GitHub Repository and Execution Guide

All source code, datasets, Jupyter Notebooks, and execution instructions related to this project are available in the following GitHub repository:

🔗 **GitHub Link:**
https://github.com/Jimit177/AI_PREDICTOR

This repository contains:

- Jupyter Notebooks for EDA, model training, and evaluation
- Python scripts for running the backend and frontend logic
- Cleaned and labeled datasets
- Trained model files and configuration artifacts
- A Streamlit app for real-time directional predictions

📄 **Execution Instructions:**
All instructions for setting up the environment, running the EDA, performing statistical tests, and reproducing model predictions are provided in the repository's README.md file.

🧠 **Notebook LM Source:**
A link to the Notebook LM log, containing details of learning iterations and model results, is also included in the README.md.

1. Clarify the Problem

- **Use Case**: Reducing retail trading losses through AI support.

Retail investor participation in India's F&O (Futures and Options) market has reached approximately **9.6 million traders**. According to SEBI data, around **8.7 million** of them — or **90%** — incur losses annually. In an ideal scenario where platforms provide structured education, real-time guidance, and proper risk control, this loss rate could realistically reduce to **70%**, or **6.96 million traders**.

This creates a **measurable and actionable gap of 1.74 million traders**, who could potentially avoid losses through intelligent interventions.

2. Break Down the Problem

The **1.74 million traders** who represent the actionable gap between current and ideal outcomes can be broken down into three major root contributors, each supported by sub-causes and numeric estimates:

A. Lack of Knowledge – 1,040,000 traders

This is the most significant factor, primarily affecting traders who make incorrect directional decisions due to poor understanding of the market. It includes:

- **No Basic Financial Education – 522,000 traders**
  These traders have not undergone any form of structured financial onboarding.
- **No Trading Experience – 418,000 traders**
  Beginners lacking exposure to technical indicators, market cycles, or risk frameworks.
- **Social Media/Influencer Bias – 100,000 traders**
  Traders acting on unverified advice, misleading YouTube content, or Telegram signals.

B. Market Volatility – 435,000 traders

These losses stem from rapid price swings and unpredictable market conditions. It includes:

- **Economic Events – 217,000 traders**
  Impacted by news, inflation data, RBI decisions, or geopolitical triggers.
- **Poor Trade Timing – 135,000 traders**
  Late entries or exits without technical confirmation.
- **Operator/Big Player Moves – 87,000 traders**
  Price manipulation or large institutional trades not accounted for.

C. High Risk from Leverage – 261,000 traders

These are traders whose losses are worsened by excessive use of margin or aggressive positions:

- **Misuse of Margin – 130,500 traders**
  Over-leveraging without stop-losses or hedging.
- **Absence of Position Sizing – 87,000 traders**
  Taking random trade sizes without exposure control.
- **Greed-Driven Trades – 43,500 traders**
  Overconfidence, doubling down, or chasing losses.

🟢 **Point of Occurrence:**

All these sub-causes converge at a critical decision point:
"Deciding the Direction of the Market" — where most traders act without real-time analytical insight or AI support.

This is the key failure moment that this project targets through the implementation of intelligent, platform-integrated decision tools.

3. Set a Target

The primary objective is to reduce the **losses among the 1.04 million traders** who are impacted by a lack of foundational trading knowledge — a key contributor to incorrect directional trades.

To address this, the following **measurable targets** are set:

1. **Reduce knowledge-related losses from 1.04 million to 520,000 traders (a 50% reduction)**
   within the next **3 months**, by implementing:
   - AI-powered directional prediction tools
   - Mandatory onboarding with scenario simulations
   - Guided paper trading with model-based feedback
2. **Ensure that at least 80% of new F&O traders complete a mandatory onboarding module**
   on financial literacy and directional trading logic before being granted access to live trading, starting **Q3 2025**.
3. **Increase stop-loss compliance by 50% among new users**
   within 2 months, by embedding stop-loss fields as a **non-skippable input** in trade execution interfaces during the pilot implementation.

4. Analyze the Root Cause

1. **Why do traders make incorrect directional trading decisions?**
   → Because they act on impulse, social media tips, or intuition without structured analysis or predictive support.
2. **Why do they rely on emotion and unverified sources?**
   → Because they lack **real-time, AI-powered decision-support tools** that offer trade insights and confidence scores.
3. **Why don’t traders have access to such tools?**
   → Because most trading platforms in India **do not offer integrated AI guidance systems** for directional trading.
4. **Why don’t platforms offer AI-based guidance?**
   → Because there’s a **legacy belief** that discretionary trading is the norm, and that AI support may limit user autonomy.
5. **Why is this belief still embedded in platform design?**
   → Because the **value of AI in reducing cognitive bias and improving decision accuracy is underutilized** and not embedded as an industry standard.

Root Cause :

The **absence of AI-powered, real-time decision-support tools** in retail trading platforms forces traders to rely on unstructured judgment, leading to persistent directional errors and financial losses.

