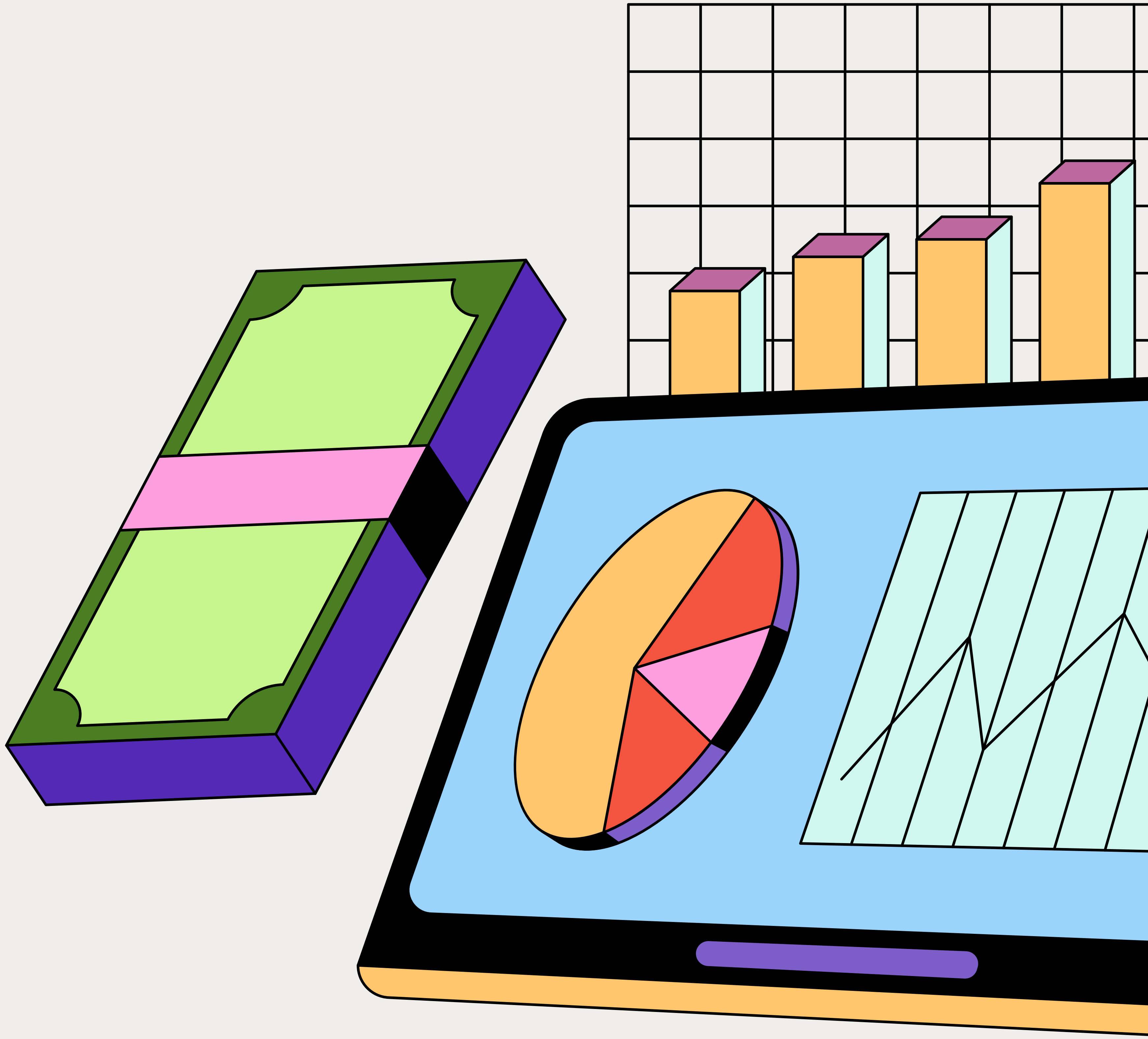
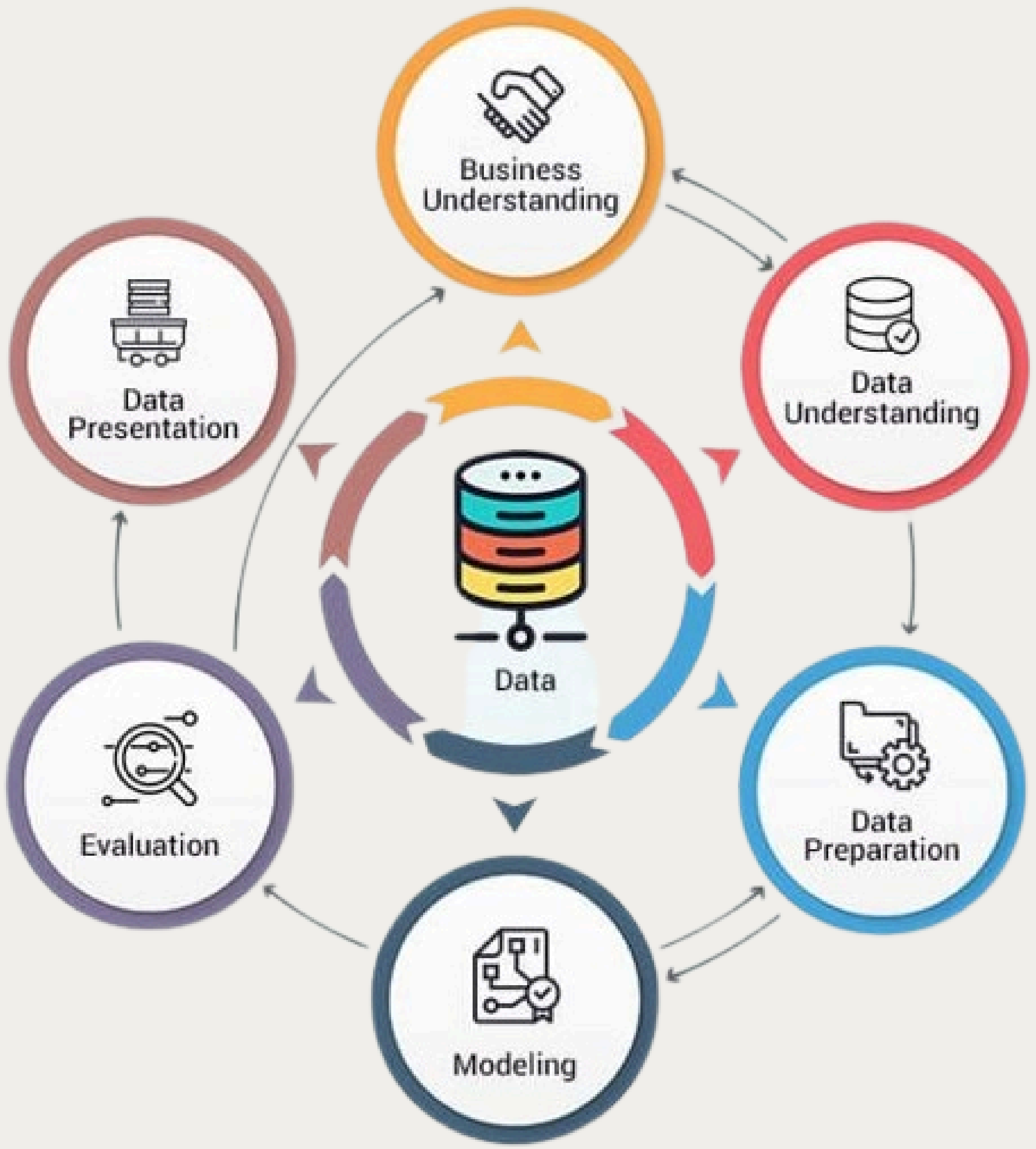
UMHackathon 2025 : Balaena Quant - Alpha Strategies Using HMM or ML

