Stock Insight from Global News Sentiments

Hirak Bandyopadhyay
Kunal Mozumdar
Rishabh Bhardwaj
Yaman Sanghavi

The Erdős Institute

December 2, 2024

Motivation

Objective : How various global events like pandemic or election results can affect stock prices?

- Predict stock prices/trends based on sentiment analysis of financial and geopolitical news.

Business Value :

- Optimize trading strategies based by forecasting stock prices accurately.
- Aids in more informed trading strategies for investors and financial institutions.

Benefactors : Financial Analysts, Investment Firms , Individual Investors, Short to Mid Term Investors.

KPIs :

- Correlations between stock prices and market sentiments.
- Identify relevant stock features for accurate stock price/trend predictions
- Predictions accuracy of future stocks prices.

Data Gathering

Stock Data

- Yahoo Finance
- Stocks : Top 5 Tech stocks (20% S & P valuation)
  - Google (GOOG)
  - Apple (APPL)
  - Microsoft (MSFT)
  - NVidia (NVDA)
  - Amazon (AMZN)

Stock News Data

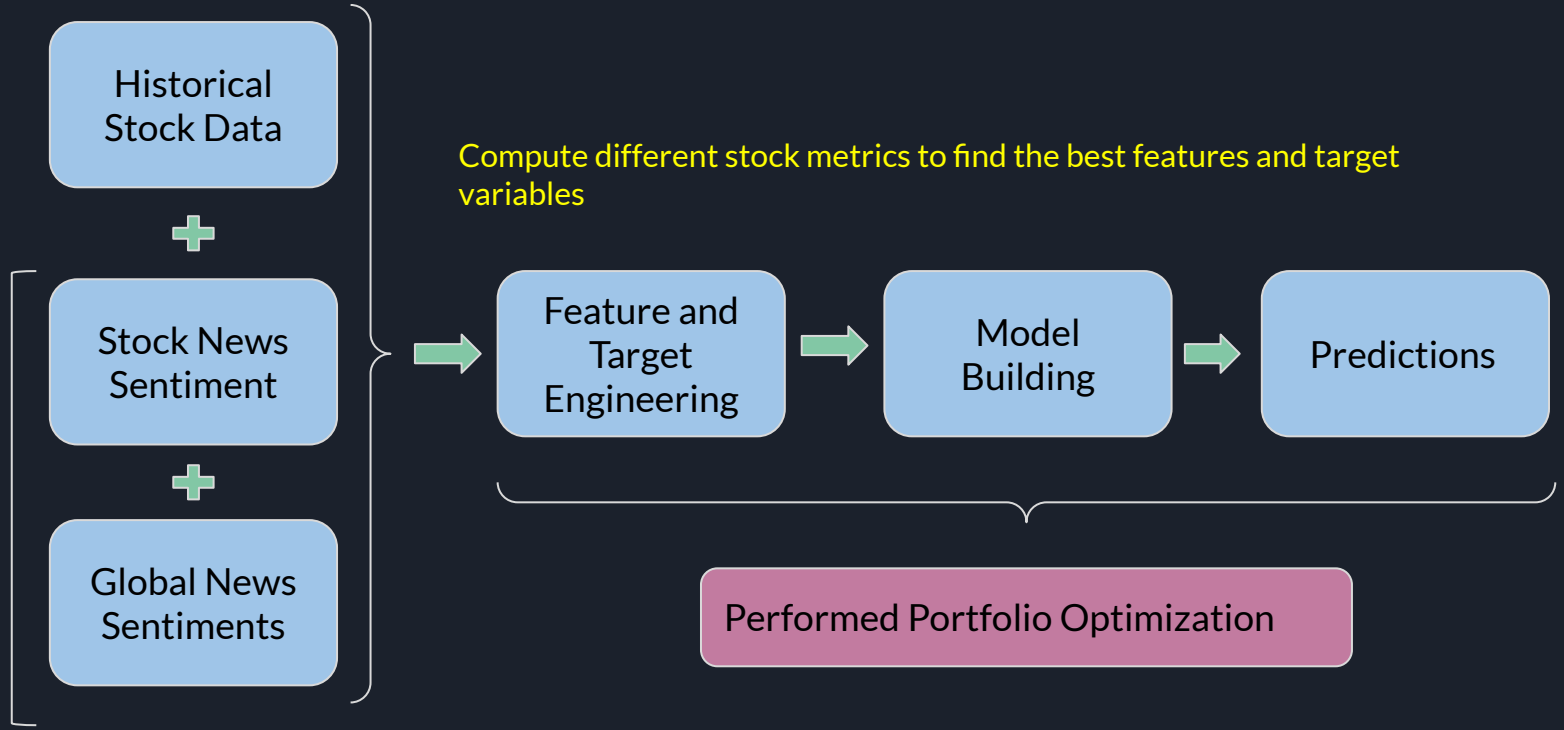
- Reddit
- Kaggle
- BBC News
- Stocks API

