



Real-time streaming application with Kafka

**Predicting stocks markets
using global Index**

Adrien & Elia's Presentation





Project Description

1	PROBLEM	Predict the market stock of Google, Amazon and Facebook using global index of EU, US and Asia
2	DATA COLLECTION	Collected data from yahoo finance of the different stock and indexes
3	KAFKA	Connection with Kafka Server, Configuration of producer and consumer, extraction and publication of closing price
4	MODELS	Feature engineering and Batch models implementation
5	CONCLUSION	Project overview and results

PROJECT TIMELINE

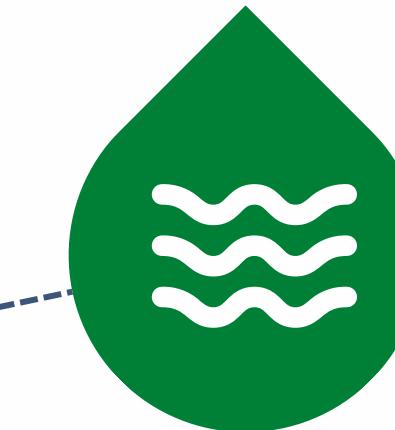
1. BATCH VS ONLINE

Overview and comparison between batch and online models for stock market prediction



3. Batch Models

Preparation of the batch models with feature engineering and presentation of the different models results



5. Results and Conclusion

Exposition of the results, discussions and Q&A



2. Data Collection with KAFKA

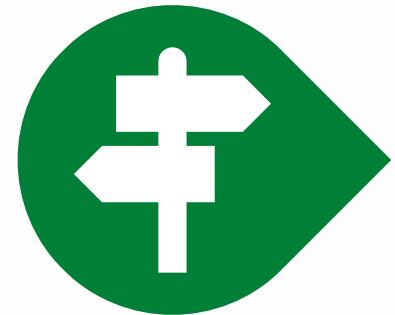
Use of Kafka for global index and stock data stream collection



