Zerodha AIML overview

- Since Zerodha's ML models hasn't been published in a catalogue, I am going out on a limb with the ML models they have used. I am including all the ML models that geenral stock brokerage applications use, and what I understood from the available features.

- ▼ AIML that Zerodha LIKELY uses
  - ▼ Real time analytics and telemetry
    - They use ClickHouse database services for storage and running streaming pipelines for telemetry. It is important with respect to AIML as AIML needs fast and accurate materialization, especially for anything real time, with very low latency.
    - They use VWAP (Volume Weighted Average Price), which is the average price of security, weighted by the volumen traded at each price. Used to measure the "true" average price over a period.

$$VWAP = \frac{\sum_{i=1}^{n} P_i * Q_i}{\sum_{i=1}^{n} Q_i}$$

      where P_i = trade price at time i
      Q_i = trade volume at time i

    - They also use order-book summaries which is the list of all buy (bid) and sell (ask) orders for a stock, along with process and quantaties

$$Imbalance = \frac{Bid\,Volume - Ask\,Volume}{Bid\,Volume + Ask\,Volume}$$

    - They also use realised volatility, which is the volatility observed over a historical period, based on certain returns. r_i = log return in each small interval (interval depends on the production. Not sure of Zerodhas interval)

$$\sigma_{realised} = \sqrt{\sum_{i=1}^{n} r_i^2}$$

- They also use session counts, which is the distinct trading sessions or login sessions per user, per a certain time window. (Not sure of this usage. I believe that if they are using realised volatility, they shouls use session counts).

▼ Fraud etc detection

- General industry practice is:
    - rules → unsupervised learning for anomaly detection → graph ML for coordination.
- Used a basic Rules + score aggregation model.
- Used Isolation Forest, which isolates anomalies by random splits.
- Used Autoencoder or LSTM - Autoencoder. Has a danged of reconstruction error.
- Graph ML, as they are extremely effective at reducing false positives.
- Not explicitly mentioned, but I am confident they have used anomaly detection.

▼ Personalization and Recommendation

- Their baseline model is offline recommendations, for user interaction, recommended based on sessions.
- They are using session and sequence models (RNN and transformers) for next trade item prediction.
- Uplift modeling that targets user who wants incremental revenue.

▼ Customer churn

- Standard algos of XGBoost and LightGBM are used.
- Uplift algo used in Personalization and Recommendation.

▼ Where we can be better

- Using Graph first fraud detection (beats ML rule book implementation, meaning, will take more resources to execute). Can use GraphSAGE + temporal LSTM or GAT + temporal LSTM.
- LighGBM predictor + convex optimizer for immediate wins. Can be evolved overtime where we use RL.
- For platform learning, we can use incremental learners and RL to analyze behavior.
- We can use an LLM that can be combined with our base ML model for well structured data to the user, and for recommendations. On the developer side, it can help us with risk detection and anomaly prediction.