

The Blavatnik School of
Computer Science and AI

The Raymond and Beverly Sackler
Faculty of Exact Sciences
TEL AVIV UNIVERSITY



Project Proposal

First Option : Stocks Prediction

Workshop in Data Science
Team 003 - Itay Mutzafi, Moran Zaks , Shaked Schnarch

Background & Topic



Motivation for this task

- Short-term decision making in trading
- Research ML performance on noisy time-series data

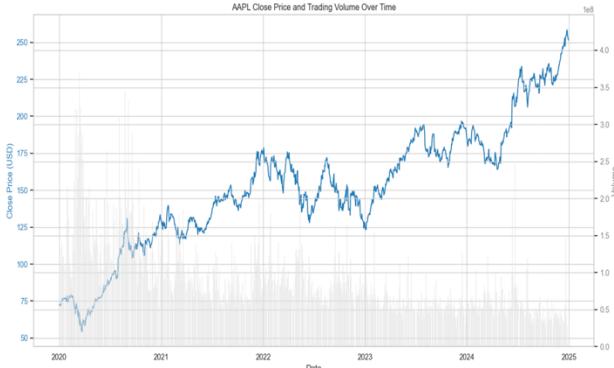


The Task

Binary Classification based on the past data, predict if Apple's stock will **increase** or **decrease** in the future .

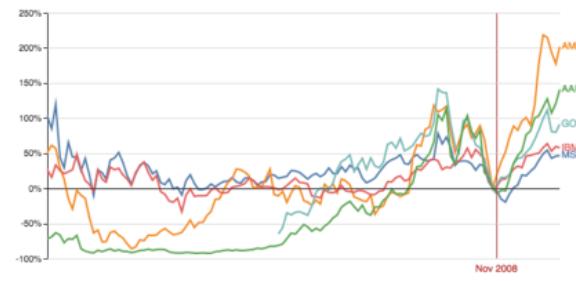
Main dataset

Yahoo Finance's API- Apple's Stock



Additional dataset

Other Stocks



Additional dataset

Google News

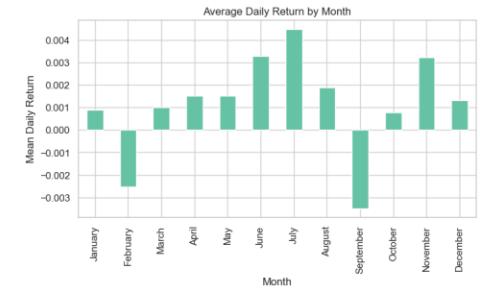
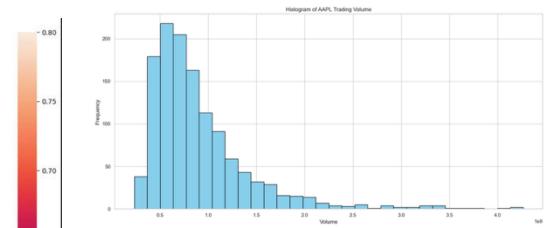
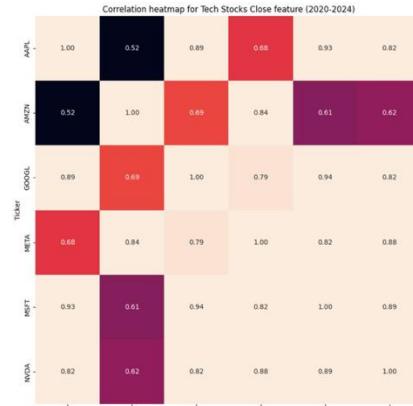
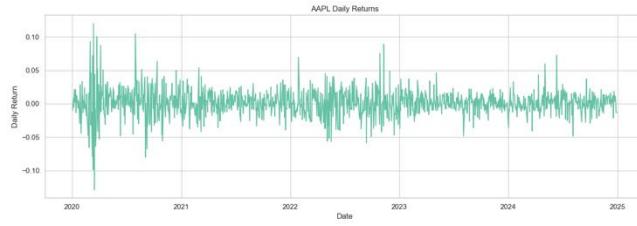


Feasibility Analysis



Initial Data Exploration

- **Total Samples:** 1275 t.s. for each stock
- **Core Features:** Open, Close, High, Low, Volume

