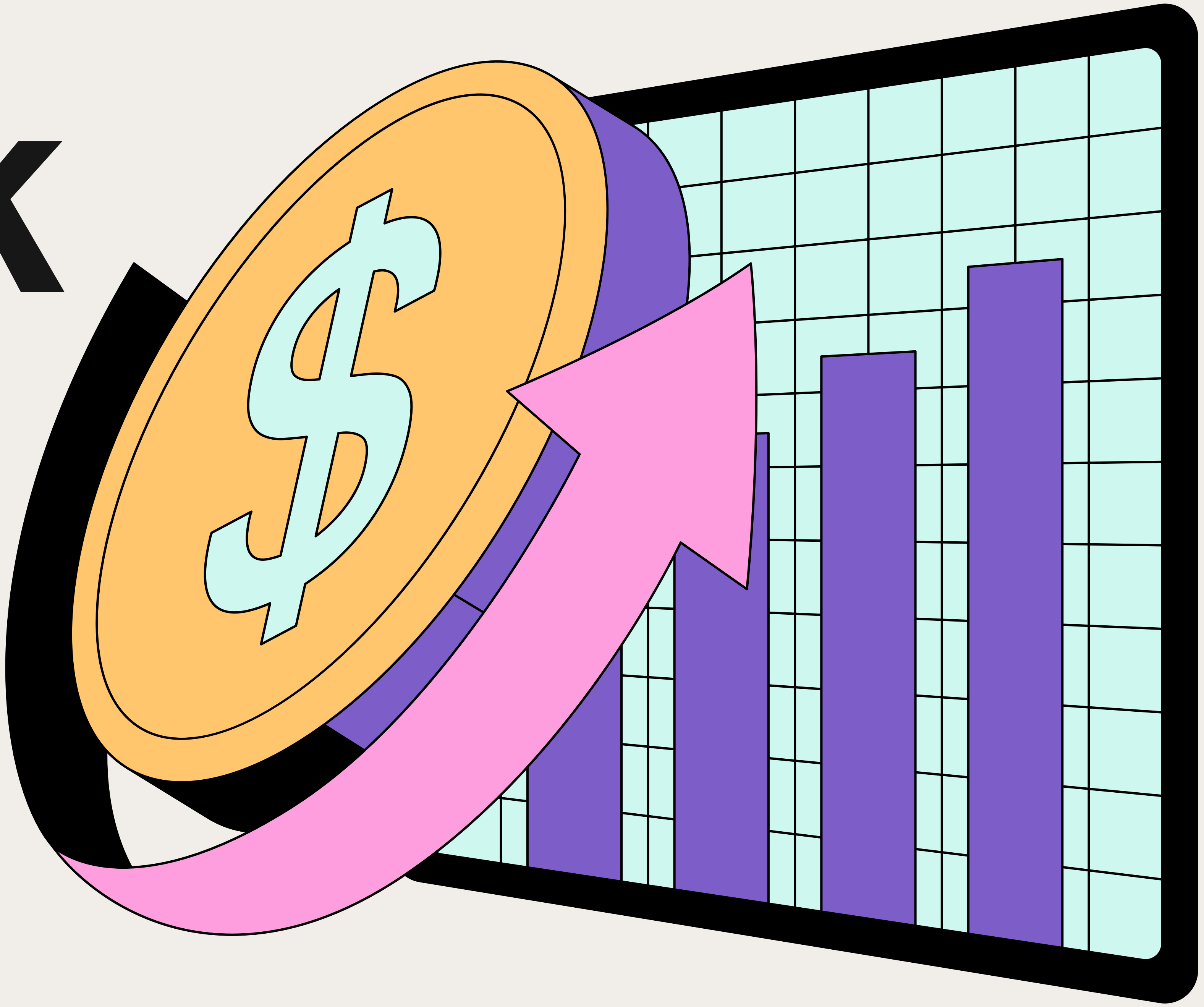
An EGARCH and XGBoost prediction of

NVIDIA STOCK PRICE VOLATILITY



Saumya Bothra

Why NVIDIA's volatility?

Daily stock prices are **extremely noisy,** but the volatility of those prices isn't.