Group Member:

1. Low Yan Cheng
2. Lim Chee Yee
3. Lim Jun Na
4. Sehneel Ansari Binti Naeem Ahmed Ansari
5. Yap Huey Shin



Presented by DA House



Selected Methodology - CRISP DM

CRISP-DM provides a structured, flexible, and proven approach for planning and executing data mining projects, making it ideal for analyzing complex datasets like miner-to-miner flows.

Reasons for Choosing CRISP-DM :

- Step-by-step framework, flexible for refining earlier stages
- Encourages testing multiple techniques on miner behavior.
- Encourages a deep understanding of on-chain transaction flows and aligns data tasks with the objectiv
- Integrates findings into trading strategies and dashboards.

Busniess Understanding

1. **Objectives**
   - Building a machine learning-driven backtesting framework to test and validate alpha trading strategies to maximizes profit
   - Maintain a Sharpe Ratio ≥ 1.8, Max Drawdown ≥ -40%, and trade frequency ≥ 3% per row
   - Ensure execution strategies that simulate real-world conditions

2. **Goal**
   - Discover a trading strategy that gives effective ***buy/sell signals*** to maximize returns
   - Use on-chain data to ***predict future price movements***
   - ***Train and evaluate models*** like HMM, and LSTM to learn market patterns

3. **Idea**

Why does the Bitcoin price often drop after miners move a large amount of Bitcoin out?

4. **Hypothesis**

If miner_outflow_total_y ***increases***, then the Bitcoin price (close) is likely to ***drop***

