ON-CHAIN DATA TRADING MACHINE

GROUP 177 : ROJAK UNI
DOMAIN : 2

UM Hackathon 2025

Universiti Kebangsaan Malaysia
The National University of Malaysia

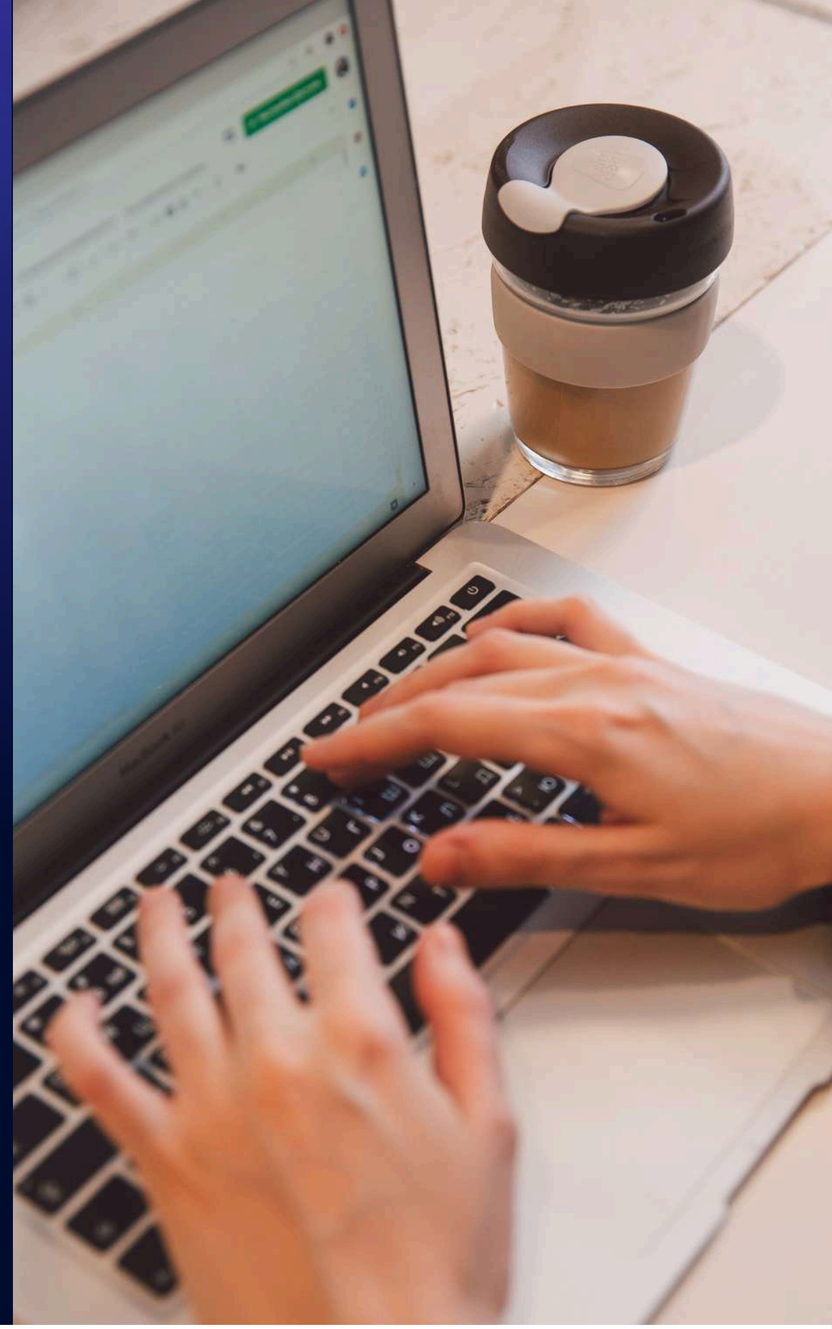
MONASH University
MALAYSIA

UNIVERSITY OF MALAYA

Balaena Quant.

IDEA REALISATION

RETRIEVE KEY DATA → HMM MODEL BUILDING → NLP MODEL BUILDING → INTE-GRATION → GENERATE ACTION → BACK-TESTING



WHY USE HOURLY ON-CHAIN-DATA?