Global News Data

- Google News (web scraping)
- Kaggle

Data Collected from 2011 - 2024 (present)

Modelling Pipeline



Used VADER

(with custom lexicon)

Features Engineered

- Sentiment scores for stock and global news using VADER and our custom lexicon.
- Other stock features:
  - $RPM = \dfrac{\text{Closing Price}}{\text{SMA}}$
  - $\text{Trend}_N = \text{Number of days the Closing Price went up in the past N days}$
  - $\text{Log Returns} = \log\left(\dfrac{\text{Closing Price}}{\text{Yesterday's Closing Price}}\right)$
  - $\text{Volatility} = \text{std}(\text{Log Returns in the past N days})$
  - Volume Traded

Target Variables to be Predicted

- Tomorrow's Closing Price
- Tomorrow's SMA
- $\text{Binary Target}_{\text{Closing Price}} = \begin{cases} 1 & \text{if Tomorrow's Closing Price} > \text{Today's Closing Price} \\ 0 & \text{otherwise.} \end{cases}$
- $\text{Binary Target}_{\text{SMA}} = \begin{cases} 1 & \text{if Tomorrow's SMA} > \text{Today's SMA} \\ 0 & \text{otherwise.} \end{cases}$
- Percentage Difference of Tomorrow's Closing Price and Today's Closing Price

