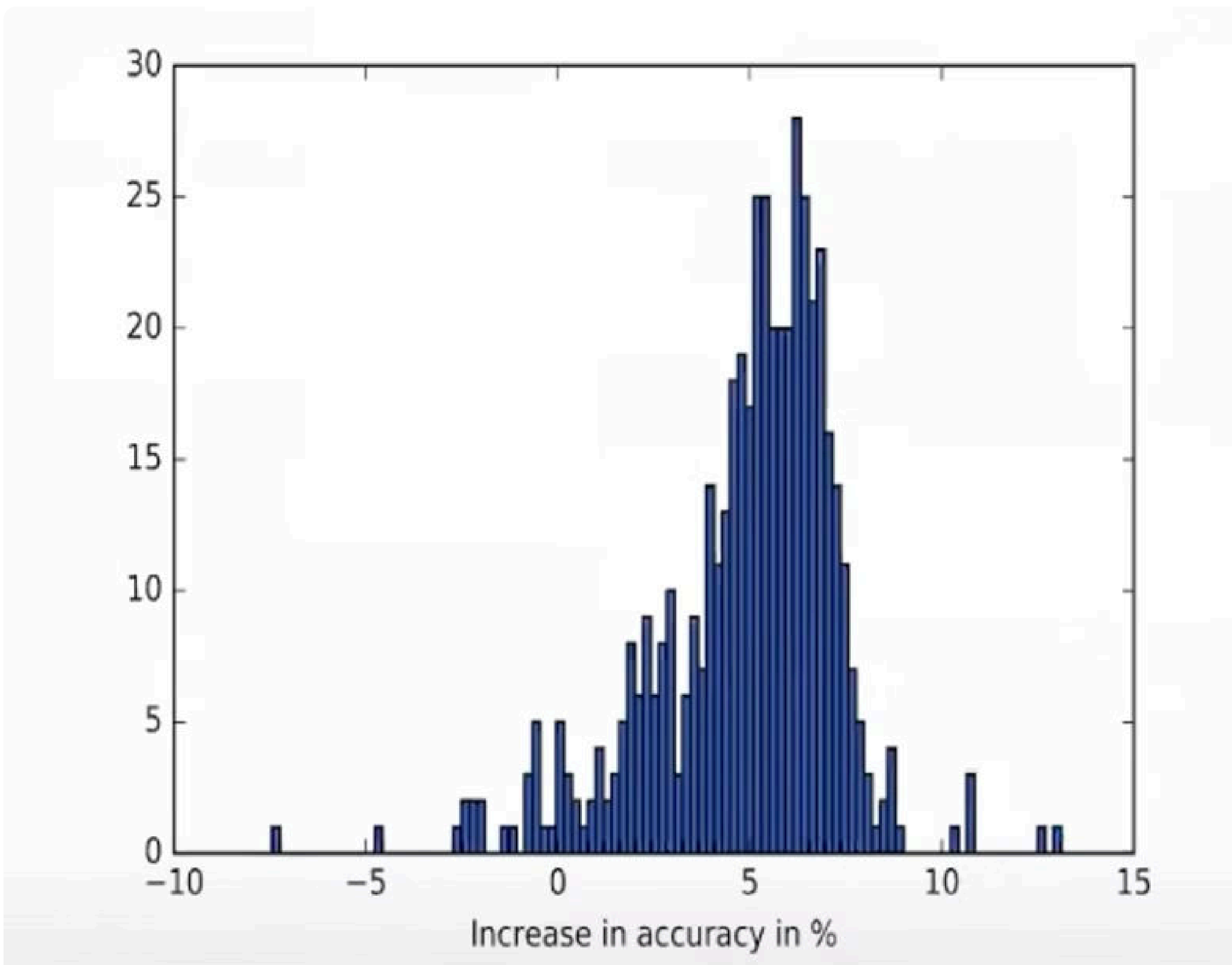
- LSTMs are a more advanced version of RNNs, that deal with some major flaws in RNNs.
- Their main idea is to selectively control the reading and writing of the memory state
- LSTMs have input and forgetting functionalities by which the neural network can better understand what information should be stored by them and what should be forgotten.