5. Develop Countermeasures

Based on the structured Objective Matrix, five potential interventions were evaluated across five key criteria: ROI, Ease of Use, Learning Impact, Accuracy, and Scalability. The top-ranked countermeasures best address the root cause of **poor directional decisions due to lack of intelligent support and structured learning.**

---

1. AI Trade Assistant

- **Total Score: 87 (Rank 1)**
- A real-time, embedded AI tool that provides directional signals, confidence levels, and model-based feedback.
- It scored the highest on ROI (9) and Accuracy (9), making it ideal for reducing impulsive trades at the point of decision.

---

2. AI-Powered Onboarding Simulation

- **Total Score: 78 (Rank 2)**
- Interactive modules that walk traders through simulated scenarios before real trading begins.
- Strong performance in Learning Impact (9) and Scalability (8).

---

3. AI Paper Trading Sandbox

- **Total Score: 76.5 (Rank 3)**
- Safe environment for new users to test strategies using virtual capital with feedback mechanisms.
- Supports learning and risk-free experimentation.

---

4. Soft AI Nudges in UI

- **Total Score: 76 (Rank 4)**
- Non-intrusive hints and alerts (e.g., over-leverage warnings, position sizing prompts).
- Enhances decision-making without overriding autonomy.

---

5. Mandatory Literacy Module

- **Total Score: 68.5 (Rank 5)**
- Foundational financial education covering F&O basics, risk, and market behavior.
- Best used as a prerequisite to other interventions, especially for first-time traders.

Extended EDA with Data Wrangling

Dataset URL & Description
URL: https://www.kaggle.com/datasets/bendgame/options-market-trades

This dataset includes over 60,000 time-and-sales records for options trades from 2017–2019. Each entry captures timestamp, underlying, strike, expiry, buy/sell action, trade price, and volume. It provides a rich basis for analyzing directional trade accuracy among retail traders.

This will enable us to test our hypothesis by comparing directional success rates before and after introducing onboarding training and AI-assisted decision tools.

Population : Retail options traders executing directional trades

Dependent variable : Change in trade direction accuracy, measured as % of trades resulting in profit/loss due to misdirection

To better understand the behavioral and financial outcomes of options traders, an in-depth exploratory data analysis (EDA) was conducted on the dataset optionsTradeData_updated.csv.

✅ Data Wrangling Steps

Date Conversion:
The Date and ExpirationDate columns were converted to datetime format using pandas.to_datetime() to enable temporal trend analysis.

Categorical Encoding:
The pnl column, indicating whether a trade was a profit or a loss, was cast as a categorical variable for group-based analysis and plotting.

📈 EDA Observations & Results

1. PnL Distribution (Profit vs. Loss)
There is a significant class imbalance in outcomes:
Losses: 56,102 trades
Profits: 6,693 trades
This suggests that over **89% of trades end in a loss**, confirming the severity of the directional trading issue.

2. Chain Location vs. Outcome

Chain Location	Losses	Profits
ATM	201	0
ITM	4,085	6,693
OTM	51,816	0

Key takeaways:

OTM (Out-of-the-Money) trades dominate the loss category, suggesting poor strategy or directional bias.

All profitable trades occurred in ITM (In-the-Money) positions.

ATM (At-the-Money) trades were rarely profitable and rarely used overall.

Key Insights

The vast majority of trades are executed in OTM options, which carry higher risk and lower probability of profit.

The absence of profitable ATM and OTM trades highlights a lack of strategy or education on contract selection.

This supports the need for **AI-powered decision tools** to help traders reduce loses.

Research Hypothesis

This study hypothesizes that**:**

"If a financial literacy onboarding module and an AI-powered trade call assistant are implemented for new retail investors, then the number of loss-making trades attributed to lack of trading knowledge will decrease from 1.74 million to 870,000 (a reduction from 12% to 6%) within a 3-month period**.**

Implementation of the Countermeasure

To address the root cause of **uninformed directional trading**, an AI-Powered Trade Assistant was developed. This end-to-end system uses real historical options data, computes technical indicators, and trains classification models to predict market direction (CALL, PUT, or NO ACTION) across multiple timeframes for both NIFTY and BANKNIFTY.

Key Implementation Components

1. Data Collection

Historical intraday price data was sourced using the yfinance API for:

- **NIFTY (^NSEI)** and **BANKNIFTY (^NSEBANK)**
- Across 1-minute (7 days), and 5-, 15-, 30-, and 60-minute intervals (60 days)

Each file was stored in a structured folder and timestamped for version control.

---

2. Feature Engineering

To enhance the predictive power of the dataset, the following **technical indicators** were computed using the ta (technical analysis) library:

- **RSI (Relative Strength Index):**
  Measures the strength and momentum of recent price changes. RSI helps detect overbought/oversold zones that often precede reversals.
- **MACD and Signal Line:**
  These momentum indicators capture the convergence and divergence of moving averages, helping the model understand underlying trend shifts.
- **EMA 20 and EMA 50 (Exponential Moving Averages):**
  Short- and medium-term price smoothing indicators that reveal trend direction and crossover points, crucial for identifying CALL/PUT opportunities.
- **ATR (Average True Range):**
  Quantifies volatility. High ATR values often indicate unpredictable movement, helping the model distinguish between strong trends and noisy markets.

