DES646: AI/ML for Designers
Group Log

Project Title: Open-Data Based India VIX Forecasting and Regime-Aware Equity Timing
Course: DES646 – AI/ML for Designers
Instructor: (To be filled)

Team Details:

- Satyam Singh (Roll No.: 220985)
- Shashwat Gautam (Roll No.: 221005)
- Vansh Mina (Roll No.: 221167)
- Hashvith (Roll No.: 221158)

High-Level Theme:

- Use only open, verifiable data from Indian markets.
- Build an explainable, regime-aware forecasting module around India VIX.
- Present outputs in a designer-friendly, narrative and visual form suitable for retail users.

---

1. Timeline of Group Activities

Week	Activity	Details / Outcomes	Members
Week 1	Problem Framing & Role Allocation	Explored multiple open-data ideas (flows, factors, VIX). Finalized project: *India VIX forecasting and regime-aware risk signal* using only public data. Defined constraints (no leakage, explainability-first). Allocated roles: Satyam (pipeline), Shashwat (features + baseline), Vansh (UX/visuals), Hashvith (ML experiments).	All
Week 2	Data Engineering Setup	Collected India VIX historical data from open sources. Cleaned timestamps, standardized column names, removed corrupt rows. Constructed daily OHLC from intraday data and computed realized volatility and intraday ranges. Documented assumptions and created a shared data dictionary for consistency.	Satyam, Vansh
Week 3	Weekly Pipeline & Features	Aggregated daily series into weekly OHLC. Defined target: next-week VIX return. Engineered core features: multi-horizon VIX momentum (2/4/8/12 weeks), rolling vol-of-vol, 52-week drawdown, and regime buckets (Very Low / Normal / Elevated / Stress). Ensured all features are computed using past data only (no leak).	Shashwat, Satyam
Week 4	Modeling, Blending & Evaluation	Implemented chronological split (train: older data, test: recent). Trained Elastic Net as interpretable baseline and gradient boosting model for non-linear structure. Designed regime-aware blending: higher weight to trees in calm regimes, Elastic Net in stressed regimes. Converted blended predictions into Risk-On / Neutral / Risk-Off labels. Evaluated using RMSE, MAE, directional accuracy and prepared final visual + textual narrative for report.	Hashvith, Vansh, All

2. Phase-wise Detailed Group Log

Phase I: Ideation and Framing

Activities:

- Brainstormed ideas within "open data for Indian markets": factor timing, mutual fund flows, index breadth, macro calendars, and volatility.
- Evaluated each idea based on: interpretability, availability of free data, clarity of problem statement, feasibility within course timeline, and potential for visual storytelling.
- Converged on India VIX forecasting: it directly encodes market fear/uncertainty and naturally fits a regime narrative.

Decisions Taken:

- Focus on *one* signal (VIX) but build it properly: strong data hygiene, no leakage, clear ML pipeline, explainable outputs.

- Position the system as a "decision assistant" that suggests regimes, not as a black-box trading oracle.

Phase II: Data Engineering and Weekly Pipeline

Key Work:

- Developed robust parsing logic to handle timestamp formats, holidays, and missing intraday bars.
- Implemented daily OHLC reconstruction and realized volatility computation.
- Defined simple filtering rules for abnormal days (e.g., extremely short sessions).
- Aggregated daily series into weekly OHLC and validated against raw patterns.

Responsibility Split:

- Satyam: Core preprocessing and weekly aggregation code.
- Vansh: Sanity checks, visualization of daily/weekly time series.
- Others: Reviewed assumptions, prepared documentation for data steps.

Phase III: Feature Engineering and Target Definition

Features Designed (Examples):

- Multi-horizon momentum: percentage change over 2, 4, 8, 12 weeks.
- Vol-of-vol: rolling standard deviation of weekly VIX returns.
- 52-week drawdown: distance from trailing high VIX to detect "calm vs. risk-on vs. panic" zones.
- Regime dummies: bucketed VIX levels and realized vol into discreet regimes.
- Range and realized-vol aggregates: proxies for intraperiod turbulence.

Target:

- Next-week VIX return y_w chosen to align with weekly decision cycles.
- Verified alignment of indices to avoid off-by-one errors in labeling.

Phase IV: Modeling, Blending, and Risk Signal Design

Elastic Net:

- Provided a transparent baseline to understand which features linearly correlate with next-week VIX moves.
- Helped detect redundant/noisy features to prune the set.

Gradient Boosting:

- Captured non-linear regime effects and thresholds (e.g., momentum matters differently at high VIX levels).

- Feature importances served as a second lens to validate financial intuition.

Regime-Aware Blend:

- In calm regimes: relied more on tree-based model to exploit subtle patterns.
- In stressed regimes: leaned toward Elastic Net for stability and smoother behavior.
- Converted blended forecast into discrete labels: *Risk-On*, *Neutral*, *Risk-Off*.

Phase V: Evaluation, Visualization, and Refinement

Evaluation:

- Used RMSE, MAE, directional accuracy, and correlation for a balanced view.
- Checked performance across sub-periods (calm vs. crisis) qualitatively.

Explainability and UX:

- Designed "Signal Cards" with: current regime, signal direction, key drivers, and a one-line plain-English explanation.
- Proposed timeline plots combining: VIX level, predicted direction, and regime shading.

Refinements:

- Removed hyper-aggressive parameter choices that caused unstable behavior.
- Tightened neutrality band to avoid noisy flip-flops between signals.

---

3. Individual Contribution Summary

Satyam Singh

- Led data engineering: minute-to-daily-to-weekly pipeline.
- Implemented target labeling and core evaluation scripts.
- Coordinated integration of code, visuals, and report.

Shashwat Gautam

- Designed feature set with financial and statistical intuition.
- Implemented Elastic Net baseline and analyzed coefficients.
- Documented methodology and mathematical formulations.

Vansh Mina

- Drove UX and visualization aspects (regime bands, signal cards).

- Helped verify data sanity and narrative consistency.
- Contributed to final report structuring and diagrams.

Hashvith

- Implemented gradient boosting / LightGBM model.
- Performed feature importance analysis and sensitivity checks.
- Assisted in designing regime-aware blending and validation.

---

4. Coordination, Tools, and Workflow

- Communication via shared messaging groups and scheduled weekly sync calls.
- Code collaboration through a shared repository / notebook environment.
- Versioned report drafts; each member annotated sections before freezing.
- Regular checkpoints: alignment on assumptions, avoidance of last-minute model changes, and verification of open-data compliance.

---

5. Challenges, Mitigations, and Learnings

Key Challenges:

- Handling missing/inconsistent data in high-frequency VIX series.
- Avoiding subtle leakage when constructing rolling features and targets.
- Balancing model complexity with interpretability.

How We Addressed Them:

- Adopted clear filtering and reconstruction rules and documented them.
- Used time-based splits; computed all statistics from training windows only.
- Used Elastic Net as an anchor model and treated tree-based model as a complement, not a black box.

Key Learnings:

- Good open-data systems are as much about disciplined engineering and communication as about algorithms.
- Regime-aware logic and explanations significantly increase trust in the model outputs.
- Collaboration with clear division of roles prevents duplication and confusion.

---

6. Declaration

We hereby declare that this Group Log truthfully represents the work carried out by our team for the DES646 course project. All members have contributed meaningfully to the project, and all external resources used are based on open/public data and course guidelines.