5. **Convert into Formula**

corr(miner_outflow_total_y, close) < 0

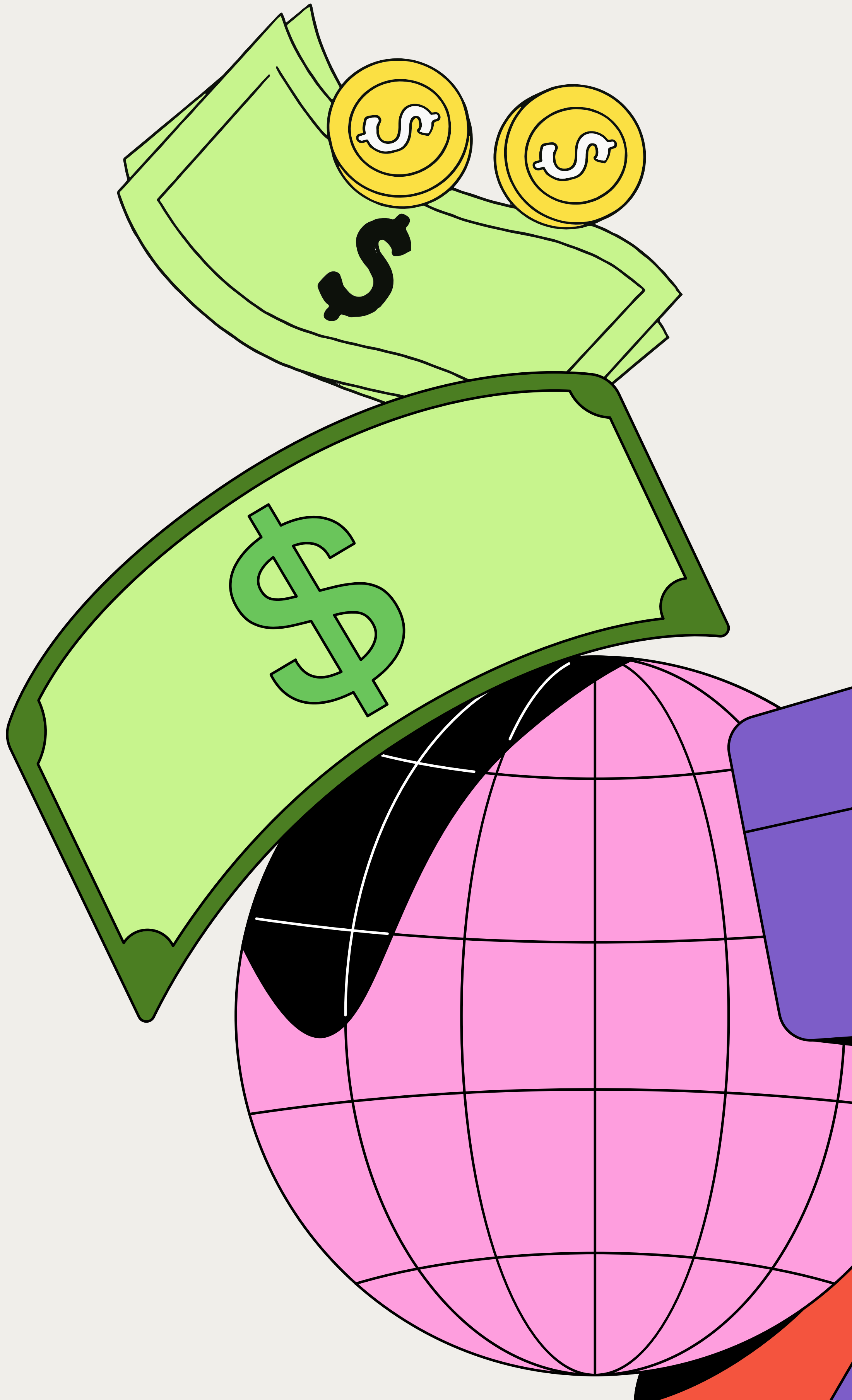
Data Understanding

1. Load and Preview Data
2. Check Data Types and Columns
3. Check for Missing Values
4. Check for Duplicates
5. Statistical Summary
6. Identify Outliers
7. Correlation Analysis

Data Preparation

1. Data Transformation
2. Merging data sets
3. Combine dataset into one dataset
4. Rename duplicate columns

Data Cleaning



1. Handling Missing Values

2. Handling Outliers

3. Rolling mean

4. Final Correlation Analysis (After Cleaning)

Modeling

1. **Define Modeling Goals**
- Detect alpha-generating patterns
- Predict market signals based on on-chain behavior

2. **Select Modeling Techniques**
- Time-series analysis for volume trends
- Machine learning models such as ***LTSM*** for signal prediction

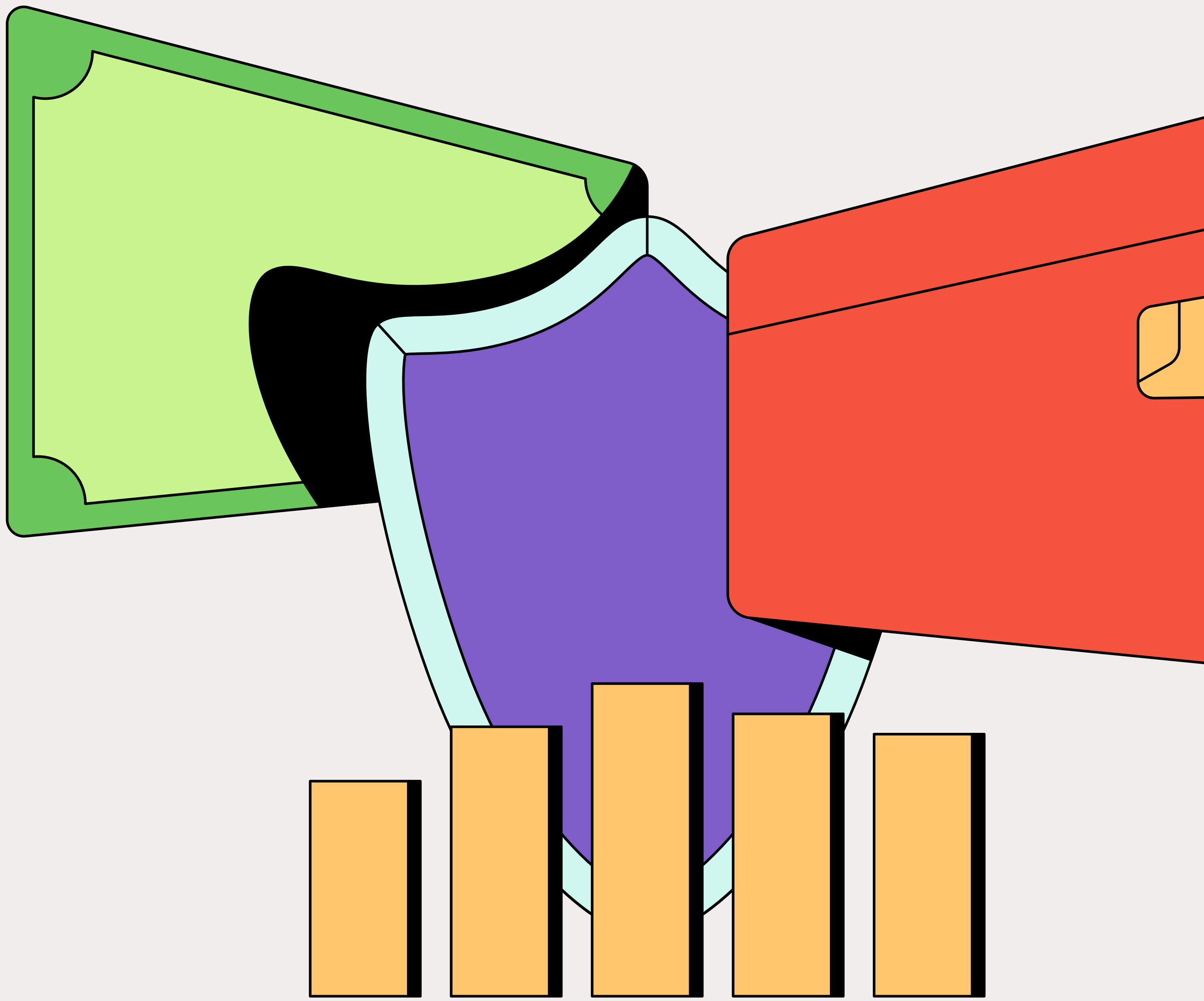
3. **Prepare Data for Modeling**
- Normalize transaction values and timestamps
- Handle missing data and smooth noisy signals