Model Evaluation

- The models are evaluated on multiple stocks as follows

$$A_i = \frac{\text{no. of times predicted price change is correct}}{TotalPriceChanges} * 100$$

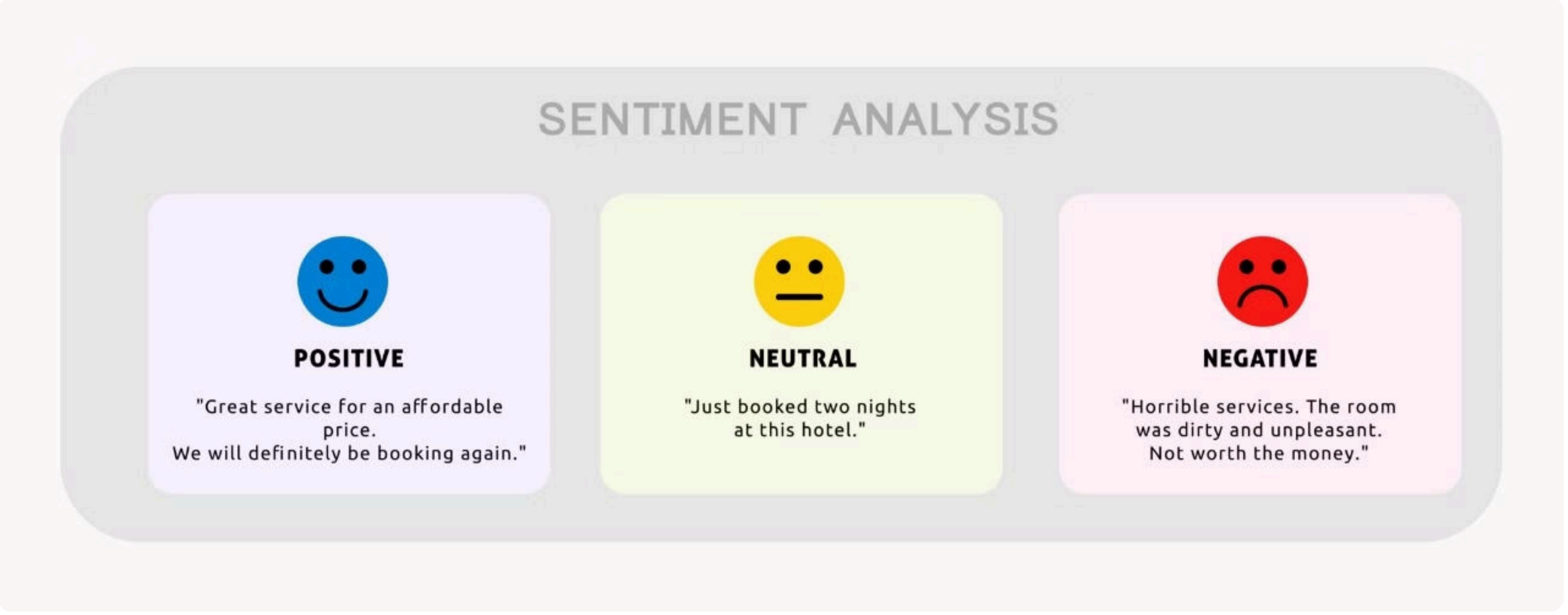


- It was found that LSTM on average performs much better than than other linear models (as shown by increase in accuracy)