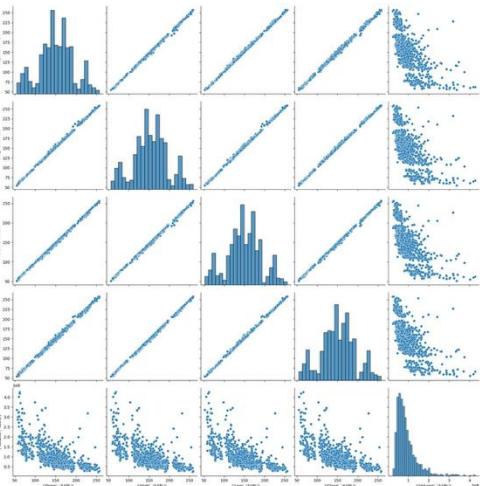


Expected Challenges:

- **Feature Engineering** - Daily return, Moving Average, Volatility etc.
- **Seasonality**
- **News Sentiment Analysis**

Why is this Feasible?

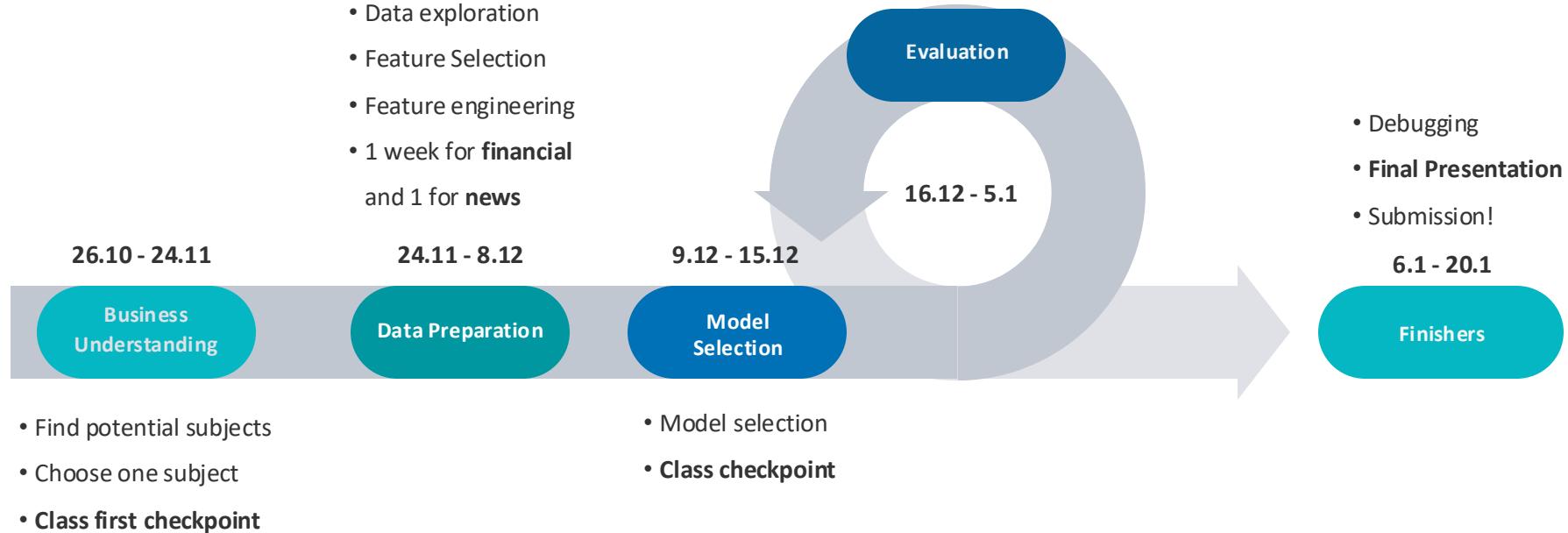
- **Classification and not exact price**
- **Highly-documented data**



Methodology

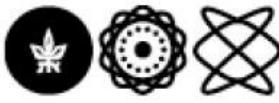


Project Work Plan



Success Criteria

- **Good model:** 55-60% accuracy prediction (Much better than random guessing).
- **Stability over time:** year-to-year accuracy drop $\leq 5\%$.



