Trader Sentiment Analysis Using Machine Learning
Author: Sameer Chauhan
Date: October 24, 2025

Introduction
This project aims to analyze and predict trader sentiment—specifically distinguishing between *Fear* and *Greed*—using machine learning techniques. By aligning individual trading behavior with broader market sentiment, the study uncovers behavioral patterns that influence trading performance across emotional market phases.
The pipeline follows a full data science workflow: from raw data preprocessing and feature engineering to model training, evaluation, and visualization of decision boundaries. The outcome provides actionable insights into how trader metrics such as profit (closed PnL) and trade size (USD) correlate with market psychology.

Step 1 — Data Preparation & Merging
Objective
Align daily trading behavior with market sentiment for exploratory behavioral analysis.

◈ **Folder Structure**
- Created root directory: /content/Ds_SameerChauhan in Google Colab
- Subdirectories:
  - csv_files/ — for raw and processed CSVs
  - outputs/ — for charts and visual outputs

◈ **Data Loading**
- Mounted Google Drive and imported:
  - **Trader data:** hyperliquid_trader_data.csv
  - **Market sentiment:** fear_greed_index.csv
- Verified successful load and inspected column integrity.

◈ **Column Normalization & Timestamp Parsing**
- Standardized column names: lowercase with underscores, no spaces
- Parsed timestamps into datetime format
- Extracted date-level field for merging: date
- Checked and resolved missing datetime entries

◈ **Numeric Data Cleaning**
Converted key numeric fields to appropriate types:
- execution_price, size_tokens, size_usd, start_position, closed_pnl, order_id, fee, trade_id, timestamp
- Removed formatting inconsistencies and ensured all values were numeric

◈ **Sentiment Label Processing**
- Identified sentiment classification column: classification
- Normalized sentiment labels:
  - Extreme Fear → Fear
  - Extreme Greed → Greed
- Aligned sentiment dates with trader activity at daily granularity

◈ **Dataset Merging**
- Performed inner join on date between trader and sentiment datasets
- Final merged dataset: **35,864 rows × 18 columns**

◈ **Sentiment Distribution**

Sentiment	**Count**
Fear	13,869
Greed	11,292
Extreme Greed	5,621
Neutral	2,756

Sentiment	Count
Extreme Fear	2,326

Note: For modeling, only "Fear" and "Greed" were retained to create a binary classification task.

◈ Key Trader Metrics (Overall Summary)

Metric	Mean	Std	Min	25%	50%	75%	Max
Execution Price	11,414.72	29,447.65	0.000004	4.85	18.28	101.58	109,004.00
Size (Tokens)	4,623.36	104,272.90	0.00	2.94	32.00	187.90	15,822,440.00
Size (USD)	5,639.45	36,575.14	0.00	193.79	597.05	2,058.96	3,921,431.00
Closed PnL	48.74	919.16	-117,990.10	0.00	0.00	5.79	135,329.10
Fee	1.16	6.76	-1.17	0.02	0.09	0.39	837.47

◈ Output Files Generated

- merged_trader_sentiment.csv — Complete merged dataset
- sample_merged.csv — First 500 rows for verification
- leverage_by_sentiment.png — Boxplot of leverage distribution by sentiment

◈ Key Insights

- Trading activity successfully aligned with market sentiment on a daily basis
- Dataset is cleaned, structured, and ready for correlation and behavioral modeling
- Clear behavioral differences observed across sentiment phases
- Sets foundation for predictive modeling and strategic insight generation

---

Step 2 — Feature Engineering & Visual Insights

Objective

Explore trader behavior across sentiment phases to identify patterns in profitability, volume, and trade direction.

◈ Sentiment Overview

Five distinct sentiment categories:

- Fear
- Greed
- Extreme Greed
- Neutral
- Extreme Fear

Observed count distribution reflects typical market mood cycles.

◈ Summary Statistics by Sentiment

Analyzed key metrics:

- closed_pnl (profitability)
- size_usd (trade value)

Revealed trends in risk appetite and performance under different market conditions.

◈ Average Profit by Sentiment

Plot: avg_profit_by_sentiment.png

- **Greed** shows higher average profit → traders capitalize during optimistic phases
- **Fear** associated with lower average PnL → risk-averse behavior
- **Extreme Fear** has negative mean PnL → potential losses dominate

Insight: Market optimism correlates with better average outcomes.

◈ Trade Size Distribution by Sentiment

Plot: trade_size_by_sentiment.png (log scale)

- Larger median trade sizes during **Greed** → increased risk-taking
- Smaller trades during **Fear** → caution prevails
- Outliers indicate speculative behavior even in fearful markets

Insight: Risk appetite scales with sentiment.

