Learning about Bitcoin, Stablecoin, and prediction

Arvind Kandala, Claire Oh, Yuchien Wang, Tully Cannon

Cryptocurrency & Circle

- The blockchain is a decentralized ledger to record all cryptocurrency transactions
- Mining is how transactions are grouped into blocks and added onto the chain
- Most of the cryptocurrency market value is contained within a few tokens

Regulations & Compliance

- GENUIS Act
  - Signed into law on July 18, 2025
  - first bipartisan federal law in US that directly regulates payment stablecoins.
  - Key rules:
    - Only licenced banks or approved non-bank issuers
    - 100% reserves in cash
    - Monthly reserve disclosures +audits
    - AML/KYC +consumer protection
- Before 2025
  - stablecoins rules were unclear
  - State money transmitter laws
  - FinCEN, Money Services Businesses (MSB) registration
  - Securities and Excahnge Commission (SEC) and Commodities Features Trading Commision (CFTC) sometimes stepped in if products looked like securities or derivatives
-

How GENUIS Impact Circle

- Get benifits from GENUIS Act
- Regulatory clarity
- Competitive advantage from others
- Better trust from institutions

Outlook for Stablecoins and Circle

- Stablecoins shifting from speculative assets to insititutinoal grade infrastructure
- Adoption expanding into
  - Corporate treasury
  - Cross-border settlement
  - Decentralized finance
- Functioning as programmable, digitally native dollars
- Integration accelerating under 2025 regulatory clarity

Market Growth

- Some forecasts predict multitrillion-dollar stablecoin market by 2028
- J.P. Morgan’s outlook:
  - Adoption still limited among institutional liquidity users
  - Payment use cases developing slowly
  - Market cap around $500B-$700B
- Growth likely steady but not explosive

Structural Impact on the Dollar System

- Growth will reinforce:
  - Dollar demand globally
  - U.S. dollar’s role as settlement currency of the internet
- Stablecoins supplement fiat rails rather than replacing them
- Firms with transparent, fully reserved models are best positioned

Circle’s Q3 Financial Performance

- USDC circulation: $73.7B (+108%YoY)
- Total revenue + reserve income: $740M (+66% YoY)
- Net income: $214M (+202% YoY)
- Adjusted EBITDA: $166M (+78% YoY)
- EBITDA margin: 57%

Circle’s Business Model and Infrastructure Expansion

- Narrow-bank economics
  - Fully reserved assets
  - Earns yield; pays no interest on USDC
- Strategic build-out
  - CCTP for cross-chain transfers
  - Programmable wallets
  - Arc for tokenized financial workflows

What Circle Staff Data Science Do?

- Analyze complex on-chain data: transaction graphs, wallet behavior, cross-chain flows, liquidity networks
- Model token distribution, detect anomalies, evaluate protocol-level risks
- Monitor stablecoin circulation, reserves, and macro-driven behavior
- Require 6+ years DS experience, including 3+ years blockchain specialization
- Use blockchain analytics platforms: Chainalysis, TRM Labs, Arkham, Elliptic
- Maintain on-chain address labels/tags and perform cross-chain transaction tracing
- Build real-time monitoring systems for USDC across multichain ecosystems
- Support products like CCTP, programmable wallets, and new chain integrations

Skills, Challenges, and Professional Context

- Apply ML, graph theory, stochastic modeling, anomaly detection, and forecasting
- Support regulatory, risk, and compliance needs (MiCA, US stablecoin legislation)
- Create AML monitoring, reserve transparency dashboards, and market-stress simulations
- Compensation: $130k–$205k+ with equity; reflects niche expertise
- Different from traditional finance DS roles (structured datasets, stable frameworks)
- Must learn smart contract mechanics, evolving chains/protocols, L2 ecosystems
- Highly cross-functional collaboration with researchers, product, compliance, leadership
- Work sits at the intersection of cryptography, economics, data science, and regulation