The Blavatnik School of
Computer Science and AI

The Raymond and Beverly Sackler
Faculty of Exact Sciences
TEL AVIV UNIVERSITY



Second Option : Detecting Autism Patterns in Brain Signals

Beyond Static Averages: Quantifying Temporal Dynamics in Calcium Imaging



Background & Topic

The Biological Problem

- **Context:** Shank3 mutation cause Autism.
- **Discovery:** Recent reveals this mutation affects OPC cells.
- **The Phenomenon:** Mutant cells exhibit a "Fading Phenotype"

The Task

Supervised Binary Classification: Predict **WT** vs. **Mutant**.

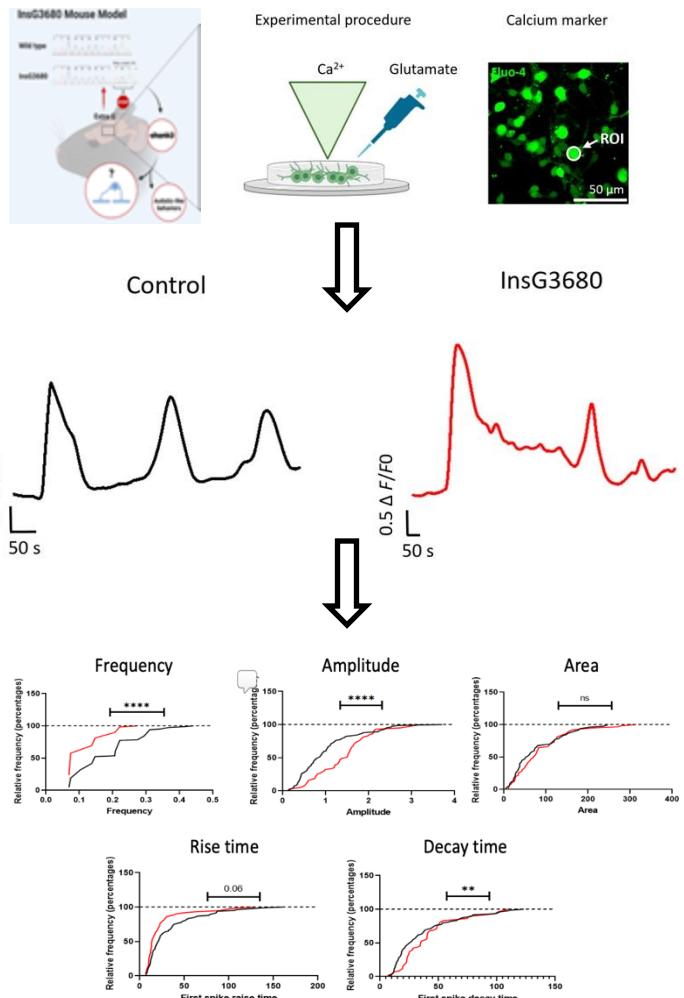
Source
Boaz Barak's Lab

Data Type
Time-Series (Calcium Traces)

Dimensions
810 Frames (15 min) x 1000 ROIs

The Baseline

- **Existing analysis:** relies on static averages
- **The Current Limitation:** High information loss.





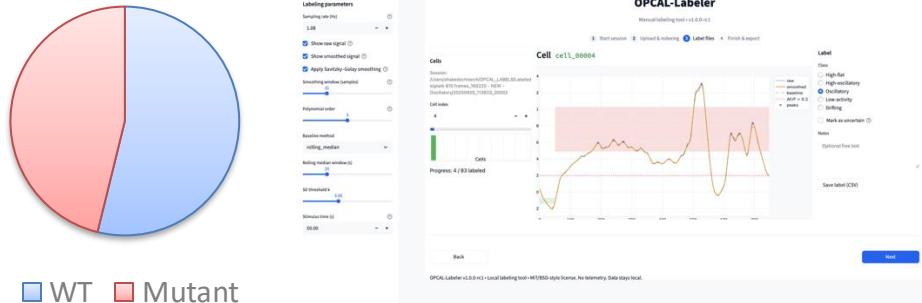
Feasibility Analysis

Curation Pipeline:



Initial Data Exploration

- Total Samples (ROIs):** N = 4,600
- Time Series Length:** 810 frames (15 mins) @ 1.1s interval.
- Key Event:** Glutamate Injection at frame ~102.