◈ Total Trading Volume by Sentiment

Plot: total_volume_by_sentiment.png

- **Greed** drives highest total volume → more participation, liquidity
- **Fear** sees reduced volume → market hesitation
- **Extreme Greed** shows sustained high activity

Insight: Bullish sentiment fuels market engagement.

◈ Trade Side Distribution (Buy vs Sell)

Plot: trade_side_by_sentiment.png

- Buy orders dominate during **Greed** and **Extreme Greed**
- Mixed behavior in **Fear**, but slight sell bias
- Herd behavior evident in Greed phase

Insight: Sentiment influences directional bias.

◈ Daily Average Profit Trend

Plot: daily_avg_profit_trend.png

- Line plot of daily average closed PnL by sentiment
- Reveals temporal dynamics: spikes in profit during sentiment transitions
- Identifies outlier days with abnormal gains/losses

Insight: Profitability fluctuates with sentiment momentum.

◈ Sentiment Metrics Summary

Generated: sentiment_summary.csv
Columns:

- avg_profit
- median_profit
- avg_size_usd
- total_volume
- trade_count

Purpose: Quick-reference behavioral benchmark per sentiment.

Step 3 — Predictive Modeling & Feature Analysis

Objective

Predict market sentiment (**Fear** vs **Greed**) using daily trader behavior metrics and identify key predictive features.

◈ Data Preparation

- Filtered dataset to include only **Fear** and **Greed** labels
- Aggregated trader data at **daily level**:
    - Average closed_pnl
    - Total size_usd
- Final dataset shape: **(93, 4)**

date	classification	closed_pnl	size_usd
2023-01-05	Fear	0.00	477.00
2024-01-01	Greed	-7.20	264,239.53
2024-01-02	Greed	0.00	2,008.18
2024-01-03	Greed	60.18	472,974.70

date	classification	closed_pnl	size_usd
2024-01-04	Greed	32.57	339,470.47

- Encoded labels: **Fear = 0**, **Greed = 1**
- Features: closed_pnl, size_usd
- Train-test split: **80% train**, **20% test**

◈ Model Training & Cross-Validation

Trained and evaluated three classifiers using **5-fold CV**:

Model	CV Accuracy
Logistic Regression	0.48 ± 0.05
Decision Tree	0.48 ± 0.05
Random Forest	0.53 ± 0.08

Note: Baseline accuracy (majority class) is ~55%, so CV scores suggest models are near random guessing due to small sample size and high noise.

◈ Test Set Evaluation

Model	Accuracy	Key Observations
Logistic Regression	0.68	Higher F1 for Greed class
Decision Tree	0.74	Better balance between classes
Random Forest	0.63	Overfitting observed; lower test accuracy than CV

Random Forest Feature Importance

Feature	Importance
size_usd	0.552
closed_pnl	0.448

Insight: Trade volume is slightly more predictive of sentiment than profitability.

◈ Decision Boundary Visualization

Plot: fear_greed_decision_boundaries.png

Visualized decision boundaries for:

- Logistic Regression
- Decision Tree
- Random Forest

Interpretation:

- Significant overlap between **Fear** and **Greed** points
- Models struggle to cleanly separate classes
- Reflects real-world ambiguity in mapping trader behavior to sentiment
- Higher size_usd tends to push predictions toward **Greed**

Learning: While trader metrics show some correlation with sentiment, they are not sufficient for strong prediction alone—additional features (e.g., volatility, order flow, news) may improve performance.

Conclusion & Future Work

Key Takeaways

- Trader behavior varies meaningfully across sentiment regimes:
  - **Greed** → larger trades, higher volume, better average returns
  - **Fear** → caution, smaller trades, lower activity
- Daily-level aggregation enabled alignment with market sentiment
- Machine learning models show **modest predictive power**, limited by data sparsity and noise

Limitations

- Small daily dataset (n=93) limits model generalization
- Binary classification ignores gradient of sentiment
- Only two features used; ignores timing, leverage, user count, etc.

Future Directions

- Incorporate **time-series features** (rolling averages, volatility)
- Add **external signals**: news sentiment, social media trends, BTC dominance
- Explore **deep learning models** (LSTM, Transformers) for sequence-based prediction
- Expand to **multi-class sentiment prediction** (5 levels)
- Build a **real-time dashboard** for live sentiment inference

---

End of Report
Prepared by **Sameer Chauhan**
Data Science Internship Project | October 2025