4. **Build and Train Models**
- Train on historical on-chain flow data
- Use rolling windows for time-series consistency

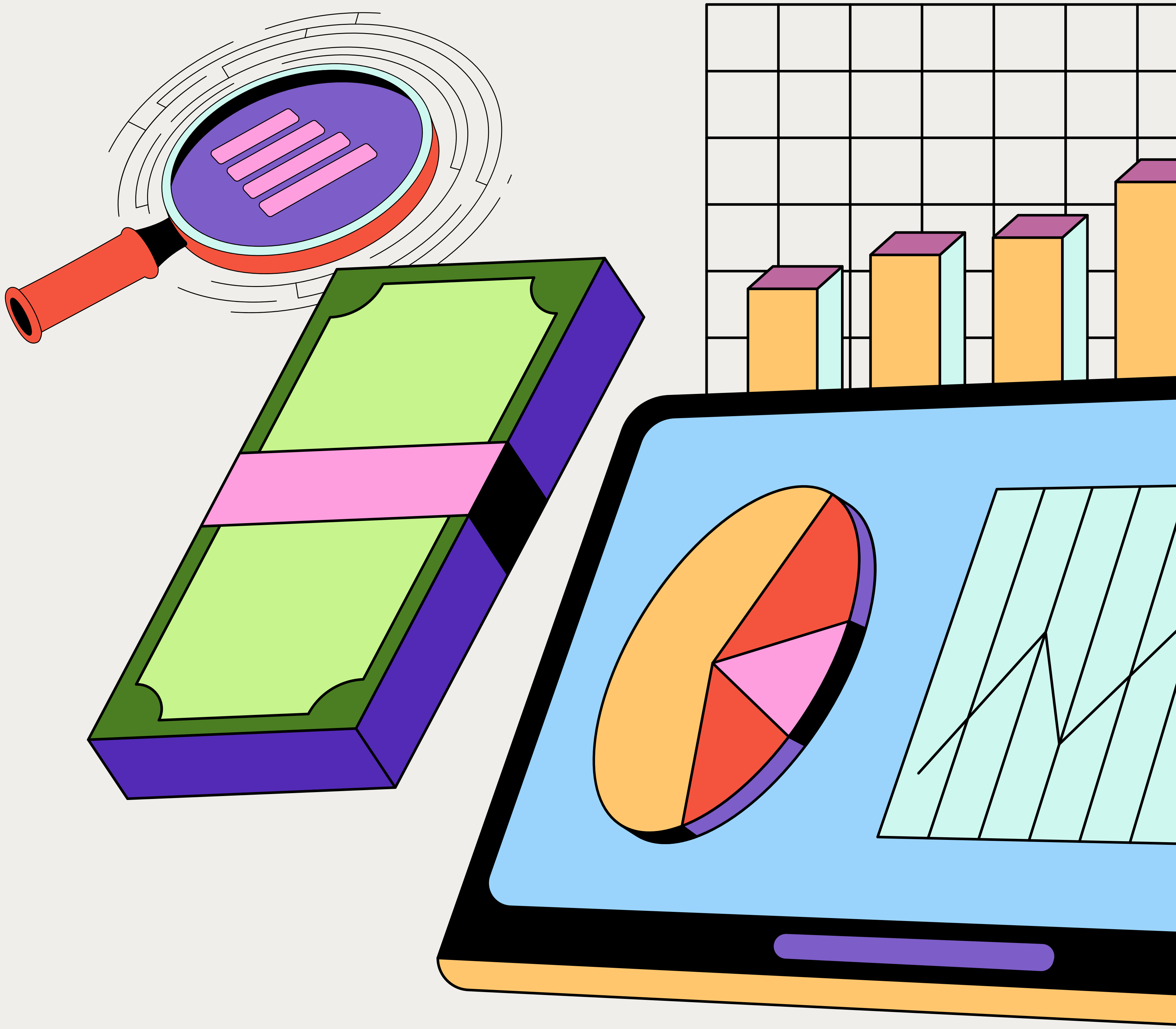
5. **Assess Model Fit**
- Evaluate with crypto-specific KPIs: Sharpe ratio, drawdown, precision of signals

Hidden Markov Models (HMM)

A Hidden Markov Model (HMM) is a time series model that infers underlying states such as mood or health from observable observations like heart rate, activity. It assumes states change in time with fixed probabilities and produce certain outputs. HMMs are used in behavior monitoring, anomaly detection, and pattern mining where the true condition cannot be directly observed.