Technical Project

Introduction and Motivation

- Voliatility reason
  - Demand and Supply
  - Market sentiment (sensitive to media)
  - Big difference with traditional assets
- Three models
  - Logistic Regression, Random Forest, LSTM

Data Sources and Acquisition

- Daily OHLCV (Open, High, Low, Close, Volume) data
  - from 11/20/2014 - 11/29/2025 for BTC-USD
  - Yahoo Finannce through Python package
- Sentiment Data
- On-Chain data source
  - 2023–2025
  - Google BigQuery
  - Metrics included:
    - Daily transaction count
    - Active addresses (unique transacting addresses)
    - Exchange netflow
    - Inflows minus outflows across major exchanges (e.g., Binance, Bitfinex)
    - Exchange reserves
    - Total BTC held in exchange wallets

Target Variable and Evaluation

- All models predict the sigh of the next day lag returns

$$y_t = \text{sign}\left(\ln\left(\frac{P_{t+1}}{P_t}\right)\right), \quad y_t \in \{0, 1\}$$

- y_t =1 indicate an up day
- 80% train and 20% test

Common Indiactors

- Returns and Volatility
- Moving Averages and Crossovers
- Momentum Indicators
- Osilliators
- Volume-based Indicators
- Lagged Returns

Logistic Regression

$$P(y_t = 1|\mathbf{x}_t) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}^\top \mathbf{x}_t)}}$$

- Accuracy: 51.1%

Table 1: Logistic Regression Performance

Class	Precision	Recall	F1	Support
Down (0)	0.498	0.894	0.639	385
Up (1)	0.602	0.152	0.242	409
Accuracy		0.511		794

Random Forest

- 100 decision trees via bootstrap
- n_estimators=100, MAX_features= log2, min_samples_leaf =1
- On-chain weight: wi = λN −i, λ = 0.995

Table 2: Random Forest Performance

Training Start	Base Accuracy	On-Chain Accuracy	RMSE (Base)
2018-06-01	46.25%	—	0.0227
2023-01-01	45.35%	44.19%	0.0274
2023-06-01	43.66%	43.66%	0.0292
2024-01-01	51.02%	44.90%	0.0345
2024-06-01	47.06%	47.06%	0.0341
2025-01-01	61.54%	53.85%	0.0257

LSTM Neural Network

- Processes 20-day sequences 13 features (as input layer)
- Layers:
  - input
  - LSTM with 32 unites
  - dropout =0.3
  - dense with 16 unites with ReLU activation
  - dropout =0.3, dense =1

Table 3: LSTM Performance (Epoch 22, Early Stopped)

Dataset	Loss	Accuracy
Training	0.6891	52.32%
Validation	0.6920	50.87%
Test	0.6969	47.48%

Cross-Model Analysis

Table 4: Model Comparison Summary

Model	Best Accuracy	Avg Accuracy	Training Time
Logistic Regression	51.1%	51.1%	less than 1 sec
RF (Base)	61.5%	47.5%	approximately 10 sec
RF (On-Chain)	53.8%	46.7%	approximately 15 sec
LSTM	47.5%	47.5%	approximately 2 min
Random Baseline		50.0%	

Bitcoin Quant Strategy

Methodology

Objective: Develop a rule-based trading strategy for the Bitcoin spot market.
Data: Daily BTC–USD prices from **30 Nov 2014 to 30 Nov 2025**.
Target variable: Daily log return of closing price

Methodology

Predictors: Up to 5 lags of past log returns.
Model: Single-hidden-layer neural network (PyTorch).
Output: Next-day log return r^t+1r^t+1.
Training: Fixed random seed (42) for reproducibility.

