

Junior Data Scientist – Trader Behavior Insights Report

Candidate: Hrutik Adsare

1. Executive Summary

This analysis identifies and quantifies a strong predictive relationship between on-chain trader behavior, market sentiment, and trade profitability. By engineering novel features that capture a trader's recent state and its interaction with the Bitcoin Fear & Greed Index, a machine learning model was developed that predicts trade profitability with ~90% accuracy (ROC-AUC: 0.96). A trader's recent performance is the primary driver of future success, while market sentiment acts as a significant modulator.

2. Methodology Overview

2.1 Data Sources & Preparation

- Historical Trader Data (Hyperliquid): 211,224 trades with features like account, size, closedPnL, side.
- Bitcoin Fear & Greed Index: 2,644 daily sentiment records.
- Data Cleaning: Standardized timestamps, handled missing values, filtered erroneous points.

2.2 Feature Engineering

Created rolling trader behavior features (win rate, average size, PnL volatility, activity level), market sentiment trends (lag, 3d mean, 7d mean), and interaction features such as sentiment x trade size and sentiment x recent win rate.

2.3 Modeling & Validation

Used time-based split (80/20) to prevent look-ahead bias. Compared Logistic Regression, Random Forest, LightGBM, and XGBoost. Evaluated using Accuracy, Precision, Recall, F1-Score, ROC-AUC.

3. Key Results & Insights

3.1 Model Performance

Metric	Score	Interpretation
Accuracy	0.898	Correct 89.8% of the time
Precision	0.890	Profit prediction correct 89% of the time
Recall	0.828	Captured 82.8% of profitable trades
F1-Score	0.858	Balanced precision and recall
ROC-AUC	0.963	Excellent discrimination ability

3.2 Feature Importance

Top drivers of profitability:

1. Recent win rate (roll50_winrate)
2. Recent PnL volatility (roll50_pnl_std)
3. Interaction between sentiment and win rate
4. Trading frequency (roll50_intertrade_mean)
5. Interaction between sentiment and trade size

4. Actionable Insights & Strategic Recommendations

- Develop a 'Trader Health' Score using rolling metrics.
- Avoid copying erratic traders with high volatility.
- Incorporate sentiment-aware strategies for dynamic position sizing.
- Monitor regime shifts in trader behavior for timing strategy changes.

5. Conclusion

This project shows trader behavior is predictable, sentiment modulates trader actions, and actionable signals exist within on-chain data. The Random Forest model achieved 0.963 ROC-AUC, providing a strong foundation for building data-driven trader evaluation and risk management tools in Web3.