Modeling Strategy

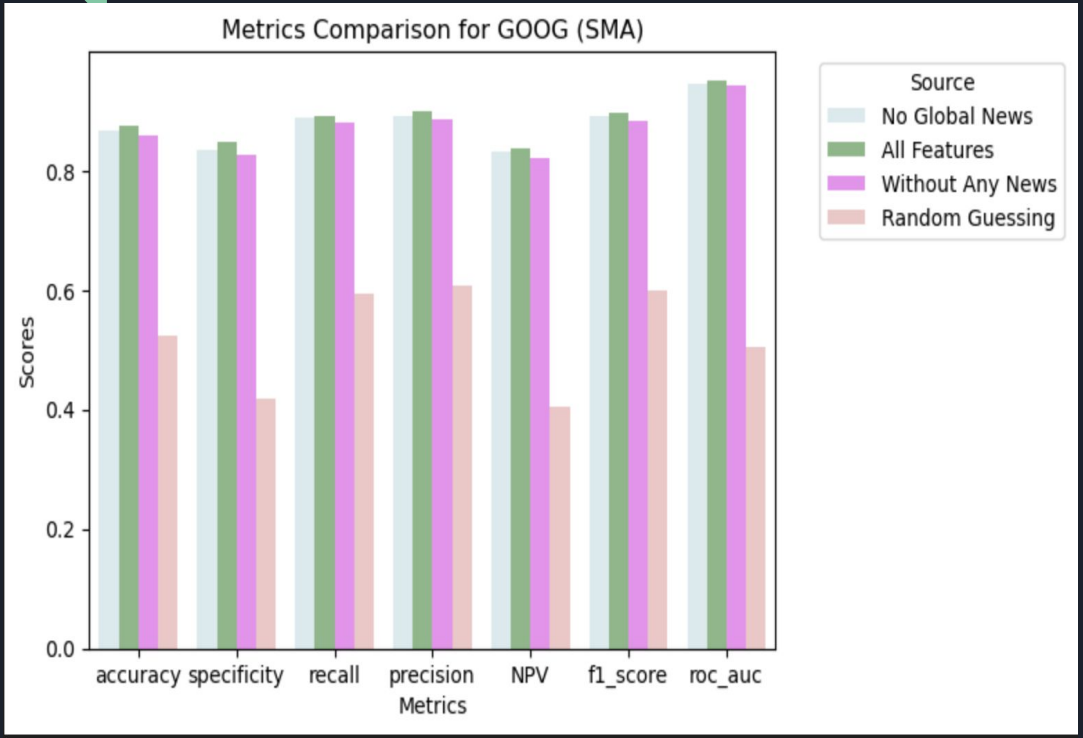
Regression Models

- Predicts Stock Prices
- Baseline : ARIMAX (Regression) and Random Baseline
- Linear Regression
- K-Nearest Neighbours
- ADA Gradient
- Random Forest
- Gradient Boosted Trees
- XGBoost

Classification

- Predicts Growth or Trends
- Logistic Regression
- ADA Gradient
- Random Forest
- Gradient Boosted Trees
- XGBoost

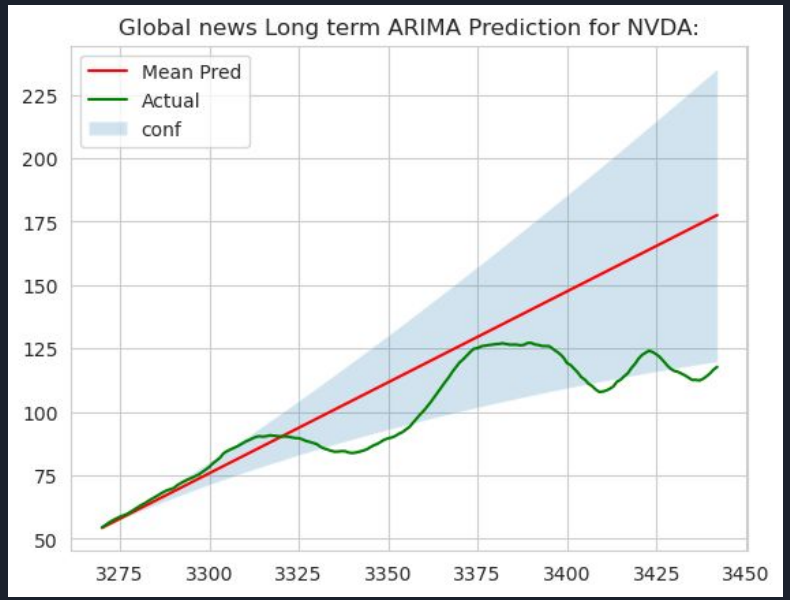
Do global news and stock news affect stock price predictions?



- News data boosts model performance across all metrics, even with global news sentiment.

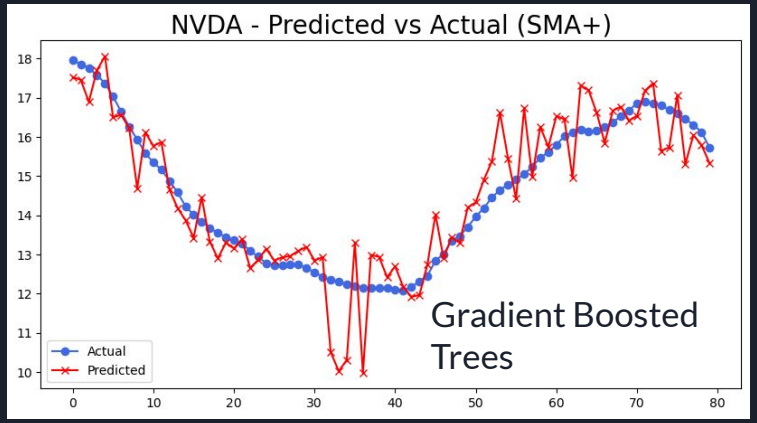
Prediction Results

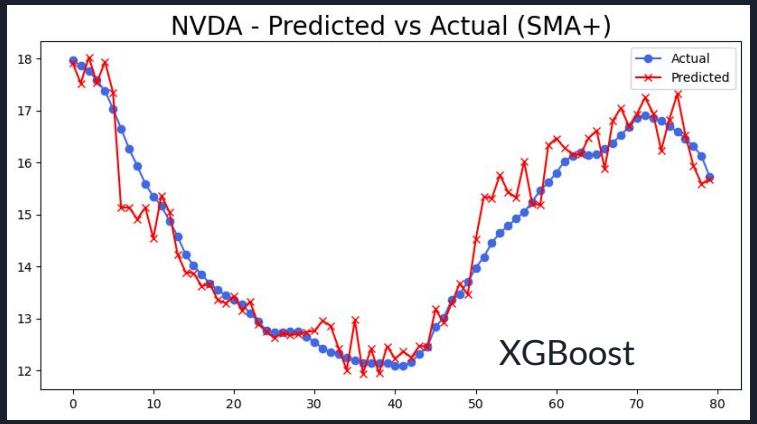
ARIMAX (Baseline Predictions)



