Trading Strategy Analysis — Project Report

Name: Madhulika Chandel
Date: 4 October 2025
Dataset: Historical Trader Data and Fear & Greed Index
Goal: To investigate how traders' behavior interacts with market sentiment.

---

1. Project Overview

Analyze how trading behavior (profitability, risk, volume, leverage) aligns or diverges from overall market sentiment (fear vs greed). Identify hidden trends or signals that could influence smarter trading strategies.

---

2. Data & Preprocessing

I decided to merge the two datasets using the **date** as a key, so that each day's trading activity could be aligned with the corresponding market sentiment. The merging could be done in two ways:

Firstly by aggregating the trades for a single day (reduce data inflation) For simplification and easy understanding I kept the aggregated dataset for visualization and generated meaningful columns like daily leverage, volume, number of trades, buy ratio and used various plots to understand the data (refer to notebook_1.ipynb or outputs for the plots).

Secondly by introducing sentiment and value as new columns for each trade of the day, I used this merged dataset for model training. For preprocessing I dropped the NaN columns, applied one-hot encoding to categorical columns and generated hour and DayOfWeek from TimestampIST.

Also applied **Log transformation** to the target column because it was heavily skewed, stabilizing extreme values and improving model performance

---

3. Modeling Approach

I tried different regression models to predict the log-transformed PnL values including **HistGradientBoosting, XGBoost, LightGBM, and CatBoost**. To find the best-performing model and hyperparameters, I used **Optuna**.

Once the best model and hyperparameters were selected, I created a **Pipeline** containing all preprocessing steps and the trained model.

Further I created a post-processing step that converts predicted PnL into **binary profitable/unprofitable labels** using a threshold derived from ROC analysis, enabling profitability classification on the basis of PnL values predicted from the model.

---

4. Evaluation

I first evaluated the regression model on the log-transformed `Closed PnL` target. The best model achieved an **R^2 score of approximately 0.88** on the test set, indicating that it captures a significant portion of the variance in the skewed PnL distribution. But when the predictions were then back-transformed to the original PnL scale,model achieved an **R^2 score of approximately 0.62** on test set.

For profitability classification, a threshold was applied to the predicted PnL values to convert them into binary labels (profitable vs. unprofitable trades). On the full test set, this approach yielded an **accuracy of 90.3%**, with the **Confusion Matrix:**

[[22582 2266]

[1824 15572]]

---

5. Key Insights

- Sentiment alone was not strongly correlated with extreme PnL values.
- **Interactions between sentiment and leverage** revealed periods of both extreme profits and losses.
- Log transformation stabilized skewness and improved regression performance.
- Threshold tuning is critical — naive thresholds give misleading results.
- The model performs well as a filtering layer for identifying likely profitable trades.
- Users can also be classified into two categories — ‘Greedy’ or ‘Fearful’ — to provide more specific insights into trading behavior.

---

6. Future Improvements

- **Explore advanced models:** Implement **Balanced Random Forest** or other **classification-specific models** to better handle skewed distributions and imbalanced profitable/unprofitable trades.

- **Feature engineering enhancements:** Incorporate **feature interactions** and **temporal effects**, such as rolling volatility, moving averages, or lagged trade metrics, to capture dynamic market behavior.

- **Optimize for business objectives:** Use **cost-sensitive metrics** to prioritize high-precision detection of profitable trades, reducing false positives that could lead to unprofitable decisions.

- **Handle class imbalance:** Apply **SMOTE (Synthetic Minority Over-sampling Technique)** to balance profitable and unprofitable trade classes, improving classifier performance on minority classes.

---

7. Conclusion

The regression + thresholding pipeline achieved **$R^2 \approx 0.62$** and **90% classification accuracy** on the test data, demonstrating solid predictive power for skewed financial distributions. The approach is flexible and can be tuned further depending on trading strategy objectives.