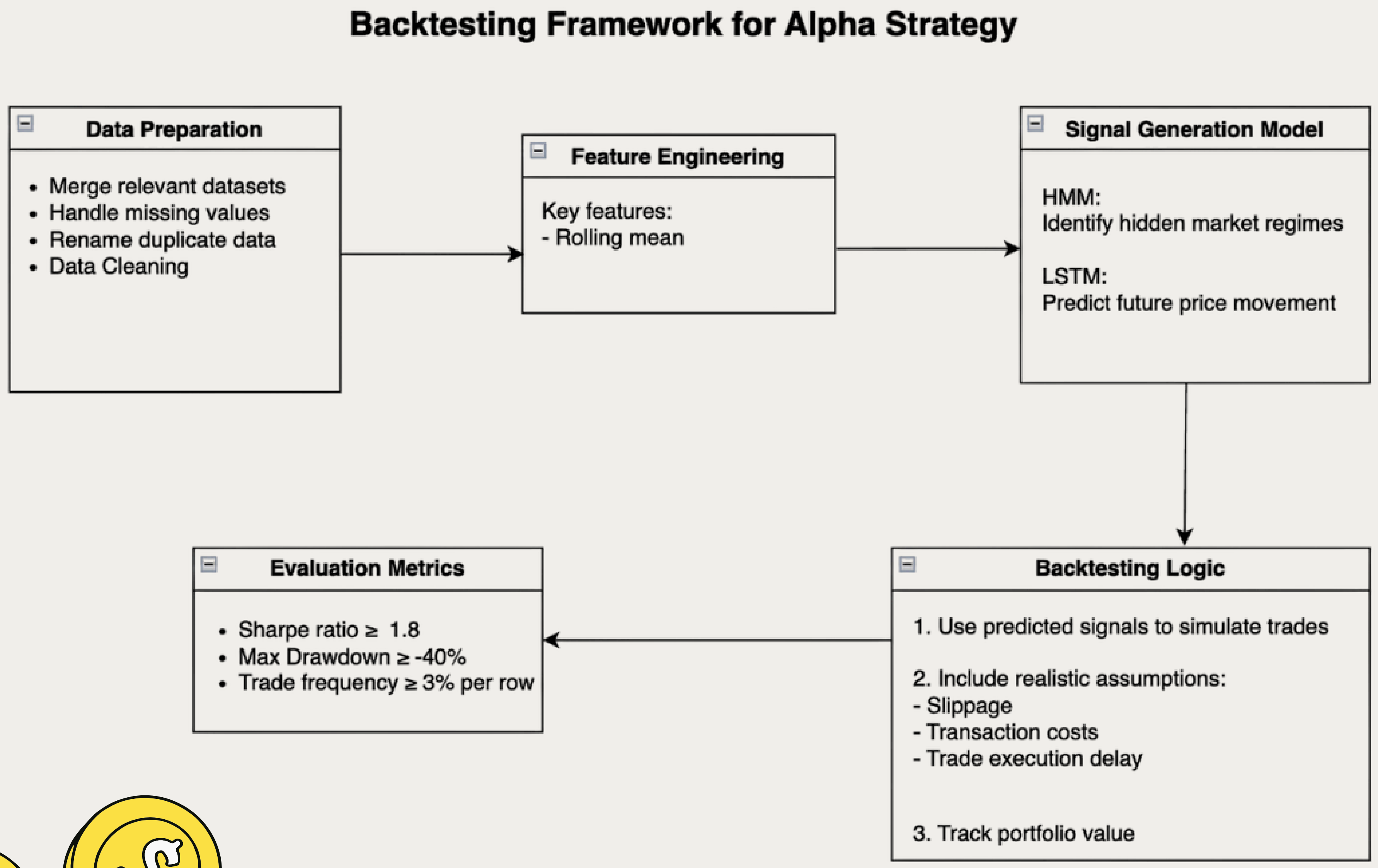


Long Short-Term Memory (LSTM)



LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) that is designed to learn long-term dependencies in sequential data such as time series or text. It uses special memory cells to store long-term dependencies, which makes it ideally suited for tasks like time-series forecasting, language modeling, and health trend forecasting.

Backtest Plan



A backtesting framework is a structured setup used to ***test trading strategies*** using historical data to evaluate how data would have performed in the past.

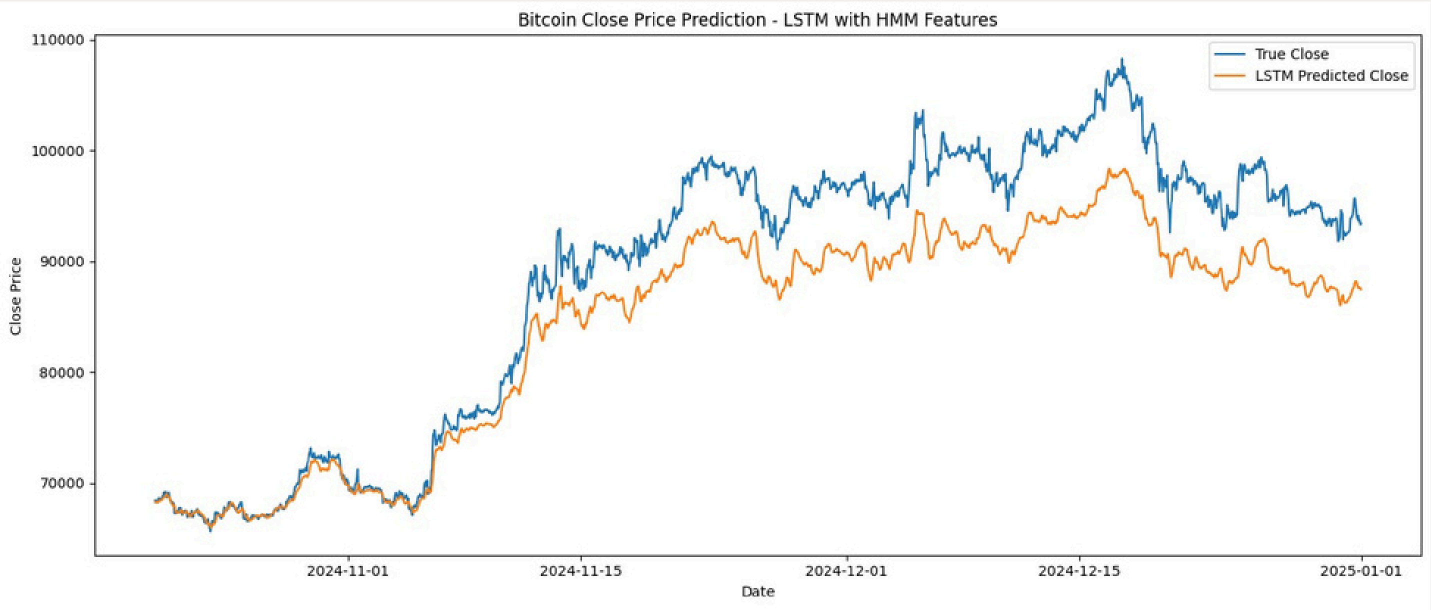
Build a model with HMM, and LSTM to generate buy/sell signals for Bitcoin trading:

- **Data Preparation:** Merged data by datetime, cleaned, normalized
- **Feature Engineering:** Focused on miner outflows, and exchange flows
- **Signal Generation:**
  - **HMM:** Detects market regimes
  - **LSTM:** Predicts future price direction
- **Backtesting:** Simulated trades with slippage, fees, and delay

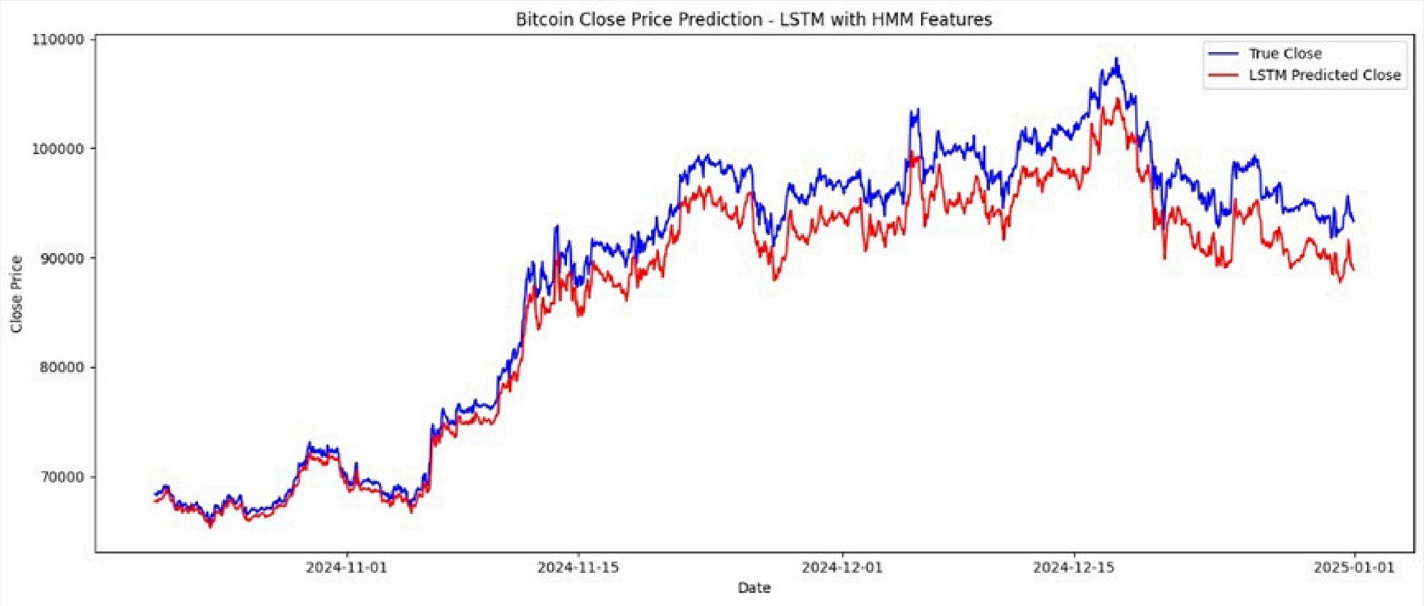
Evaluation

We use HMM with LSTM model to predict trading signal such as buy, hold and sell signal and proceed to backtest

Results of Model Training:

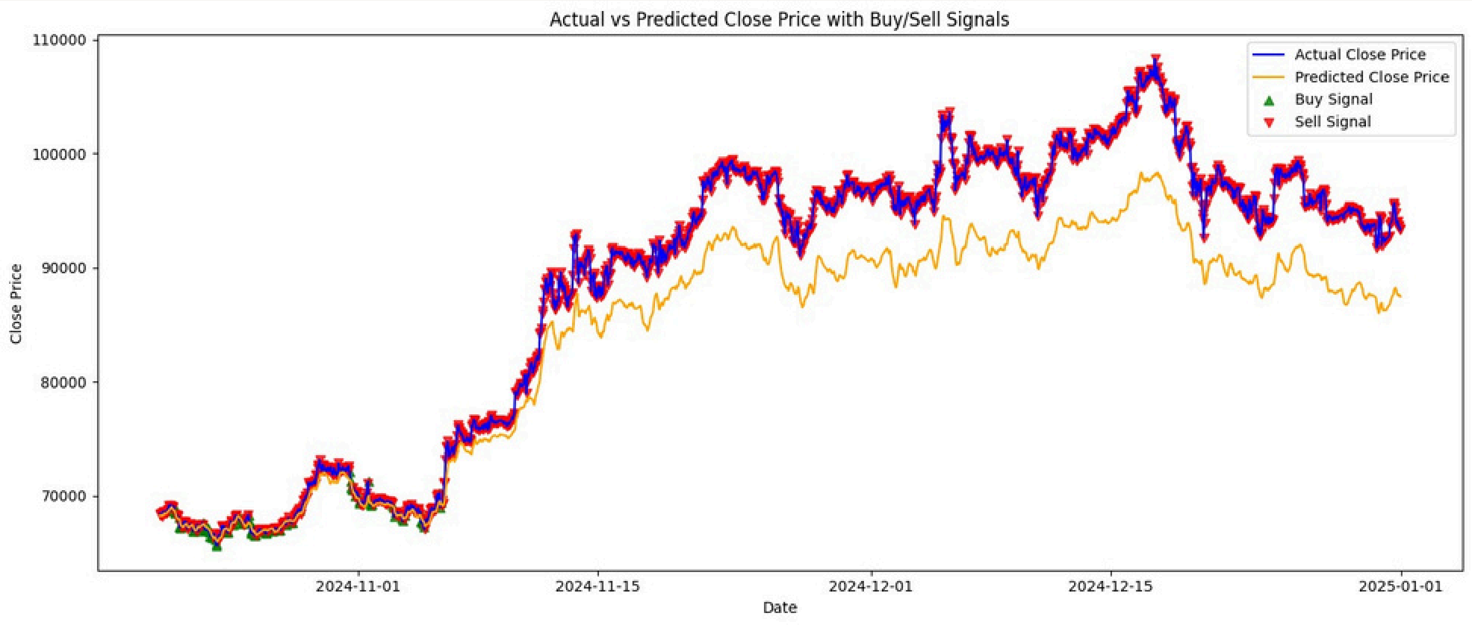


model run **25 times** of the entire training dataset:
Initial learning trends but may have higher prediction errors with less refined signals