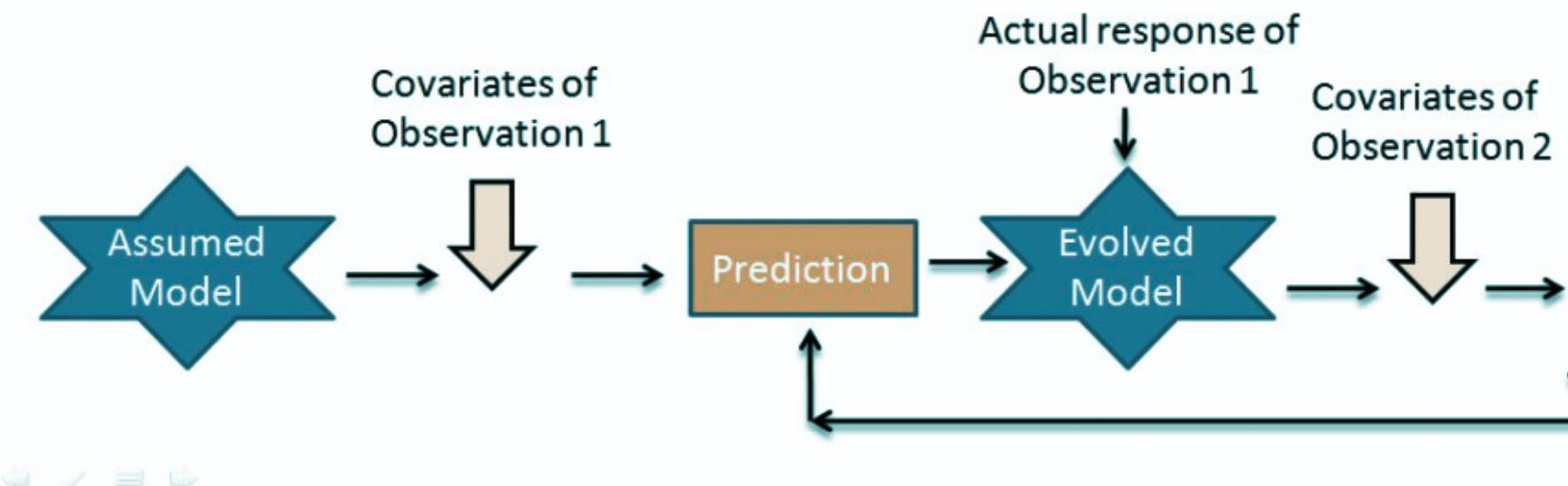
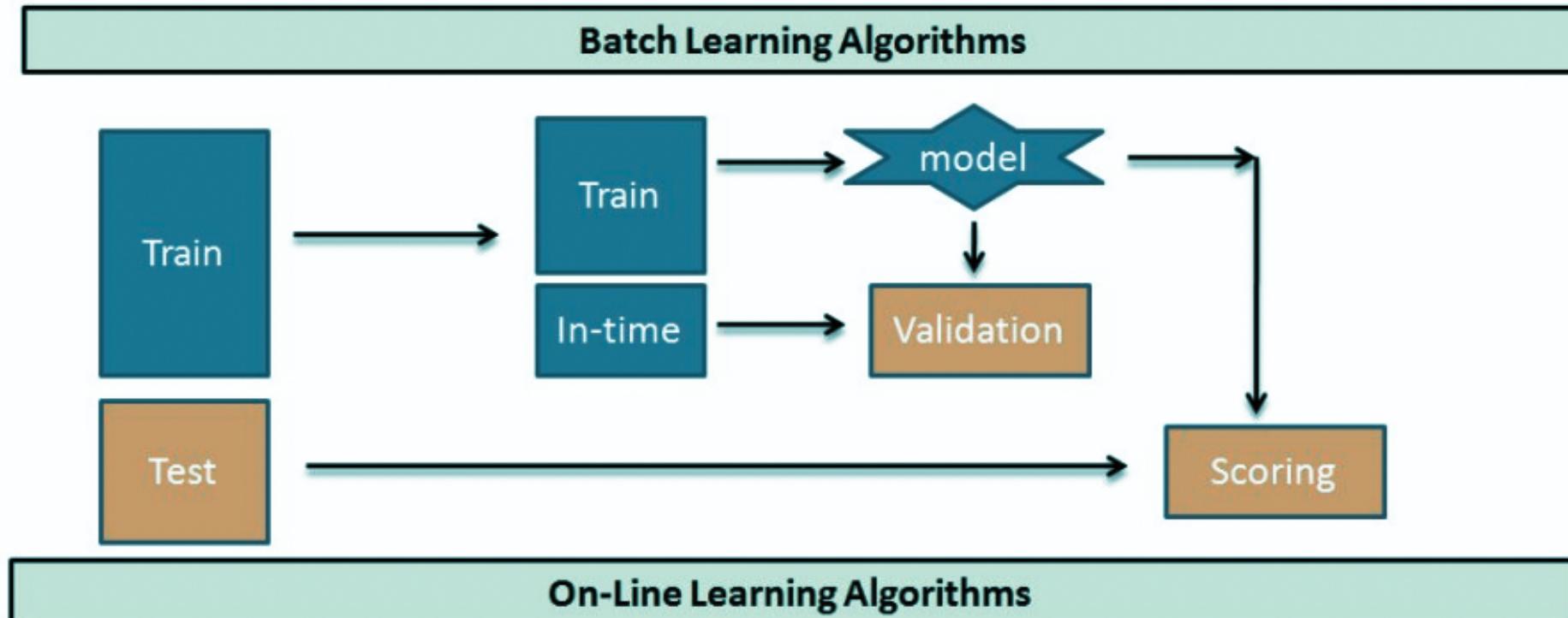
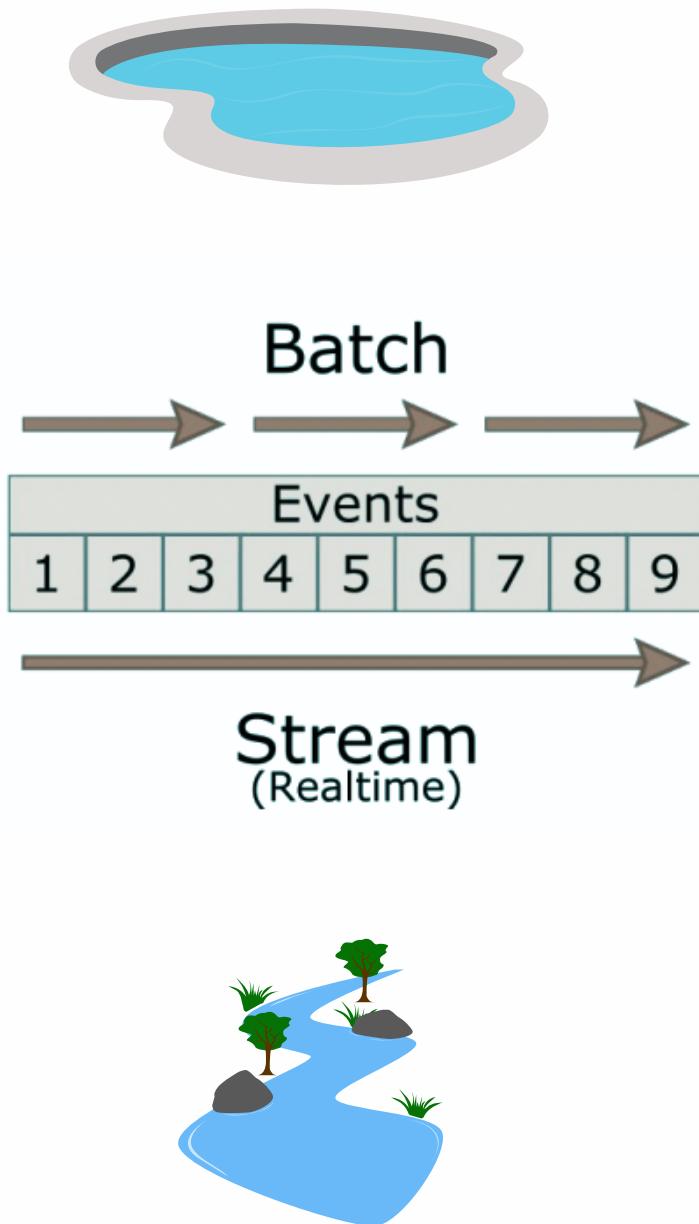
4. Code Demo