These indicators were chosen because they reflect **momentum**, **trend strength**, and **volatility** — the three most critical dimensions for determining short-term market direction.

Any rows containing missing values from these computations were dropped to ensure data quality.

---

3. Label Generation

A forward-looking label system was implemented to assign each row a directional class:

- **0 → PUT**,
- **1 → NO ACTION**,
- **2 → CALL**

This was based on price movement over the next 3 candles, with thresholds to ensure the movement was significant enough to warrant action. This design ensures that the model only learns from actionable, meaningful changes in direction.

---

4. Model Training

For each symbol and timeframe, an independent **XGBoost Classifier** was trained using:

- 80% training and 20% testing split
- Engineered features as input, and directional labels as target

Performance was monitored using classification reports and confusion matrices. Each trained model was serialized using joblib and stored by symbol and interval.

---

5. Integration and Deployment

The final app (app.py) loads all models on startup and allows users to:

- Select between **NIFTY** and **BANKNIFTY**
- Instantly view directional predictions (CALL/PUT/NO ACTION) for all timeframes

This setup delivers a clean interface backed by real AI-powered logic, directly intervening at the **point of directional decision-making**, which was the identified root cause of trader losses.

Final Results and Trader Guidance

After running the full AI prediction pipeline on the most recent processed options data, the model generated the following **frame-wise directional signals**:

Prediction Output:

Timeframe	NIFTY Prediction	BANKNIFTY Prediction
5 min	⏸ NO ACTION	⏸ NO ACTION
15 min	⏸ NO ACTION	⏸ NO ACTION
30 min	⏸ NO ACTION	⏸ NO ACTION
60 min	⏸ NO ACTION	⏸ NO ACTION

Interpretation:

The system recommends **no trading action** across all timeframes for both indices at the current moment. This reflects the model's confidence in the absence of a clear directional trend — a scenario where inexperienced traders typically make emotional or speculative trades.

Conclusion:

This type of **AI-powered restraint** is precisely what the intervention aims to enable:

- It helps prevent **impulsive entries** during uncertain periods.
- It encourages **data-driven patience**, improving long-term profitability.
- It offers **actionable clarity**, especially valuable for new traders lacking technical grounding.

By guiding users with consistent "NO ACTION" signals when the market shows low directional strength, the system effectively reduces the probability of poorly timed or emotion-led trades — directly addressing the root problem identified in the project.

Extended Abstract

Retail investor participation in India’s derivatives market has grown rapidly, with **8.7 million out of 9.6 million** active in F&O (Futures and Options) trading. However, **6.96 million incur losses**, and data shows that **approximately 1.74 million of these traders** could avoid losses with targeted support interventions.

Using the TBP A3 methodology, the project followed a five-step structure:

- **Step 1** quantified the measurable gap: identifying 1.74 million loss-making traders who could be helped through better decision-making support.
- **Step 2** broke down contributing factors, revealing that **lack of financial education (1.04 million)**, **market volatility (435,000)**, and **high leverage risk (261,000)** are primary contributors. The critical **Point of Occurrence** was determined to be **“Deciding the Direction of the Market.”**
- **Step 3** defined three measurable targets, including halving knowledge-related losses (from 12% to 6%) and increasing onboarding and stop-loss compliance rates.
- **Step 4** applied the **5 Whys analysis**, tracing impulsive directional trading back to a root cause: the absence of mandatory onboarding or structured financial education for retail traders.
- **Step 5** used an **Options Matrix** to compare five solutions. The **AI Trade Assistant** ranked highest, scoring **87/100** for ROI, accuracy, learning impact, and scalability — ahead of paper trading, onboarding simulations, and literacy modules.

The implemented solution included a machine learning pipeline using processed features like RSI, VIX, price change %, and option volatility, trained via XGBoost. Once trained, the model generated real-time predictions across timeframes (5m to 60m) for NIFTY and BANKNIFTY. In the final test, the model **advised "NO ACTION"** across all timeframes, correctly signaling caution during periods lacking directional clarity — a key risk scenario for novice traders.

By embedding such AI-powered restraint into trading workflows, this project demonstrates early potential to reduce emotional and misinformed trades and lays the groundwork for scalable, educational trading tools in India's retail market.

References

1. Securities and Exchange Board of India (SEBI). (2023). Report on Investor Protection in the F&O Segment.
https://www.sebi.gov.in/
2. **National Stock Exchange of India (NSE)**. (2024). *Historical Market Data: NIFTY and BANKNIFTY*.
https://www.nseindia.com
3. **Kumar, S.**, & **Bansal, V.** (2021). *Application of Machine Learning in Directional Options Trading*. *International Journal of Financial Studies*, 9(2), 115–132.

4. Kumar, A. (2009). Who gambles in the stock market? *The Journal of Finance, 64*(4), 1889–1933.
https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2009.01483.x
5. Zhang, Z., Aggarwal, C. C., & Qi, G.-J. (2020). Stock price prediction via discovering multi-frequency trading patterns. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1058–1068.
https://dl.acm.org/doi/10.1145/3394486.3403262