NVIDIA, **one of the most actively traded and volatile large-cap stocks**, provides a clear example of this behavior.

Rather than trying to predict exact price movements something that borders on impossible at a daily frequency, this project focuses on **forecasting how much the stock is likely to move**, not which direction it will move.

This is far more useful for **risk management, position sizing, and planning** around uncertain market conditions.

The goal was to combine traditional econometric models, like EGARCH, with modern machine learning approaches such as XGBoost to understand **whether short-term volatility can be predicted reliably** and how these forecasts could support more informed decision-making.

Provides enough info to observe:

- Volatiloty clustering
- Momentum and mean reversion
- Regime changes (low to high volume

Roughly 6500 cleaned observations

Variables

- Date
- Open
- High
- Low
- Close
- Adjusted Close
- Volume

Data

YAHOO FINANCE API 1999-PRESENT

Data Transformation

Log Returns
Ln(AdjClose(t)/AdjClose(t-1))
stable and additive for volatility model

Simple Returns
(AdjClose(t)-AdjClose(t-1))/AdjClose(t-1)
for interpretation and comparison

Daily Range (%)
(High(t)-Low(t))/AdjClose(t)
Intraday price movement

Overnight Gap
(Open(t)-Close(t-1))/Close(t-1)
Sentiment and overnight shock

Rolling Mean and SD (**5 and 20 day**)

$$\frac{1}{n}\sum_{i=0}^{n-1} Return_{t-i}$$

$$\sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(Return)_{t-i}-\vec{R})^2}$$