SMA Prediction

Time (days)





SMA Prediction

Time (days)

Portfolio Optimization

Feedback on day *d+1*

Import Metrics:

- Model Accuracy
- MSEs
- Predicted Price (P_{d+1})
- *Actual Price* (S_{d+1})
- Predicted Probabilities $Prob^{d+1}_{up}$ and $Prob^{d+1}_{down}$

Trading Strategy:

- **Buy *k* stocks if:** $Prob^{d+1}_{up} > Threshold_{up}$
- **Sell *k* stocks if:** $Prob^{d+1}_{down} < Threshold_{down}$
- **Hold otherwise.** Update cash and positions on day *d*

Portfolio Returns:

$Returns_{d+1} = Cash_{d+1} + k_d S_{d+1}$

$$k = k_0 \left(\frac{\text{Accuracy}_{\text{SMA}}}{\sqrt{\text{MSE}_{\text{SMA}} + \text{MSE}_{\text{P}}}} \right) |\text{SMA}_{d+1} - P_{d+1}|$$

Portfolio Growth and Model Performance

- Total investment of $100,000
- 30% GOOG and AMZN, 20% AAPL and 10% for MSFT and NVDA
- Trained our models during the timeline November 2020 to October 2023
- Ran the portfolio simulation from May 2024 to November 2024

Model	Portfolio Growth
Buy and Hold	11.46%
Logistic Regression	20.23%
XGB	23.61%
Gradient Boosted Trees	29.11%

What to do next?

- Extend the methodology to other S&P 500 stocks and sectors.
- Explore applications for risk assessment and ETF investment strategies.
- Develop advanced models using deep neural networks (DNNs).
- Improve the portfolio strategy by introducing a robust risk management strategy
- Account for factors like market volatility and adjusted returns for real-world factors.

The Team

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

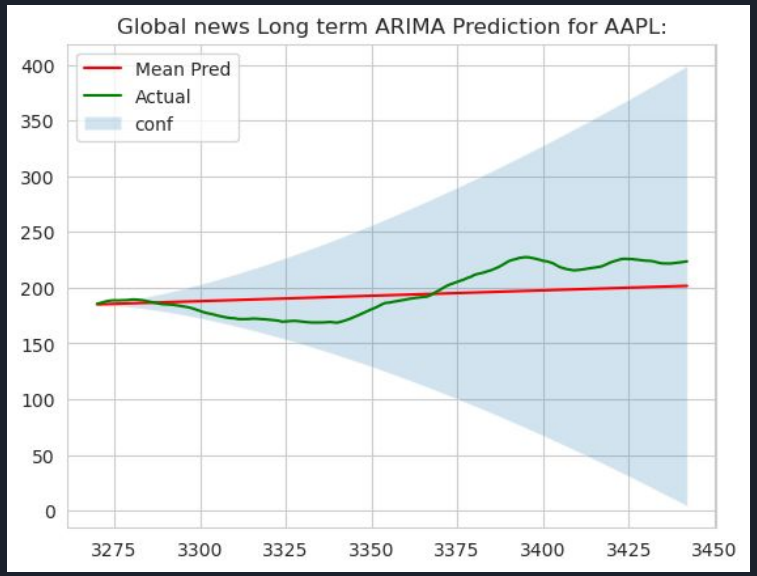
Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur

Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum

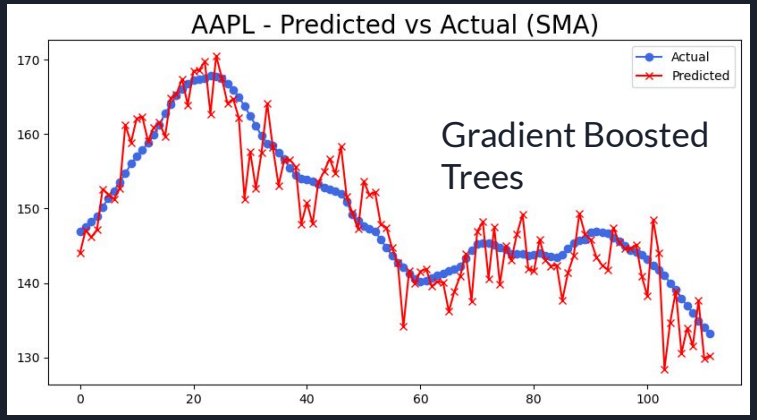
Prediction Results

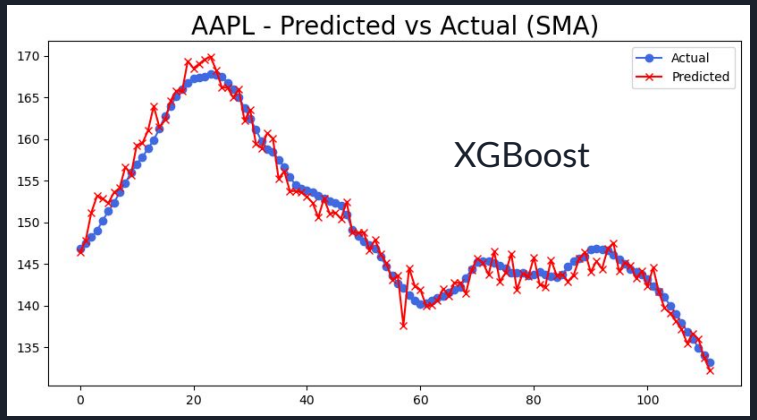
ARIMAX (Baseline Predictions)



SMA Prediction

Time (days)





SMA Prediction

Time (days)

Project objective:
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do

Modelling Strategy

– Baseline : ARIMAX (Regression) and Random Baseline

– Linear Regression

– K-Nearest Neighbours (R)

– ADA Gradient (R)

– Gradient Boosted Trees (best prediction) (R/C)

– XGBoost (best prediction) (R/C)

– Logistic Regression (C)

– Random Forest (C/R)

Understanding the market

Target audience

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt

The competition:

- Lorem ipsum
- Dolor sit amet

Market trends

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

Client Implications:

- Incididunt ut labore et dolore
- Consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

Client Implications:

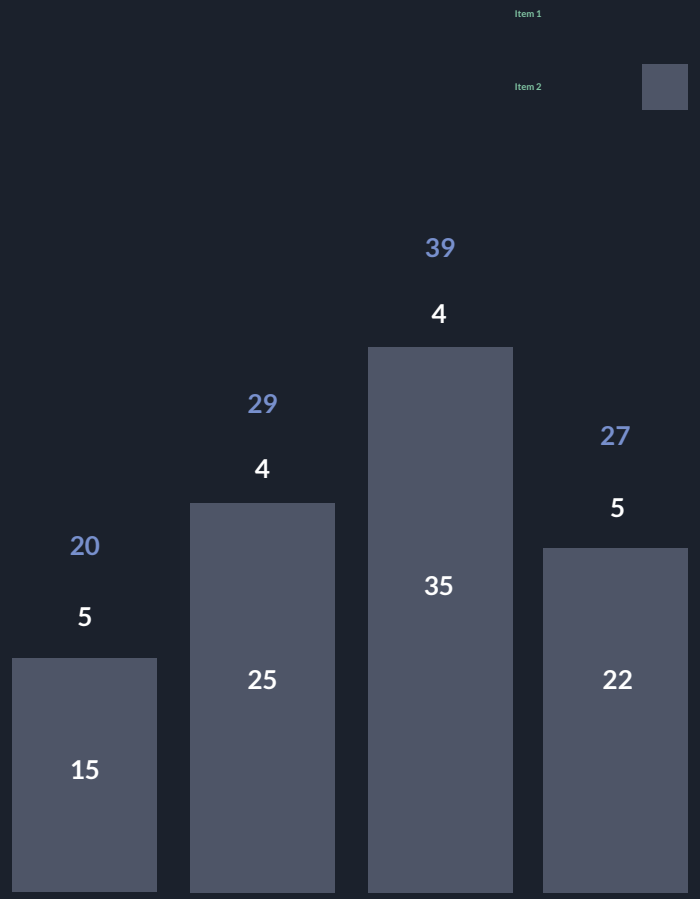
- Incididunt ut labore et dolore
- Consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore