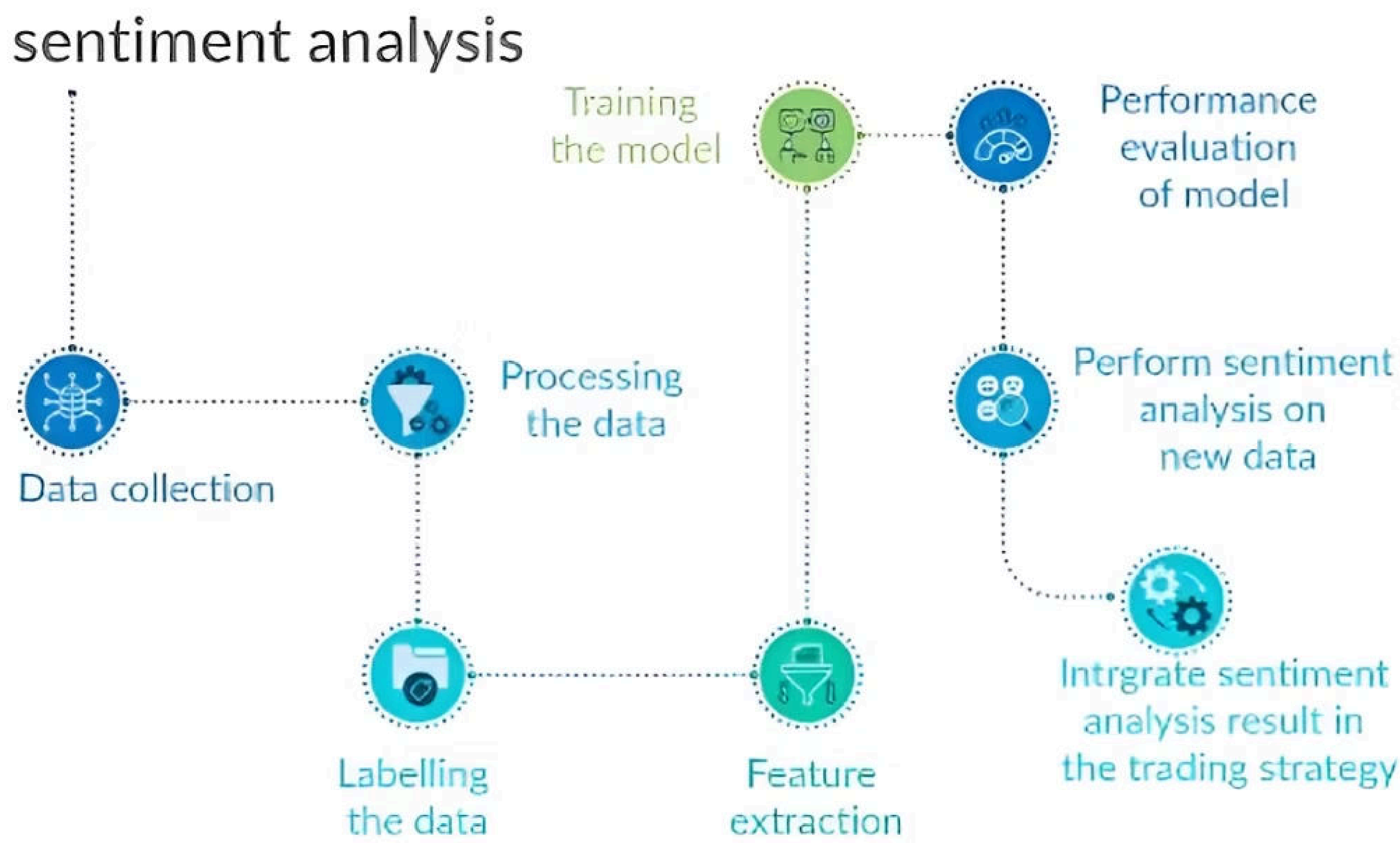
Sentimental Analysis of Stock Market using NLP

Data Analysis: NLP helps analyze news articles, social media, and financial reports to gauge market sentiment towards stocks.