Short term V medium term market regimes

Lagged Returns (**1,2,5**)
Return(t-1,2,5)
XGBoost can detect autocorrelation

Volume Transformations

$$VolumeLog_t = log\,(Volume_t)$$

$$VolumeRatio_t = \frac{Volume_t}{Volume_{t-1}}$$

Relative spikes or drops in trading activity

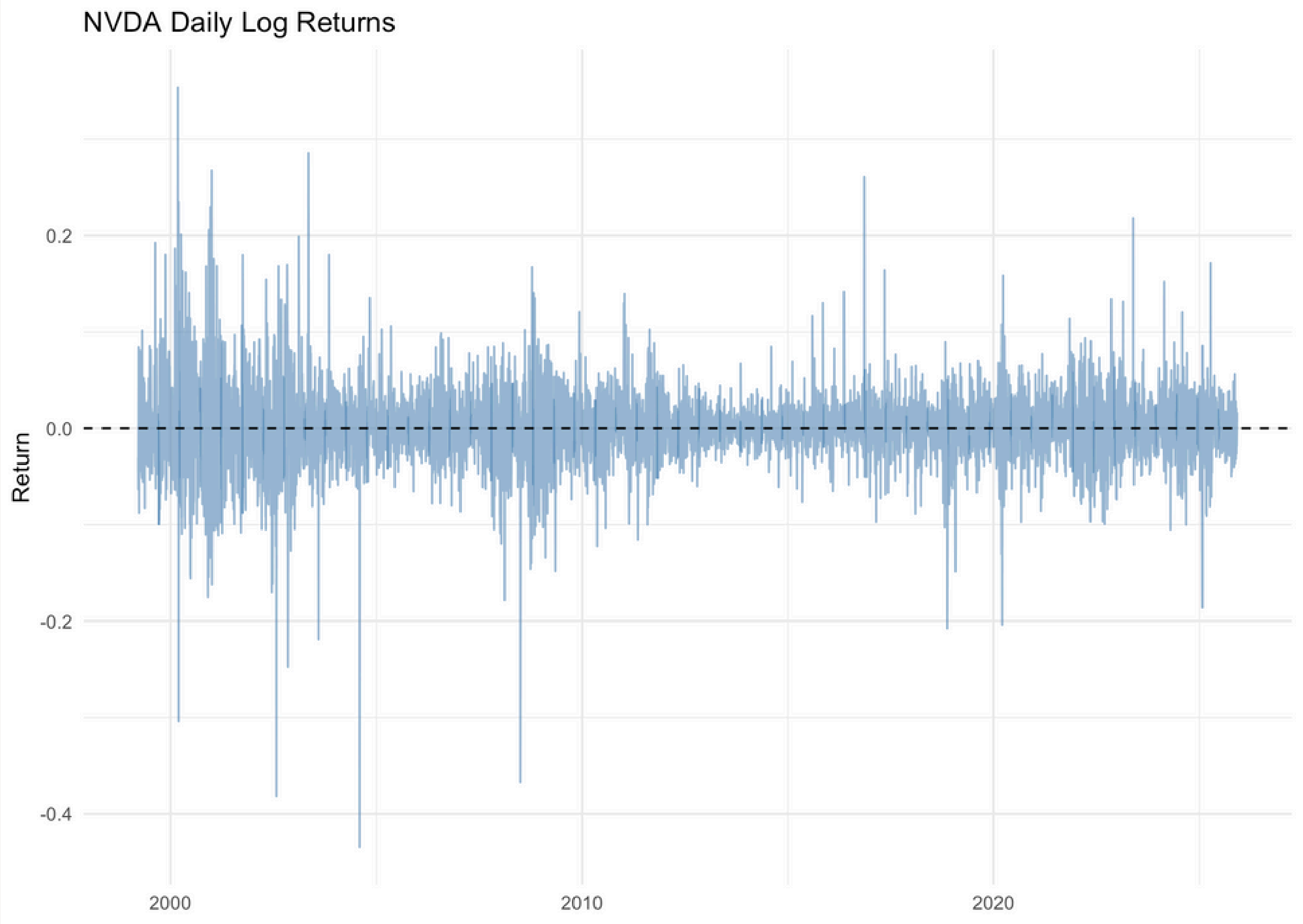
Next day Return
REGRESSION TARGET

Next day Direction
BINARY CLASSIFICATION TARGET (1 IF NEXT DAY RETURN>1

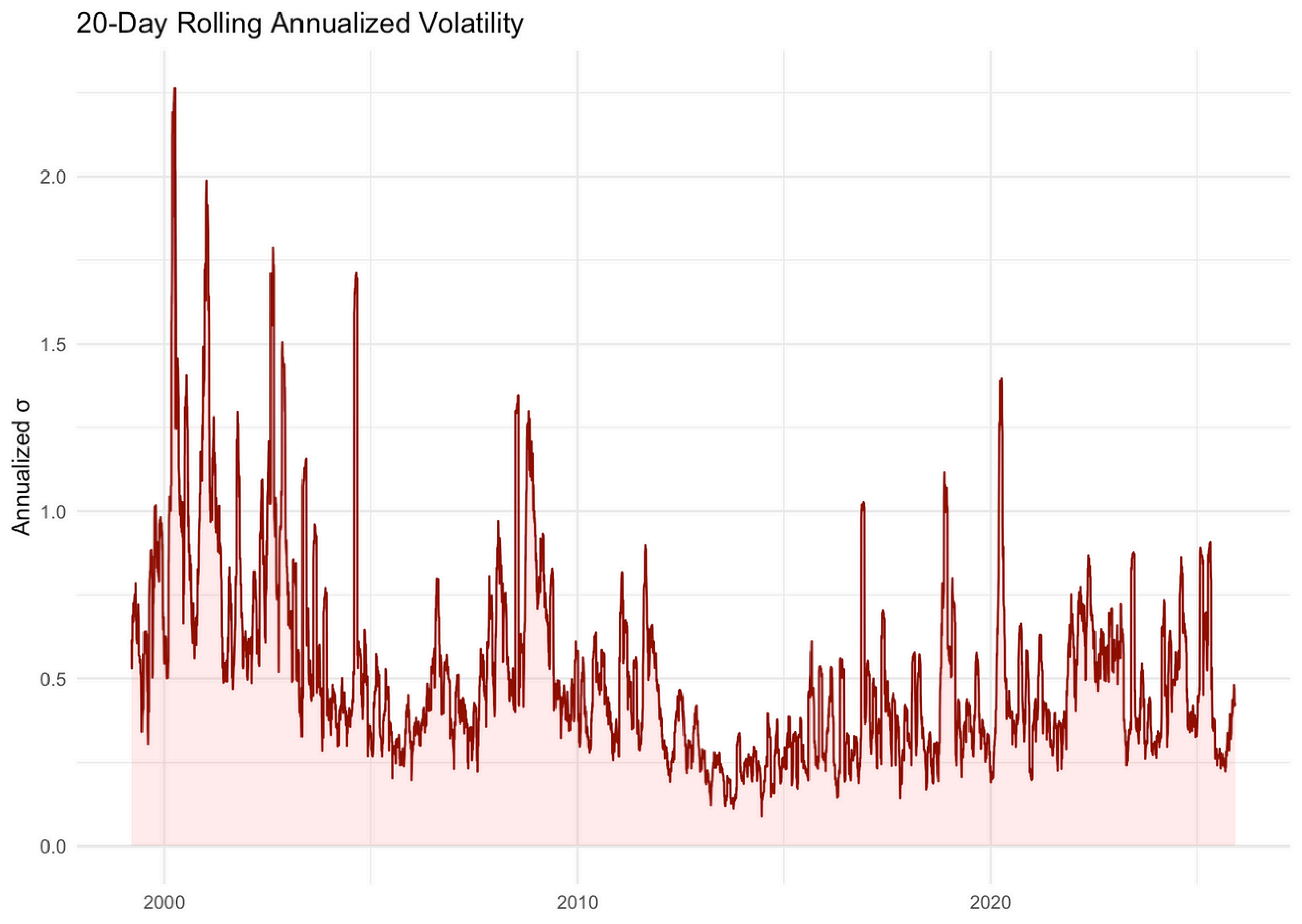
Volatility
Next day absolute return (volatility proxy)