Capture Short-Term Market Signals

- Detect rapid changes in behavior, like sudden whale transactions or spikes in gas usage.
- React to arbitrage opportunities that only exist for minutes or hours.
- Respond to market sentiment shifts faster than with daily data

More detailed Alpha Strategies

- Hourly data gives us more datapoints, which helps with training and pattern recognition.
- It enables feature engineering
- Enables model intraday volatility, flow patterns, or specific wallet behaviors.

Front-Running & MEV Awareness

- Hourly data gives us a clearer timeline of events.
- Not missing microstructure dynamics

FEATURE ENGINEERING

Features Engin...

Future Improvement

Features Engineering

- 📌= Category
  - 👉 = Features

Objective: "Enhance the input DataFrame by calculating and adding various features derived from existing on-chain metrics (inflow, outflow, transaction counts, etc.) to provide potentially predictive information for analysis or modeling."

- 📌Inflow / Outflow ratios
  - 👉inflow_outflow_ratio
    - Calculation: "Total Inflow / Total Outflow"
    - Purpose: Indicates the balance of coins moving onto exchanges (potential selling pressure) versus off exchanges (potential buying/HODLing pressure). Extreme values or sharp changes could signal shifts in market sentiment or supply dynamics.
    - Ratio > 1 suggests more inflow, while < 1 suggests more outflow.
    - Potential signal:
      - High values might precede price drops; low values might precede price increases.

ENSEMBLE MODEL

XGBoost

Logistic Regression

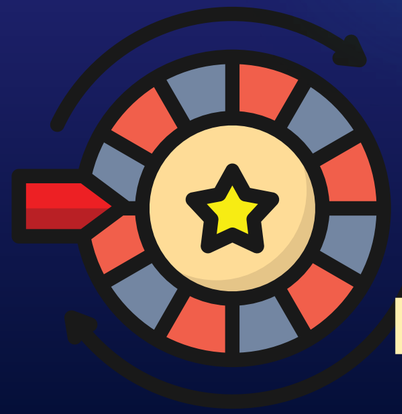
Random Forest

MLP

What they do?

- Each ML model provide its prediction validation accuracy
- Based on the prediction validation accuracy, calculate their emsemble weights
- Using emsemble weights to predict price direction

ENSEMBLE LOGIC



Provides probability of prediction of price rising up

If prediction>0.55 indicate bullish market

If prediction<0.45 indicate bearish market

Calculate amount to invest based on position size

HMM MODEL BACKBONE

Features input for Visualisation

What we tried after preliminary round?

Loop through distinct features combinations

Do simple testing and training datasets analysis

Outcome & Decision

- **Improve consistency of data and prevent unpredictable situation**
- **Taking base combination to minimize noise (flow mean, flow total, transaction count flow) - sticking to the one we used in prelimary round**
- **Plotting features against time interval and feature correlation matrix for numerical representation of features relationship**

2. Optimising Model Selection

01

Model selection

1. Statistical approach :

- Fit the best feature combination to AIC BIC

2. Faced a problem :

- The BIC and AIC score get lower while state increases, so the optimal states of regime may not work as expected, manual assign

02

Regime Classification, Distribution plots

1. Statistically backed regime classification for data

2. 5-year time series datapoint Visualisation

- Enable us to easily identify data with extreme flow means
- Better understanding of regime characteristics

03

Summary Metrics for Further Regime Characteristic Identifications

1. Get mean ,min and max

2. Calculate stds to understand each regime better

3. Regime Transition Handling

01

Keeptrack regime transition by count

1. Get the transition counts , where system moves from one regime to another.

2. Convert to probabilities:
- Transition probabilities a->b = counts b / total counts

02

Convert to regime transition probability

Similar to counts, easier to use for probabilistic modelling or as transition matrices in HMMs

03

Summary Metrics Identify Regime Characteristic Identifications

1. Each regime is characterized by statistical properties:
- Stability, Duration and Frequency

2. Enables semantic understanding, the representation of the regime's plain text meaning

4. Transition Precision Simulation

1. Train and Test Set Splitting from Original Datasets

2. Visualisations to indicate precision

3. Regime Prediction Accuracy

4. Regime Transition Detection Accuracy

NLP PIPELINE

1. RETRIEVE TEXT DATA FROM API KEY / DATASET

2. STORE RAW DATA

CSV

3. IMPORT AND LOAD CSV USING PANDAS

4. CLEAN AND PREPROCESS TEXT

5. FEED DATA INTO MODEL AND TRAIN USING THE DATA

6. APPLY MODEL TO NEWS ARTICLES AND MAKE PREDICTIONS

IN-DEPTH IDEA

(FEEL FREE TO EXPLORE OUR PRELIM SLIDES)

UM Hackathon 2025 | Universiti Kebangsaan Malaysia The National University of Malaysia | MONASH University MALAYSIA | UNIVERSITY OF MALAYA | Balaena Quant.