Code demonstration





Batch VS Online



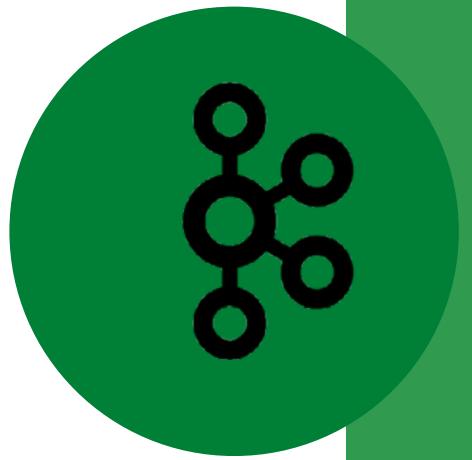


Data Collection



YahooFinance Data:

- Amazon, Google and Facebook daily close price for 3 years
- Data of indexes: S&P500 (US), CAC40 (EU) and Nikkei (Asia)
- Data of stocks: GOOGL, META, AMZN



Kafka Integration

- Requests YahooFinance to download the datasets
- Data preparation
- Kafka Topics for each stock (target to predict)
- One producer and one consumer to manage them all



Producer

Configuration

- Connection to a local Kafka broker (localhost:9092)
- JSON serializer to convert messages into JSON strings

Download

- Uses yfinance library to download historical data for stocks and indices
- 3 years data range

Publish

- For each timestamp in the data, iterates over stock symbols
- Extraction of closing price of the stock and index prices and publication to the appropriate topic



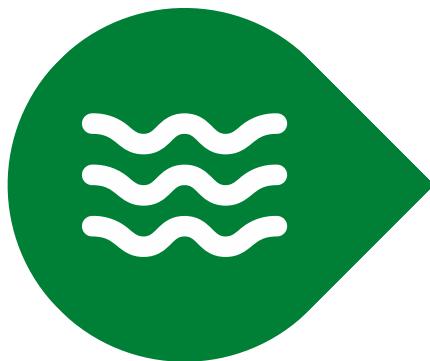
Consumer

Configuration

- Connection to a local Kafka broker (localhost:9092)
- Subscription to topics of stocks

Streaming

- Extraction of streaming data from the broker
- Saves to separate CSV files for each stock as datasets
- Contains the columns : timestamp, symbol, stock price, and index prices (S&P500, CAC40, Nikkei).