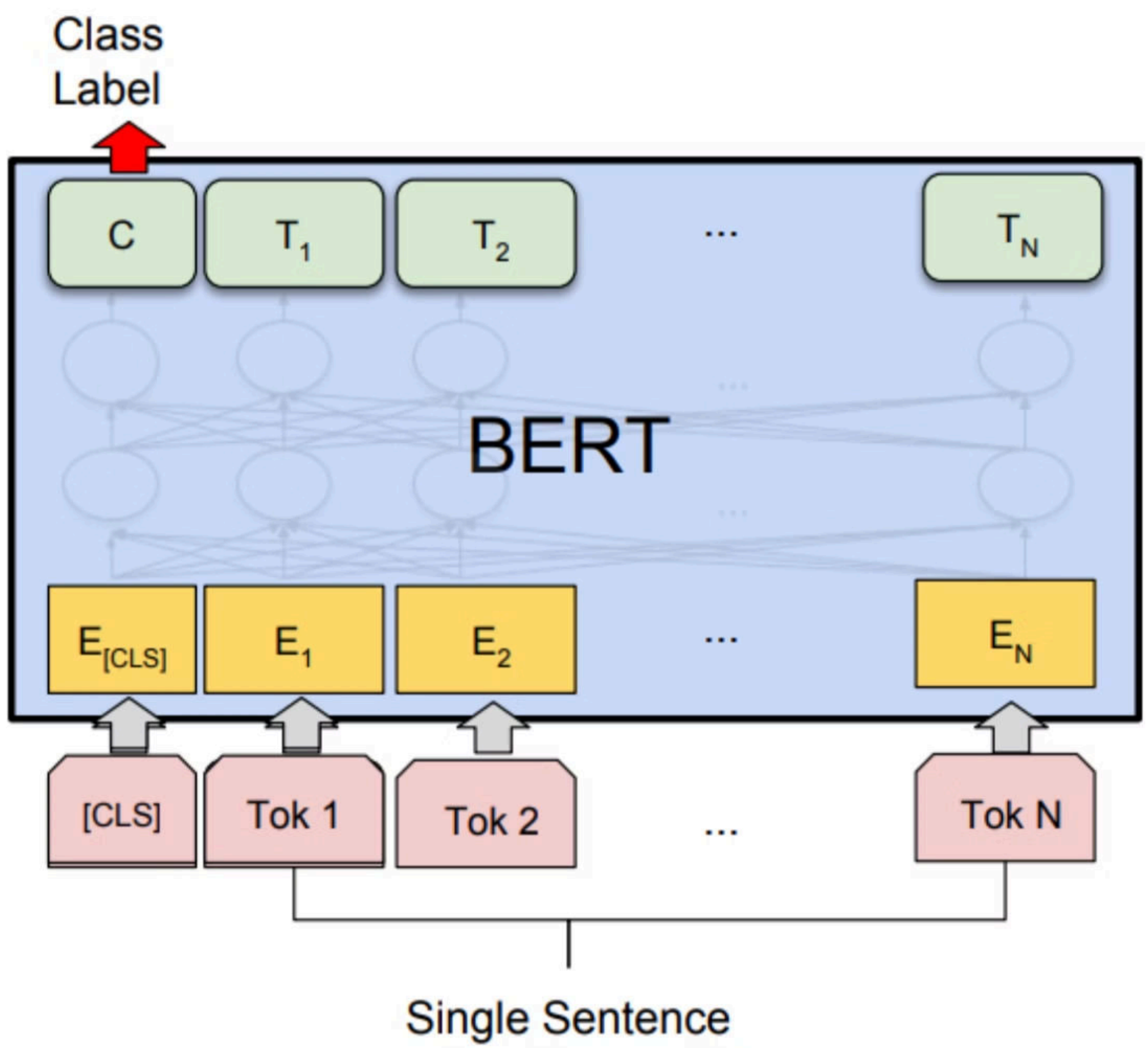
Insightful Decisions: By leveraging NLP techniques, investors can make informed trading decisions based on sentiment analysis.