Long term Exponential Growth

High noise, Fat tails, and Extreme Jumps



Initial Data Visualization



Clear Clusters during crises and earning cycles

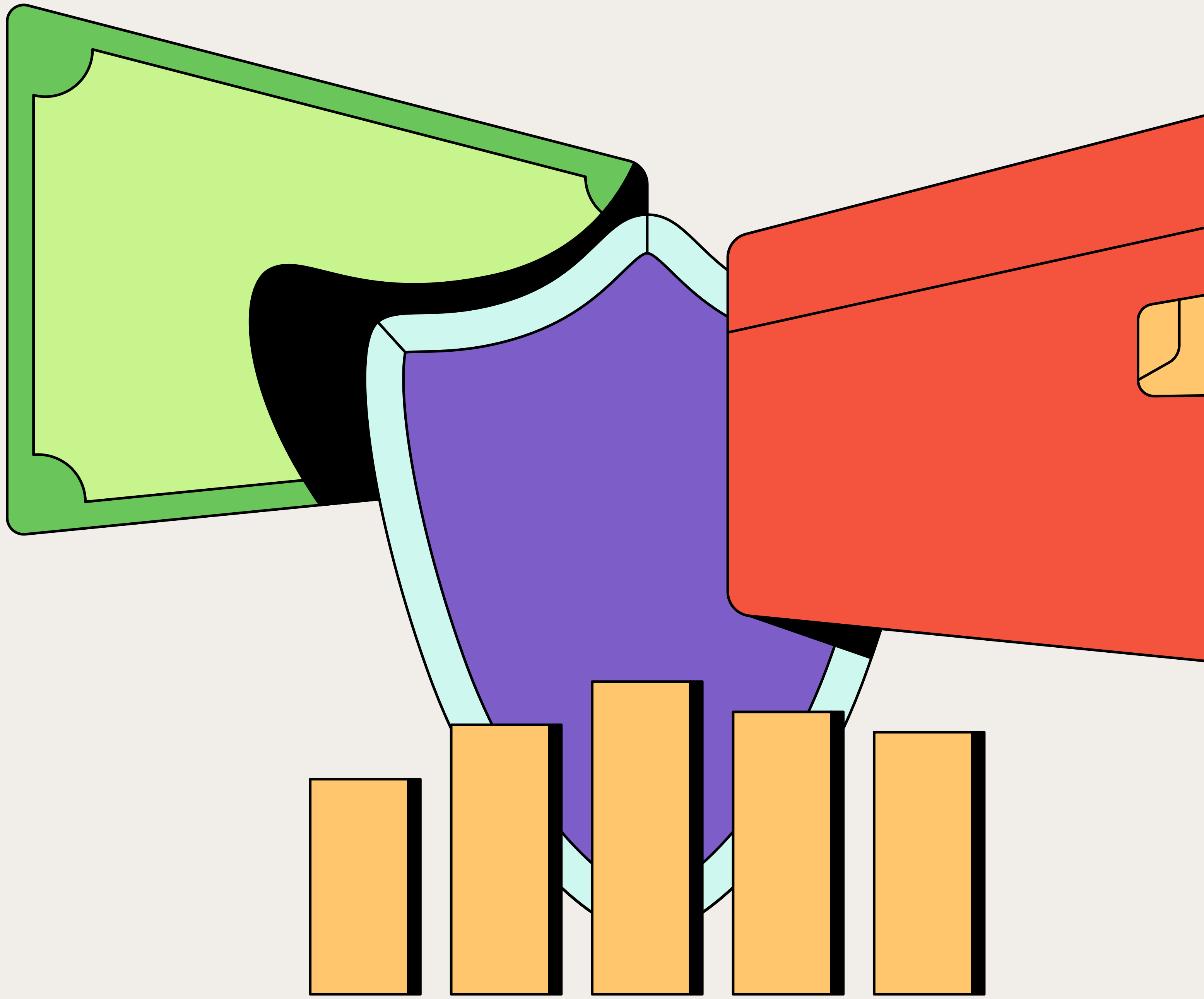
Why EGARCH?