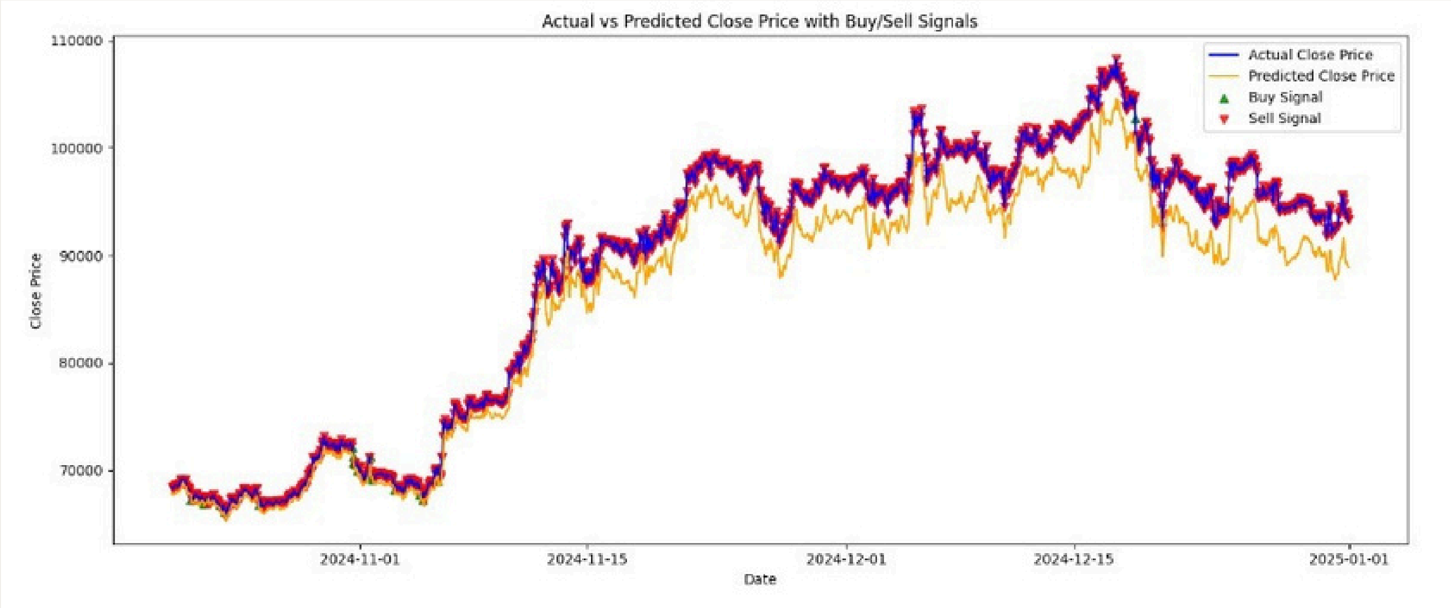
model run **50 times** of the entire training dataset:
Improved alignment between actual and predicted prices, suggesting better model accuracy with more training

Results of Buy Sell Signals:



model run **25 times** of the entire training dataset:

- Sharpe ratio: -0.60
- Max Drawdown: -29.37%
- Trade frequency: 30.55%



model run **50 times** of the entire training dataset:

- Sharpe ratio: 0,95
- Max Drawdown: -8.80%
- Trade frequency: 14.03%

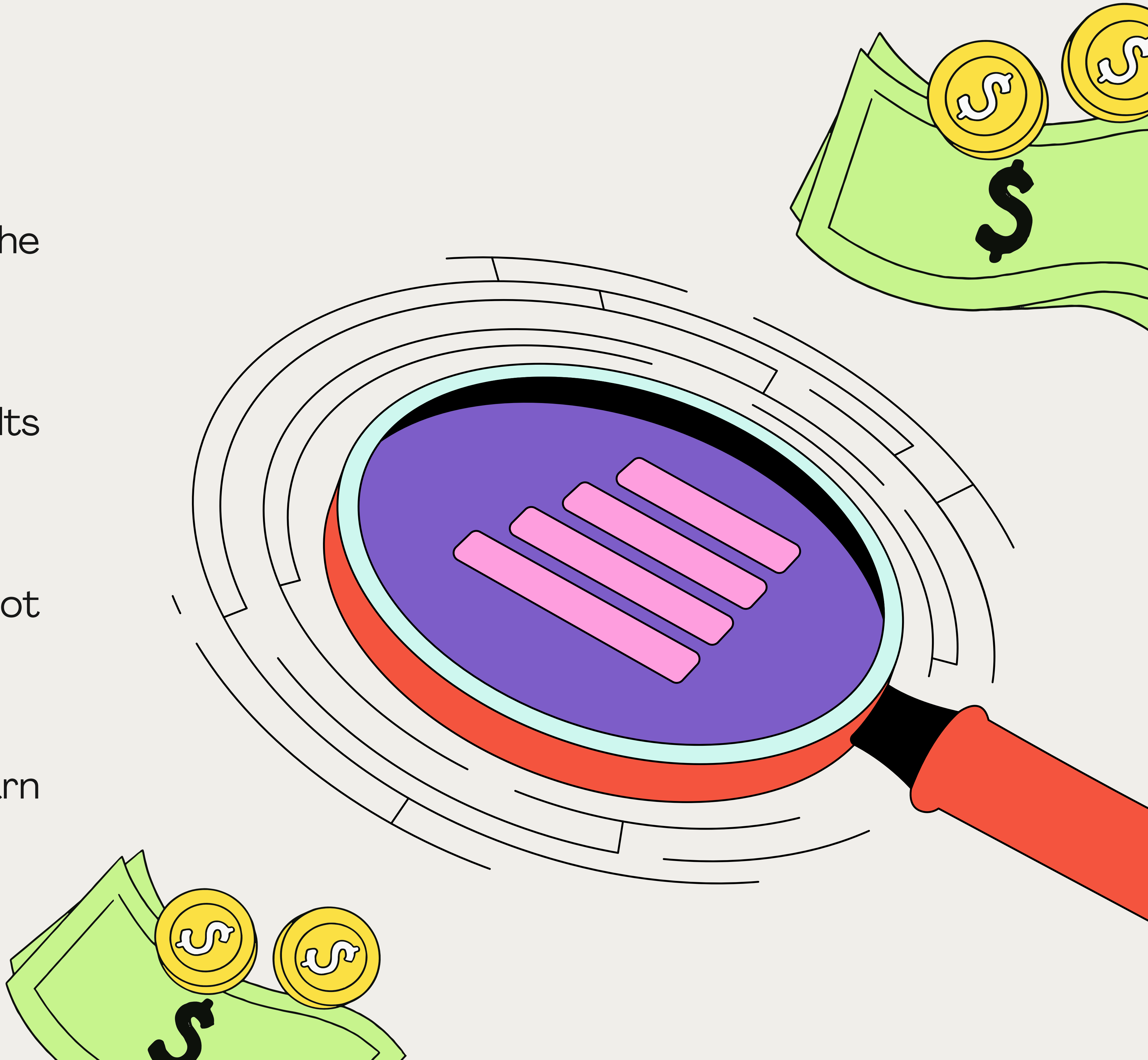
Conclusion

We use a combination of HMM and LSTM to predict the buy-sell signal and observe that:

When we trained the model **50 times**, it gave better results and more accurate buy-sell signals.

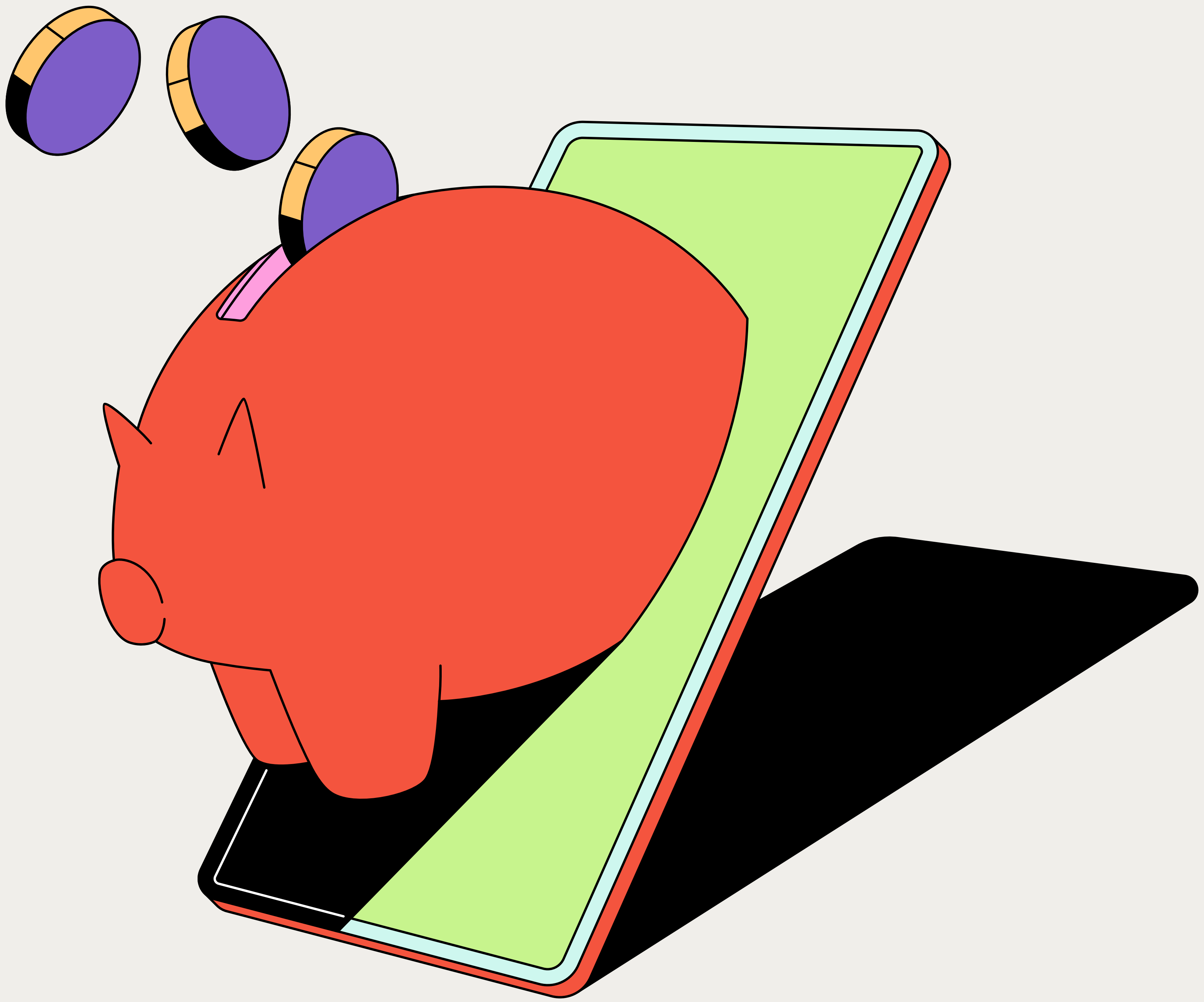
But when we trained it only **25 times**, the signals were not as good.

This means that training the model more times helps it learn better and make smarter predictions.



HOME ABOUT CONTENT OTHERS

THE END



Thank You For Your Attention