Batch models

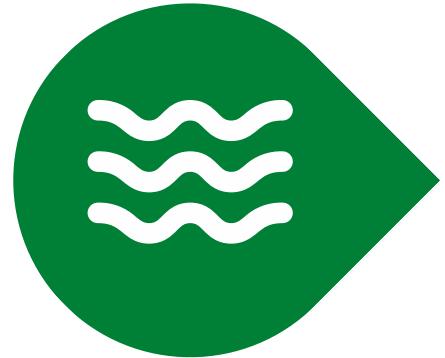
Models



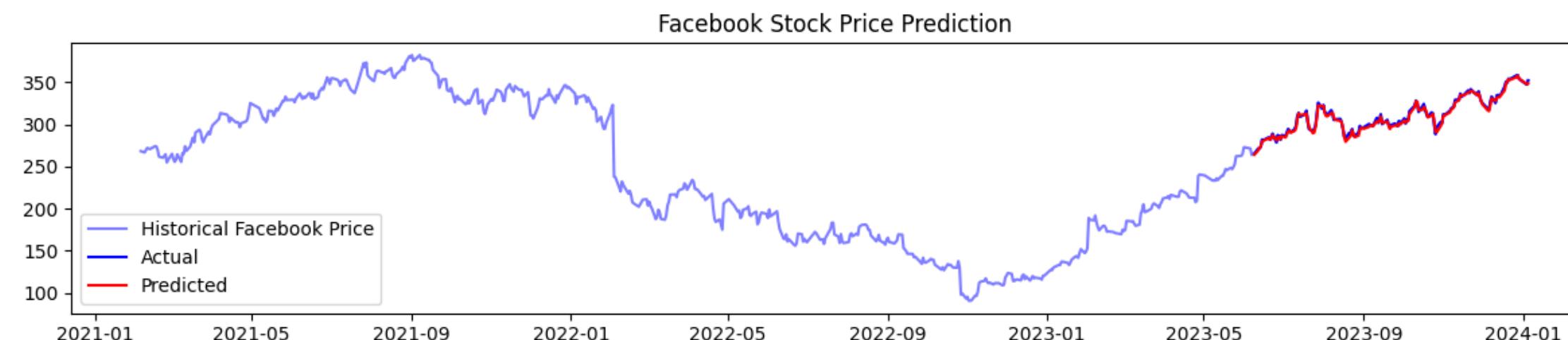
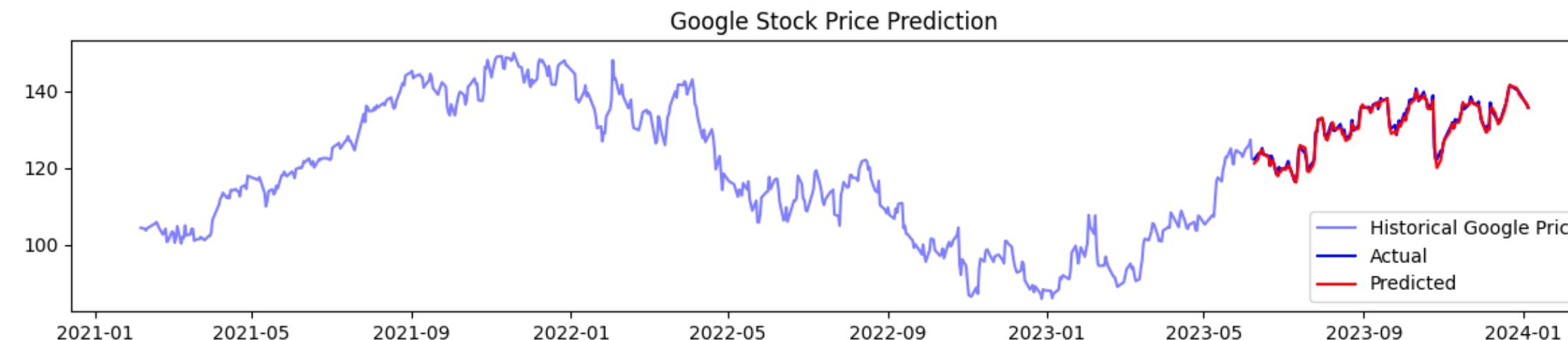
Feature Engineering