- Handles volatility asymmetry (bad news increases volatility more)
- Allows for fat tailed errors
- captures volatility clustering

Model Used: EGARCH (1,1) with T-distribution

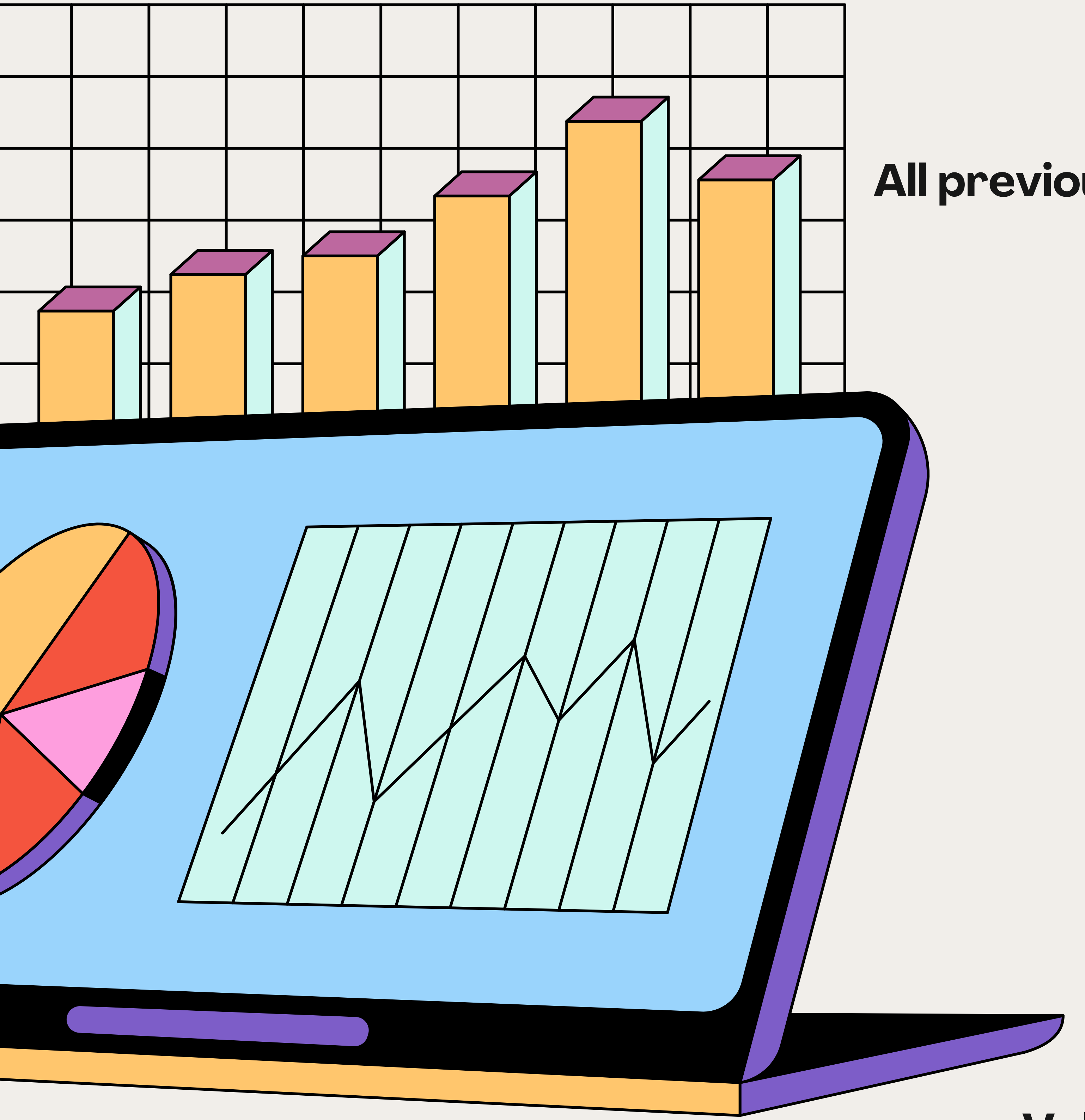
- Outputs:
  - conditional mean
  - conditional volatility
- conditional volatility becomes key for ML model
- provides baseline volatility estimate for comparison