Trend analysis

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor

Client Implications:

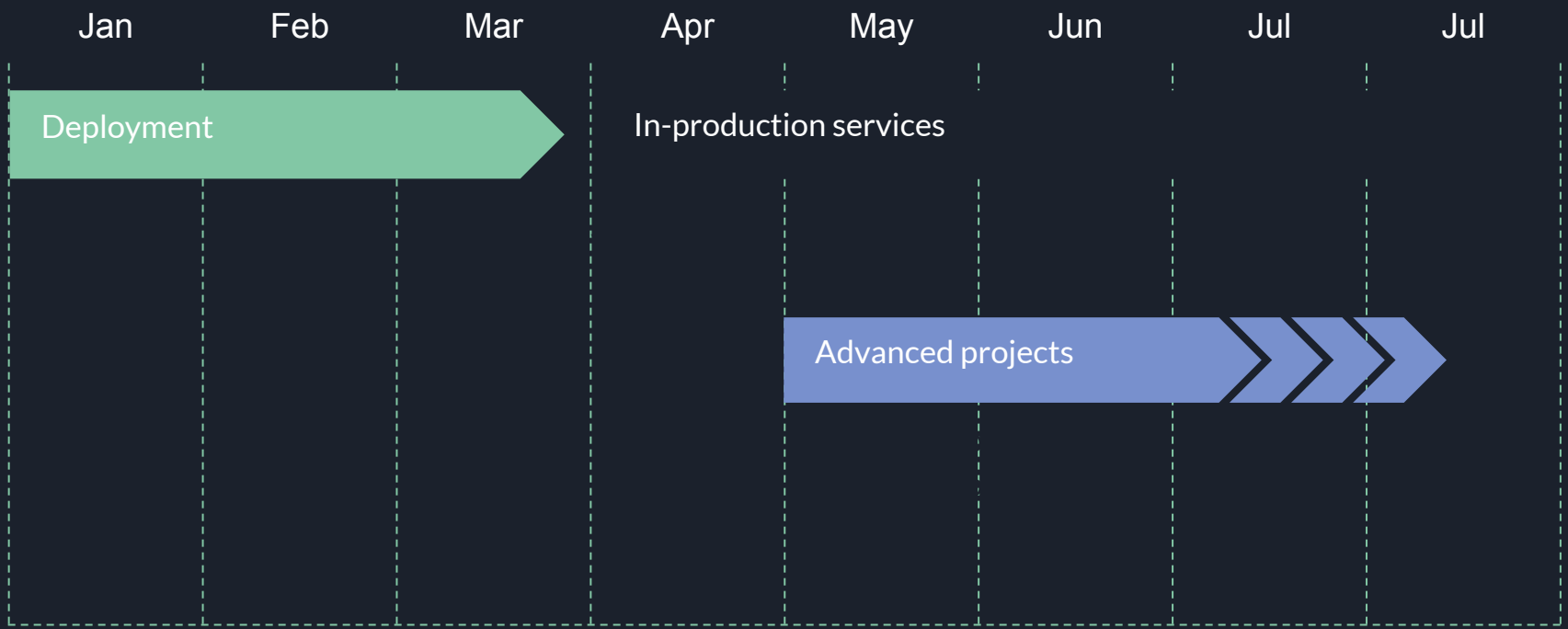
- Incididunt ut labore et dolore
- Consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore



Proposed deliverables

Deliverable 1	• Lorem ipsum dolor sit amet • Sed do eiusmod tempor incididunt ut labore
Deliverable 2	• Lorem ipsum dolor sit amet • Sed do eiusmod tempor incididunt ut labore
Deliverable 3	• Lorem ipsum dolor sit amet • Sed do eiusmod tempor incididunt ut labore
Deliverable 4	• Lorem ipsum dolor sit amet • Sed do eiusmod tempor incididunt ut labore

Timeline


Jan
Feb
Mar
Apr
May
Jun
Jul
Jul
Deployment
In-production services
Advanced projects