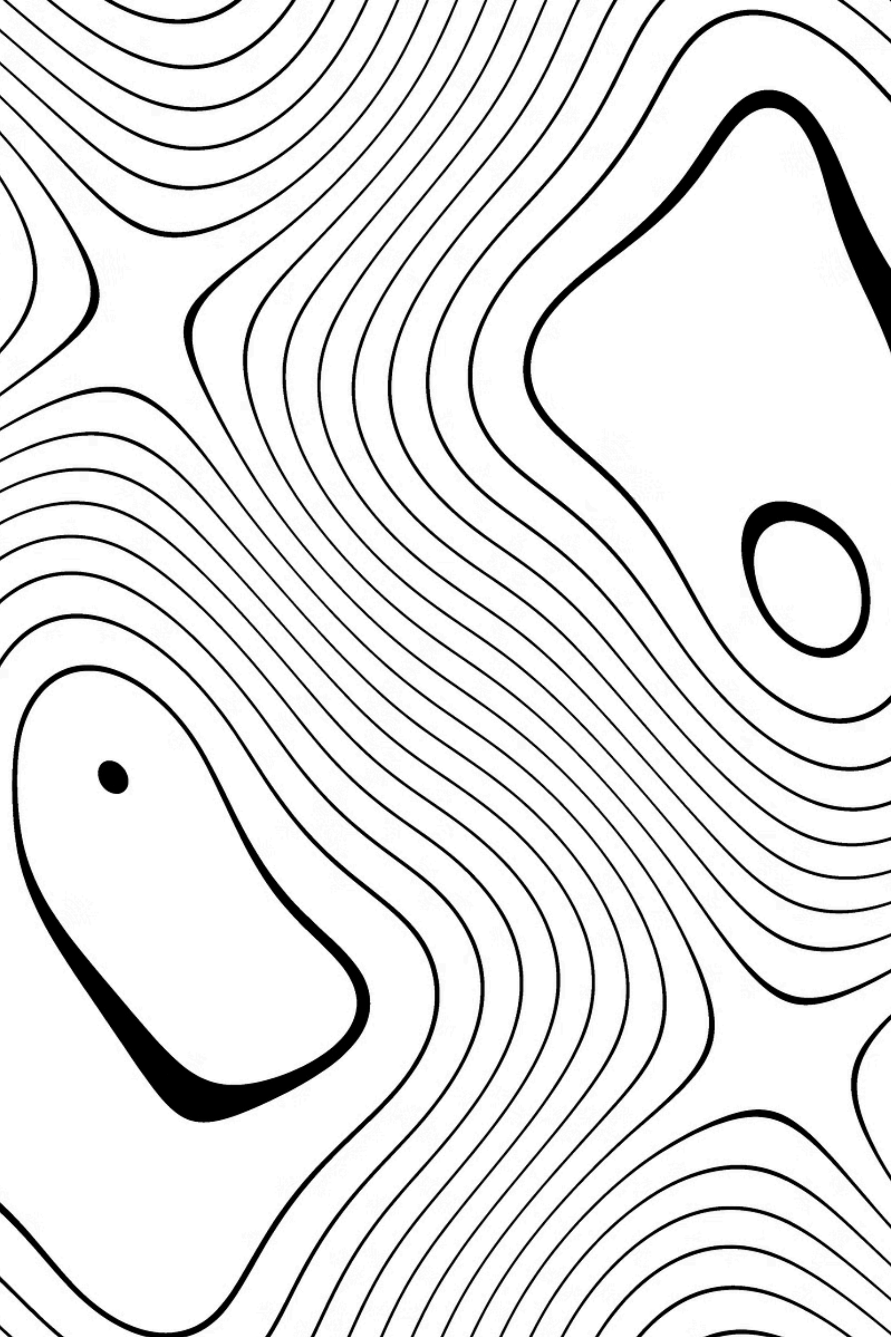
The Sentiment Analysis Framework



Sentiment Analysis using BERT



- We can assign a sentence to belong to the 'positive', 'negative' or 'neutral' class. We can also get numerical scores for these classes and use them in trading strategies.



A trading strategy using sentiment data, and back testing

Sentiment Strategy (Deployed)

SOURCE OF IDEA:

Fluctuations in Coin Currencies like Doge, Shiba Inu, etc.

Elon Musk Twitter Handle

HYPOTHESIS:

A **Spike** in Social media Activity about the stock indicates an upcoming strong price movement.

FIRST IMPLEMENTATION:

Social volume > yesterday's Social Volume then, entering a position based on the intensity of sentiment in the market.

snt_social_volume : Social Media Volume of sentiment

snt_value : Mood of sentiment

snt_buzz : Intensity of Sentiment

Journey toward more Signal Accuracy:

Detecting a Spike

The first spike in social volume gave a good enough entry point in our hypothesis.

We assumed a 40% sudden rise in volume as the spike.

KEY POINTS IN FINAL IMPLEMENTATION:

To make it more Robust:

We set a moderate threshold value of sentiment.
Mean of past x days instead of the ts_delay, value for sentiment data.