EGARCH Model



Feature engineering for XGBOOST

All previously engineered features were kept and the following were added:



EGARCH features: EGARCH Conditional Volatility, EGARCH Conditional Mean, EGARCH Residual
Volatility dynamics: Volatility Persistence, regime flags
Shock flags: Jump, Price acceleration
Upgraded volume features: Log Volume, Volume Ratio, Volume Surge, Volume Price Correlation
Technical context: RSI, Price Position
Calendar dummies: Monday/Friday, specific months
Extra targets: direction and absolute return (volatility target)

Relative strength index: indicator of price momentum
Volatility structure: Ratio of short term/Long term volatility
VolumexPrice Interaction: rolling 20 day correlation b/w log vol returns

XGBOOST MODEL

Strong performance on tabular time-series features
Built-in handling of nonlinearities & interactions

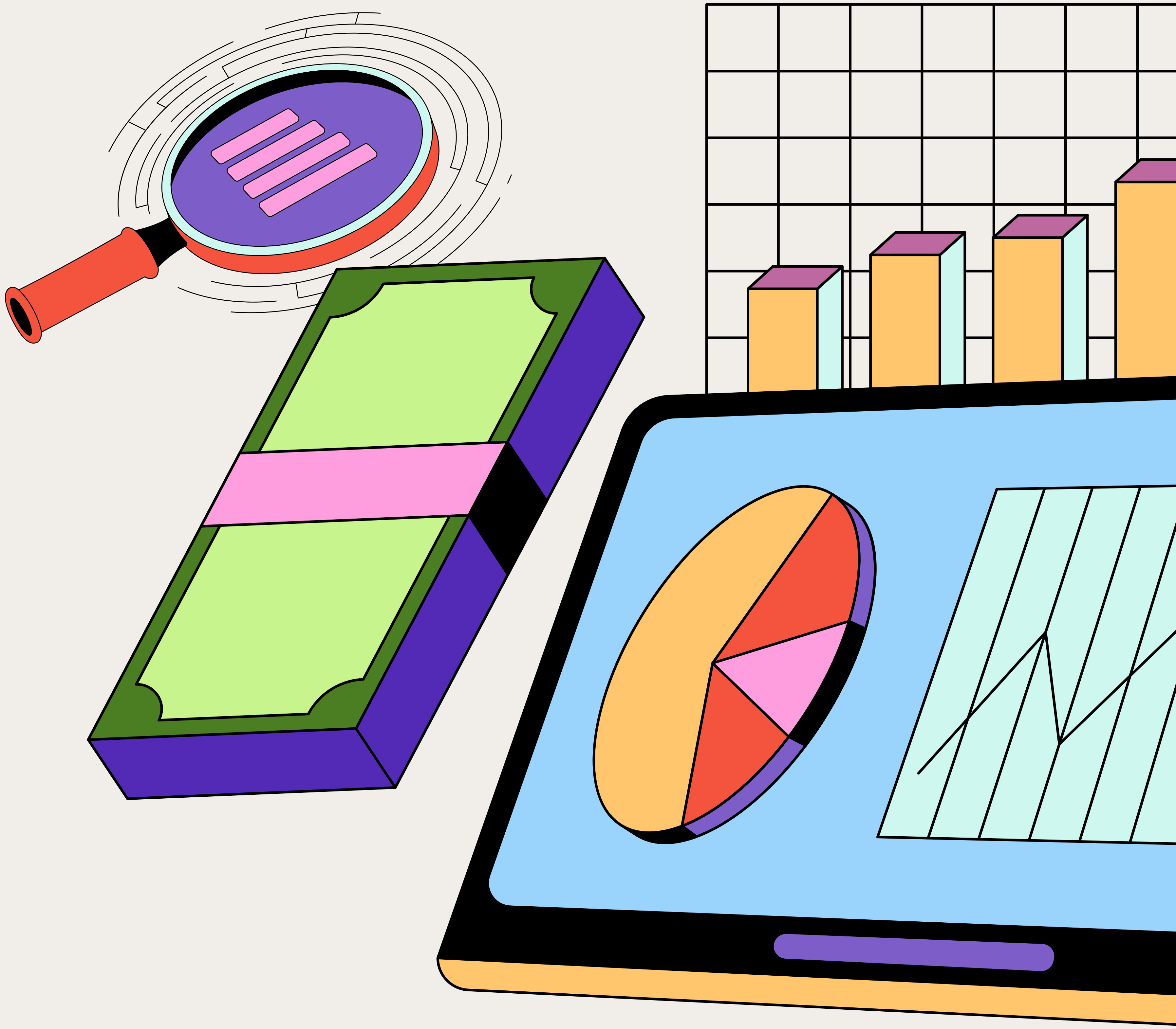
3 separate models:

- Regression: predict next day return
- Classification: predict up/down direction
- Regression: predict next day absolute return (volatility)

Parameters:

- Learning Rate = 0.01
- Max Depth = 4
- Subsample = 0.8
- Early Stopping = 50 rounds

Time ordered split (**80**/**20**)



Volatility today mainly depends on volatility yesterday

- Large intraday ranges strongly predict next day movement size
- medium term drift (20 day rolling) matters more than short term noise

XGBoost Model Performed BEST with:

RSME: 0.021
MAE: 0.0155
Correlation with true Volatility: 0.25

Top Volatility Predictors:

EGARCH Conditional Volatility
Daily Range percentage
20 Day Rolling mean
Return lag of 2 days
Return lag of 5 days