HMM IMPLEMENTATION (SIGNAL)

Regime-Based Implementation: Key Points

- Simple Regime MappingMaps regimes 0-5 to adjustment factors (0.5-2.0)
- Higher values for high-flow regimes, lower for low-flow regimes
- Trading AdjustmentsModifies signal thresholds based on current regime
- Scales position sizes up/down according to regime
- Adapts sell percentages inversely to regime aggressiveness
- Code IntegrationUses existing 'current_regime' column in data
- Minimal changes to strategy implementation
- Adds regime visualization to backtest charts
- Safety MechanismsCaps position sizes and sell percentages within safe limits
- Maintains core strategy logic while adding regime adaptivity

BACKTEST DEMONSTRATION

UM Hackathon 2025

Universiti Kebangsaan Malaysia
The National University of Malaysia

MONASH University
MALAYSIA

UNIVERSITY OF MALAYA

Balaena Quant.

CONTRIBUTION ON BUSINESS SUCCESSES

COMPETITIVE EDGE

Uncovers implicit market indicators faster than competitors, leading to early-mover advantages in trading

ENHANCED PATTERN RECOGNITION

HMMs improve detection of hidden market trends and helping identify profitable opportunities not visible through traditional analysis.

BETTER REGIME DETECTION

Recognizes shifts between bull, bear or sideways markets allowing adaptive strategies to reduce losses and capitalize on trends.

RISK MITIGATION

Adaptive models adjust to changing conditions to reduce exposure to volatile or unfavorable market regimes.

OPTIMIZED TRADING SIGNALS

Machine learning refines buy/sell triggers using extracted features to increase accuracy and maximizing returns.

TARGET AUDIENCE

01 **Quantitative Hedge Funds & Asset Managers**

02 **Proprietary Trading Firms**

03 **Institutional Investors (Pension Funds, Sovereign Wealth Funds)**

03 **High-Frequency Trading (HFT) Firms**

05 **High-Frequency Crypto Traders**

06 **FinTech & Algorithmic Trading Startups**

WHY CHOOSE US?

	OUR PRODUCT	TOKEN METRICS	DELPHI DIGITAL	SENTIMENT
REGIME DETECTION	• 74.83% of prediction accuracy test regime • 67.83% of prediction accracy training regime • 99.9% of regime transition detection	• Provides AI-driven trading signals and alerts • Does not publicly disclose specific accuracy metrics for regime detection	• Focuses on research and analysis • Regime detection accuracy metrics are not specified	• Utilizes machine learning for behavioral and on-chain indicators but does not provide specific accuracy figures for regime detection.
DATA SOURCES AND INTEGRARION	• Integrates on-chain data sources like CyboTrade	• Employs over 80 data parameters for its AI algorithms, though specific data sources are not detailed.	• Provides in-depth research, potentially utilizing various data sources, but specifics are not disclosed.	• Focuses on behavioral and on-chain data but lacks clarity on integration with external data providers.
SIGNAL FREQUENCY	• Delivers real-time signals with a frequency of at least 3% trade signals per data row via API.	• Provides real-time AI trading alerts through Telegram and Discord channels.	• Focuses on periodic research reports rather than real-time signal delivery	• Provides behavioral and on-chain indicators. • Real-time signal delivery specifics are unclear.
SUBSCRIPTION -BASED	• Basic (With HMM): Free • Premium (With NLP and LSTM): $19.99/month	• Basic: Free • Advanced: $39.99/month • Premium: $199.99/month	• Delphi Pro: $190.00/month • Premium: $999.00/month	• Sanbase: $44.00/month • SanAPI: $149.00/month

Not a direct offer. They sell raw data, not model-based insights.

UM Hackathon 2025 | Universiti Kebangsaan Malaysia The National University of Malaysia | Monash University Malaysia | University of Malaya | Balaena Quant.

THANK YOU

Here are the things that need to be done

QNA TIME!

Ask us anything!