Limitations of ML in finance

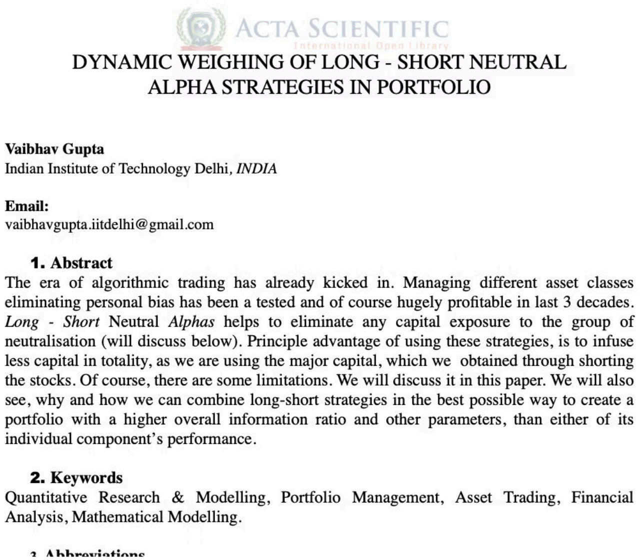
- ML models heavily rely on vast and diverse datasets to capture meaningful patterns and make accurate predictions.
- In scenarios where the available dataset is limited, ML may not outperform traditional models .
- In situations where interpretability and explainability of the model's decision-making process are crucial, traditional models that provide more transparency and interpretability may be preferred over ML models.

Therefore, while ML is a valuable tool in finance, its effectiveness depends on the availability of large datasets and the specific requirements of the financial scenario.

Quick Brain Teaser:

If we run two strategies of 25% annual returns each, what will be the net portfolio annual returns?

Managing Portfolio with Multiple Models



ACTA SCIENTIFIC
International Open Library

DYNAMIC WEIGHING OF LONG - SHORT NEUTRAL ALPHA STRATEGIES IN PORTFOLIO

Vaibhav Gupta
Indian Institute of Technology Delhi, *INDIA*

Email:
vaibhavgupta.iitdelhi@gmail.com

1. Abstract

The era of algorithmic trading has already kicked in. Managing different asset classes eliminating personal bias has been a tested and of course hugely profitable in last 3 decades. *Long - Short* Neutral *Alphas* helps to eliminate any capital exposure to the group of neutralisation (will discuss below). Principle advantage of using these strategies, is to infuse less capital in totality, as we are using the major capital, which we obtained through shorting the stocks. Of course, there are some limitations. We will discuss it in this paper. We will also see, why and how we can combine long-short strategies in the best possible way to create a portfolio with a higher overall information ratio and other parameters, than either of its individual component's performance.

2. Keywords

Quantitative Research & Modelling, Portfolio Management, Asset Trading, Financial Analysis, Mathematical Modelling.

How to merge two or more strategies?

Is it altogether necessary?

Follow up Question:

Merging Two Correlated vs Two Un-Correlated Models. Which is more profitable?

What are LONG-SHORT Neutral Strategies?

Lets take an example:

S.No.	Stock Name	Weights in Neutral Strategy I	Weights in Neutral Strategy II	Net Weight
1	A	0.7	-0.4	0.3
2	B	0.3	-0.6	-0.3
3	C	-1	1	0
Book Size	(Say in millions $)	2*	2*	0.6*

Let's understand with another example of same kind.

S.No.	Stock Name	Weights in Neutral Strategy I	Weights in Neutral Strategy II	Net Weight
1	A	0.7	0.6	1.3
2	B	0.3	0.4	0.7
3	C	-1	-1	-2
Book Size	(Say in millions $)	2*	2*	4*

Conclusion

In the session we covered:

- What is ML?
- Datasets used in finance
- Time series analysis
- Sentiment Analysis
- Trading Strategy and Backtesting
- Diverse Portfolios

Scan the QR for asking questions on slido

Thank You!

Thank you for joining us today. We hope you found the session on Machine Learning in Finance informative and engaging. If you have any further questions or would like to continue the discussion, please feel free to reach out to us. Stay tuned for more exciting topics in the future!