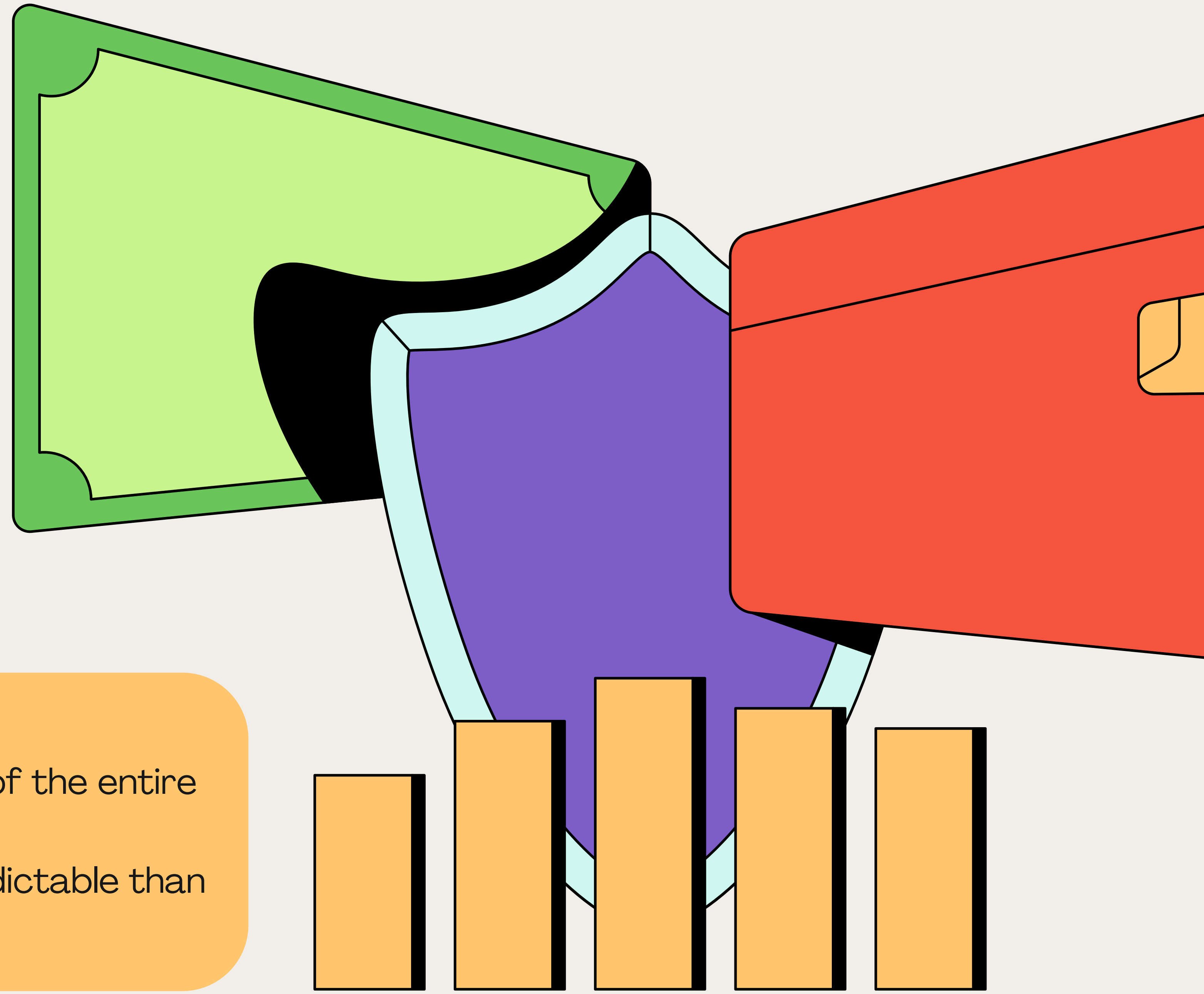
Volatility Summary

RMSE	CORRELATION	NOTE
0.0321	0.04	expected short term prices are unpredictable

ACCURACY	RECALL	NOTE
52%	75%	Model is essentially random, and model flags upwards volatility and moves liberally.

Model Evaluation

- Meaningful correlation
- Best- performing component of the entire system
- Confirms volatility is more predictable than returns



Strategy insights

A simple rule-based trading strategy was tested:

- Buy if
  - Predicted return > 1%
  - Predicted volatility < 70th percentile
- Sell if predicted return < –1%
- Otherwise hold cash

Results:

- Strategy return ≈ 3.7%
- Sharpe ≈ 0.25
- Max drawdown ≈ 3.6%

- model helps avoid high volatility days
- does not give a reliable directional alpha
- useful for **risk control** not stock picking

Trading Implications

This model shouldn't be used for buy/sell strategies because it **doesn't reliably predict next-day returns or direction.**

However, it is quite useful for risk. **It gives a daily estimate of expected movement size** (**volatility**) and a way to classify days as calmer or more turbulent. That can directly drive trading choices like:

- Position sizing (smaller exposure on high-vol days),
- Setting wider or tighter stops,
- Deciding when to avoid trading around noisy regimes,
- Budgeting risk across a portfolio instead of betting on a single directional call.

Dont trade more, trade SMARTER: use the volatility forecast to control downside and manage exposure, and treat any directional signal as weak support at best, not the main decision-maker.