Feature Stages

Stage 1 – Price-based Rules: Only lagged returns
Stage 2 – Market Condition Features: Add volume, volatility, liquidity, momentum
Stage 3 – On-chain Information Integration: Add active addresses, new addresses

Trading Strategy Evaluation

Stage 1 – Price-based Rules

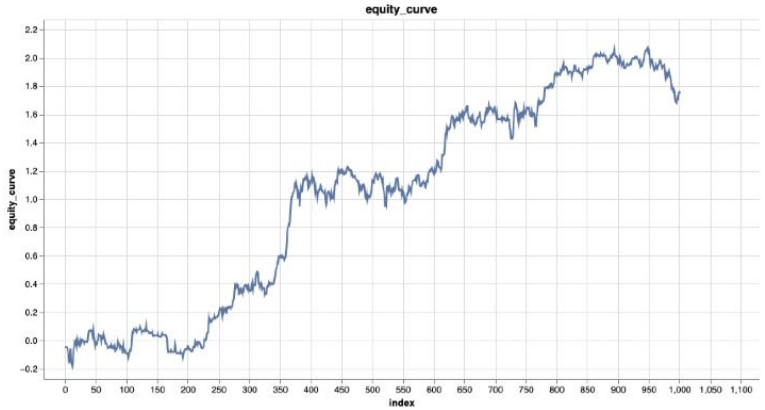

Figure 1: Effect of Equity Carve-Out Signals on Close Log Return (Lag 1)

Table 5: Price-Based Lagged Return Performance

Feature	Win Rate	Expected Value (EV)	Sharpe	Max Drawdown
Close_log_return_lag_1	0.5125	0.0014	0.9378	-0.3856
Close_log_return_lag_2	0.5055	0.0013	0.8773	-0.3878
Close_log_return_lag_3	0.5045	0.0012	0.7913	-0.4522
Close_log_return_lag_4	0.5055	0.0013	0.8773	-0.3878
Close_log_return_lag_5	0.5055	0.0013	0.8773	-0.3878

Trading Strategy Evaluation

Stage 2 – Market Condition Features

Table 6: Top 5 Performance Metrics for Selected Feature Combinations

Features	Win Rate	EV	Sharpe	Max Drawdown
Close_log_return_lag_1, bb_high, bb_low	0.6337	0.008505	5.9698	-0.118534
Close_log_return_lag_1, ROC_7d, bb_low	0.6297	0.008342	5.8396	-0.118536
Close_log_return_lag_1, Mom_7d, bb_high	0.6206	0.008171	5.7041	-0.133623
ROC_7d, bb_high, bb_low	0.6165	0.007857	5.4589	-0.106730
Close_log_return_lag_1, ROC_7d, bb_high	0.6095	0.007840	5.4453	-0.133623

Trading Strategy Evaluation

Stage 3 – On-chain Information Integration

Table 7: Top 5 Performance Metrics for Selected Feature Combinations

Features	Win Rate	EV	Sharpe	Max Drawdown
Close_log_return_lag_1, bb_high, bb_low	0.6276	0.008514	5.9773	-0.124473
ROC_7d, bb_high, bb_low	0.6085	0.008043	5.6035	-0.106730
Close_log_return_lag_1, Mom_7d, bb_high	0.6095	0.007925	5.5117	-0.133623
Close_log_return_lag_1, ROC_7d, bb_high	0.6065	0.007905	5.4959	-0.133623
Close_log_return_lag_1, ROC_7d, RSI_14d	0.6236	0.007804	5.4177	-0.168461

Limitations

- Bias
- Data quality and outlier
- Model limitation
- No cross-validation

Recommendation and Conclusion

- Bitcoin’s daily directional movement may be fundamentally unpredictable from historical data alone.
- Price behavior is heavily influenced by exogenous news events, not past patterns.
- Future work should explore:
  - Higher-frequency data to capture market microstructure
  - Regime-aware models to address non-stationarity
  - Alternative forecasting tasks (e.g., volatility, multi-horizon predictions)
  - Methods that exploit different timescale patterns
- Model evaluation should prioritize economic metrics, not accuracy:
  - Sharpe ratio, maximum drawdown, transaction-cost effects
  - Even ~55% accuracy can be unprofitable once trading costs are included