



# Amazon.com stock market movement forecasting using Logistic Regression

Andres Gherzi Sayan (A20539425)



# Amazon.com Inc.

- Founded in 1994, went public in 1997
- 5th biggest company in the world, by market cap
- Cloud Services and e-Commerce
- Why choose this company?



# Scope of this Project

## What we are doing

- Forecasting the stock movement based on historical data
- Preprocessing of data
- Comparing logistic regression with LSTM
- Statistical tests and measurements

## What we are NOT doing

- Not forecasting the exact value of stock, only the ups and downs
- Results are not necessarily better than other methods

# Objectives and Results

- 1: Define the data - Dataset, constraints and correlations
- 2: Preprocess the dataset - For logistic model and encoding for classification
- 3: Design a Logistic regression with memory
- 4: Implement a LSTM RNN
- 5: Compare both models using Accuracy and Loss

Objectives 1 and 2 for MidTerm presentation.

Objectives 3 to 5 for Final presentation (30 November)

# Methods and Tools

- Python
- TensorFlow, Keras
- Accuracy tests: Accuracy, loss and MSE
- Compute resources: GoogleColab

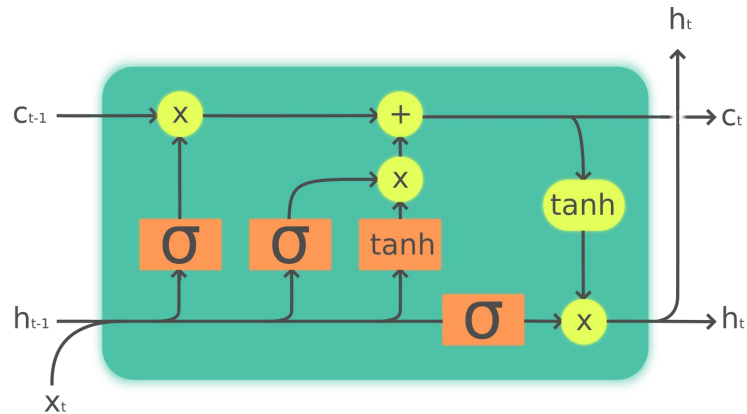
The logo for Google Colab, featuring the word "colab" in a lowercase, rounded, orange font.

TensorFlow

# Theoretical Framework

- Why a RNN model?
- What is the Vanishing Gradient problem?
- How to take advantage of them?
- Which dataset are we using? [Kaggle: Amazon stock 1999 - 2022](#)
- Time series forecasting or time series analysis?

# LSTM Cell



Legend:

Layer



Pointwise op



Copy



# Feature Augmentation and Selection

## Feature Augmentation

- Mean Average: The average of the last N periods. And then drawing a line with those values to visualize it.
- We will use MA for 50 and 200 days as new features.

## Feature Selection

- The dataset consist of 7 columns: date, open, close, high, low, volume and adjusted close.
- We will remove the adjusted close and the volume to simplify the analysis.