Our DS Innovation: Temporal Feature Eng.

- Proposed Features:** Decay Slope (rate of decline), Early-vs-Late activity ratio, Stability across time.
- Group-level split** by animal to prevent data leakage.

Why is this Feasible?

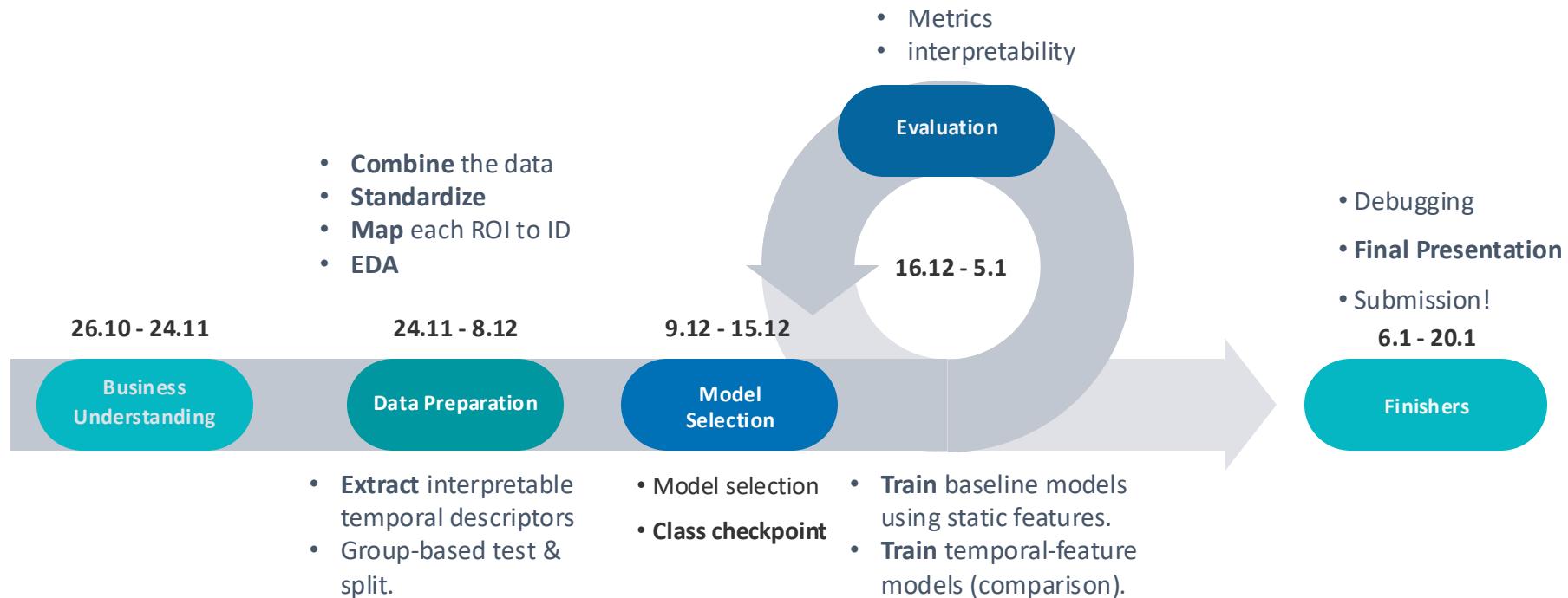
- Scientific Validity:** The "Fading" effect is a documented biological mechanism.
- Baseline Evidence:** even baseline achieved partial separation.



Methodology



Project Work Plan



Success Criteria

- **Quantitative:** Improvement over the static baseline.
- **Qualitative :** Identification of **new** temporal features that capture the fading signature.

Stocks vs. Brain Signals



Stocks

- ✓ Excellent documentation and reliable API
- ✓ Include data integration between two data sources
- ✓ Feasible and intuitive
- ✗ Hard to achieve strong results (noisy data)
- ✗ Complex feature engineering



Brain Signals

- ✓ Has clear and well-justified scientific value
- ✓ Highly feasible classification task
- ✓ Chance of clear and strong results
- ✗ Domain-specific challenge
- ✗ Hard to visualize for general audience

Preference: focus on stocks