Feature Engineering applied to the stock markets

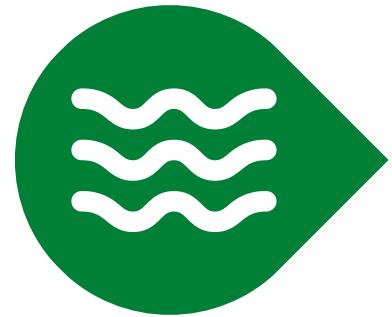
- **Lagged Features:**
 - Shift the time series data by one or more time periods, creating 'lags'
 - Useful in identifying trends or patterns over time
- **Rolling Window Statistics:**
 - Rolling Mean / Standard Deviation:
 - Smoothen the data and helps identified trends / Measures volatility
- **Relative Strength Index (RSI):**
 - Momentum oscillator measuring speed and change of price movements
 - Calculated with a 14 time-step window
- **Moving Averages :**
 - Simple Moving Average (SMA) and Exponential Moving Average (EMA)
 - Averages the closing prices over a specific period, smoothing data prices
 - Window lengths: 5, 10, and 20 time steps



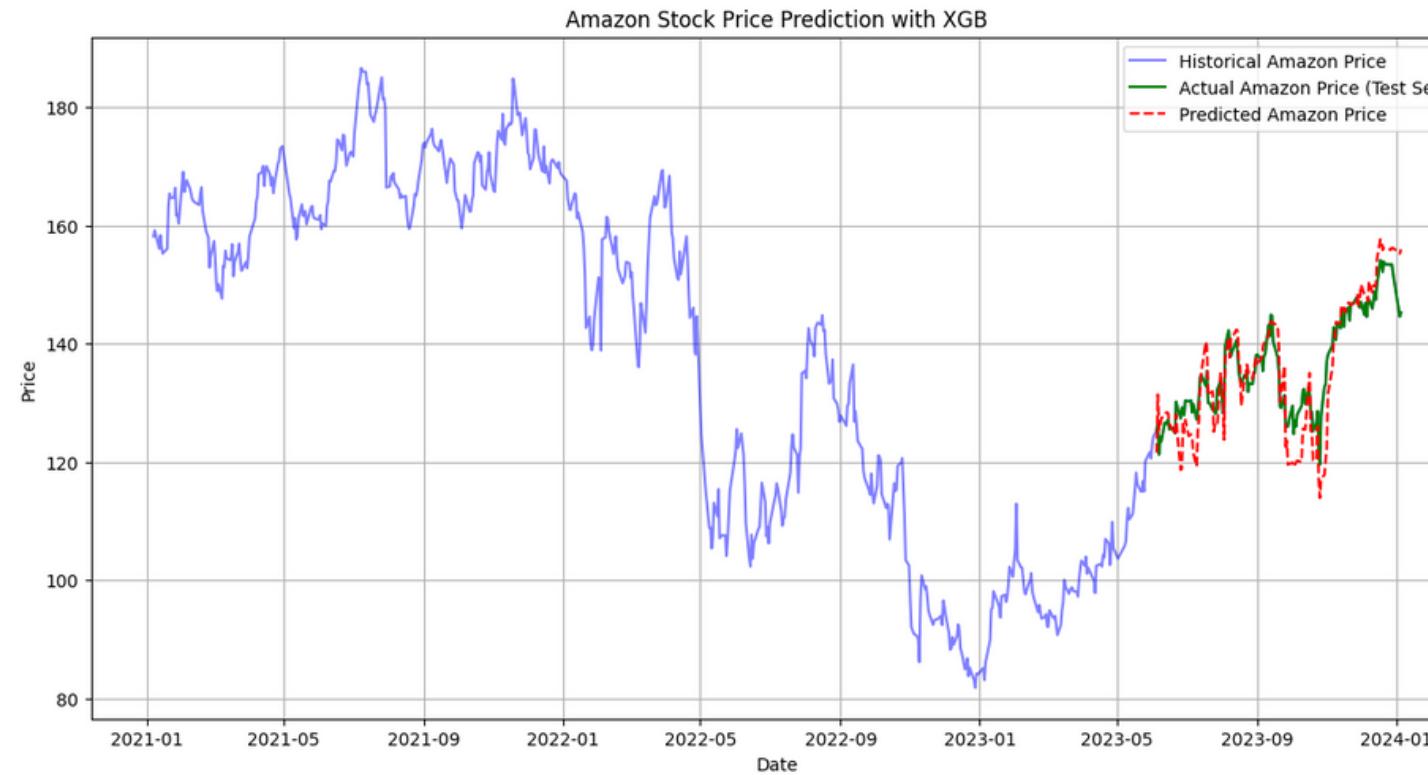
Linear Regression



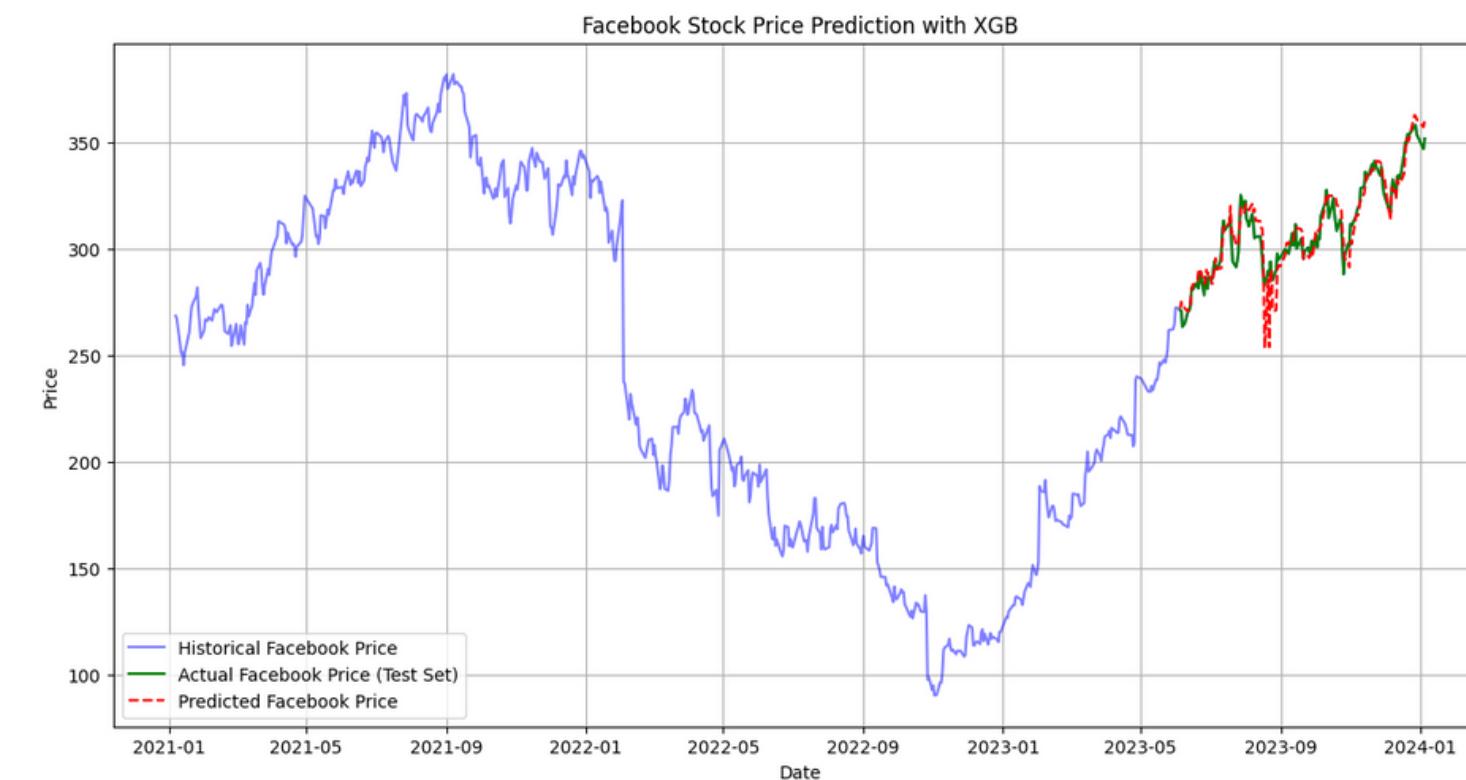
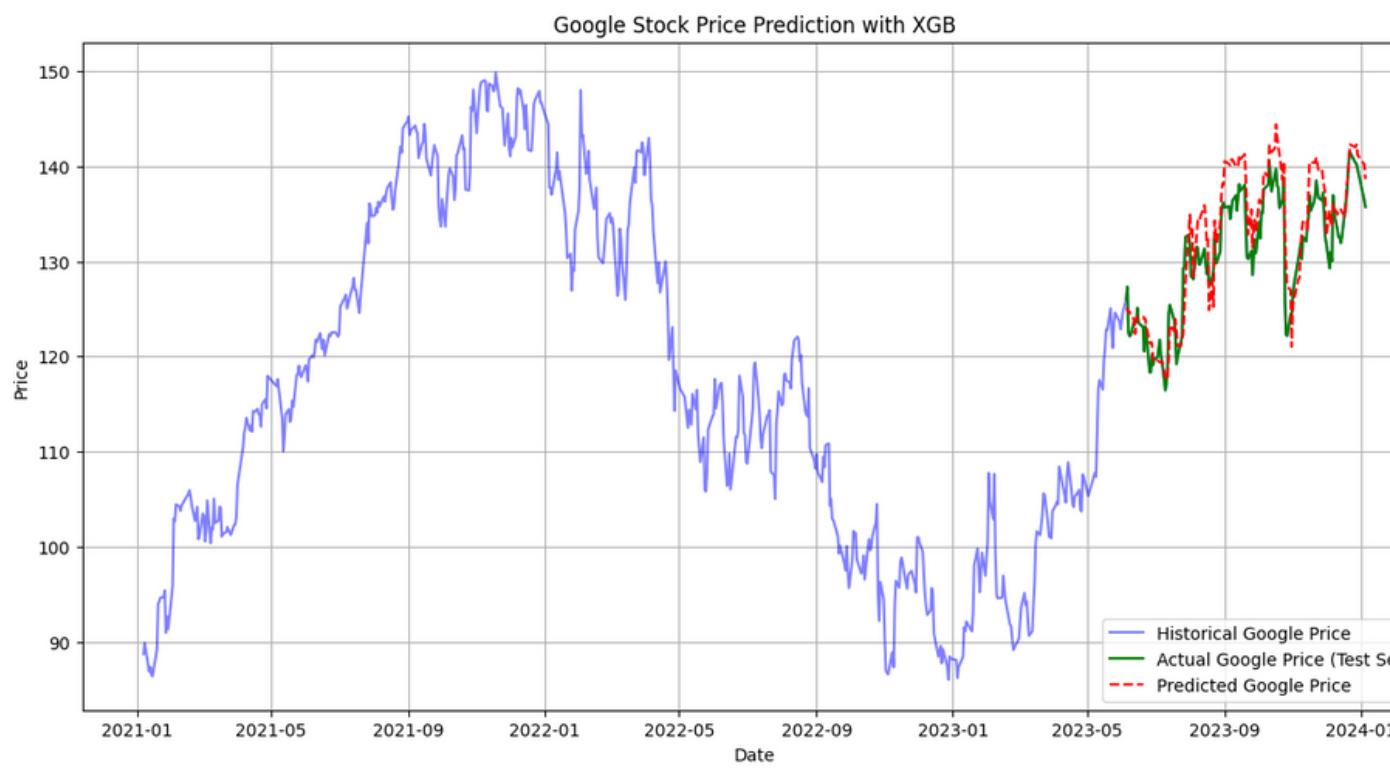
	Stock	MSE	R^2
0	Google	0.663349	0.983301
1	Facebook	4.696356	0.988772
2	Amazon	0.842700	0.988057

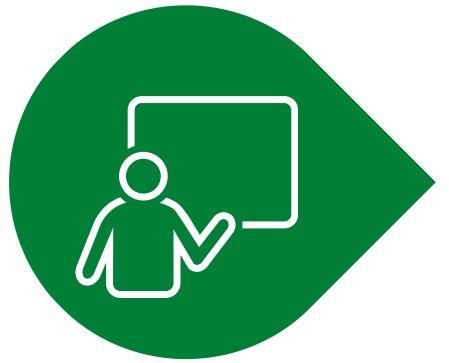


XGboost



Company	MSE	R-squared
0 Google	9.616639	0.758456
1 Facebook	49.455982	0.891694
2 Amazon	20.847378	0.712248





Code Demo



Conclusion



- Integration of Kafka consumer and producer
- Feature engineering: Indicators calculation
- Models implementation (Regression and XGboost)



- Implementation of Online Models for comparison
- Hyper-parameter tuning