## Normalization

- It is crucial for time series models to use normalization

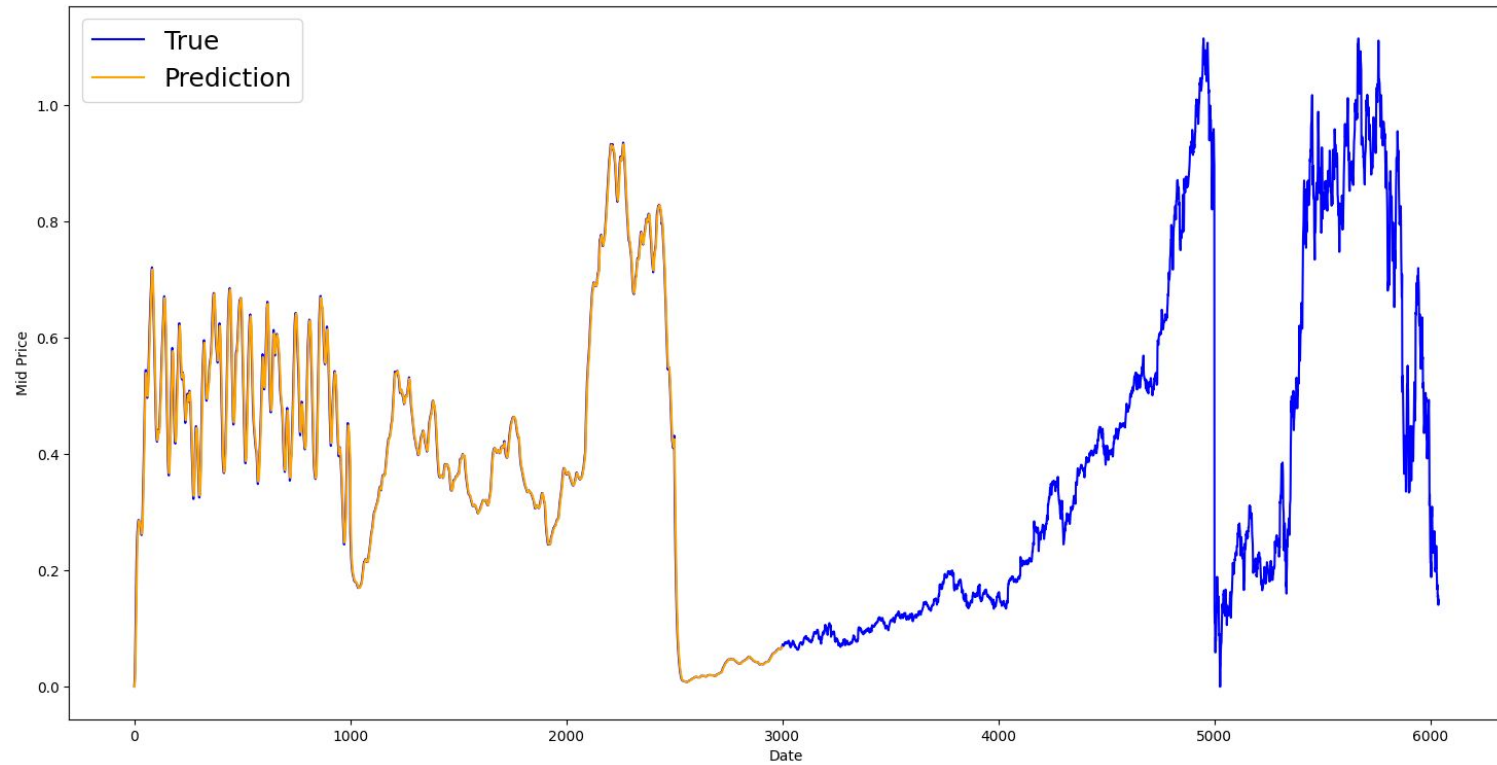


# Results

- Google Colab with own model (logistic regression) and benchmark from literature (LSTM):  
<https://github.com/aghersisayan/MachineLearningIIT>

## Conclusions:

- LSTM behaves better, with a MSE of less than 0.01
- Logistic regression achieves an accuracy of 50%
- While a more complex model is better at predicting movement, simple models can also be used.
- Datasets are the most important part of ML models



# References

- [Google Colab](#)
- [How LSTM networks solve the problem of vanishing gradients | by Nir Arbel | DataDrivenInvestor](#)
- [Vanishing Gradient Problem in RNNs | by Sagar Patil | Medium](#)
- [The Exploding and Vanishing Gradients Problem in Time Series | by Dr Barak Or | Towards Data Science](#)
- [A Technical Guide on RNN/LSTM/GRU for Stock Price Prediction | by Chris Kuo/Dr. Dataman | The Startup | Medium](#)
- [Amazon \(AMZN\) - Market capitalization](#)
- [Forecasting: theory and practice - ScienceDirect](#)
- [When should I use an RNN LSTM and when to use ARIMA for a time series forecasting problem? What is the relation between them? - Quora](#)
- [MAE, MSE, RMSE, and F1 score in Time Series Forecasting | by Ottavio Calzone | Medium](#)
- <https://www.kaggle.com/code/sriharshaeedala/amazon-stock-price-prediction-with-arima-98>
- <https://www.kaggle.com/datasets/sriharshaeedala/amazon-stock-price-from-1999-to-2022>
- [Recurrent neural network - Wikipedia](#)
- [How to Develop LSTM Models for Time Series Forecasting - MachineLearningMastery.com](#)
- [Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras - MachineLearningMastery.com](#)
- [A Quick Deep Learning Recipe: Time Series Forecasting with Keras in Python | by Yang Lyla | Towards Data Science](#)
- [Machine Learning to Predict Stock Prices | by Roshan Adusumilli | Towards Data Science](#)
- [Understanding LSTM Networks -- colah's blog](#)