Predicting VIX Futures Direction Using Machine Learning

Financial Volatility Forecasting, Feature Engineering, and Model Automation

Project Overview

- **Objective:** Forecast next-day VIX Futures direction using financial, technical, and sentiment-based indicators.
- **Scope:**
  - Machine learning modelling for financial volatility forecasting
  - End-to-end automation of feature engineering, modelling, and evaluation
  - Integration of economic and news sentiment factors with technical indicators
- **Impact:** Enhanced data-driven insights and predictive accuracy for volatility analysis in financial markets.

Key Contributions and Leadership

- Helped establish foundational data workflows, version control systems, reproducible analytic pipelines, reusable functions, technical documentations, and analytics best practices.
- Designed and developed the full machine learning infrastructure from scratch, including:
  - Automated Python pipelines for data ingestion, transformation, and modelling
  - Rolling window generation, evaluation, and reporting workflows
  - Internal Python packages for reusable feature engineering and model comparison
- Conducted research and experimentation on new methodologies and model types, ensuring continuous improvement.

Data and Feature Engineering

- **Data Sources:**
  - Historical financial and macroeconomic time-series
  - News sentiment data and technical indicators (e.g. RSI, moving averages)
- **Feature Engineering:**
  - Engineered high-signal features combining technical, macro, and sentiment-based variables
  - Enhanced predictive performance and robustness through careful feature selection and transformation
- **Goal:** Improve signal-to-noise ratio in volatility prediction.

Modelling Techniques

Models developed and tested:

- Time-series forecasting: **ARIMA**, **GARCH**, **LSTM**
- Classification and regression: **XGBoost**, **Polynomial Logistic Regression**, **LightGBM**

Approach:

- Combined statistical and machine learning approaches for short- and long-term prediction
- Integrated external economic and sentiment indicators into model frameworks
- Used iterative testing and validation to ensure generalisability and consistency

Automation and Infrastructure

- Built fully automated Python pipelines covering:
  - Data ingestion, cleaning, and transformation
  - Feature generation and versioning
  - Rolling model training and evaluation
  - Report generation and insights delivery
- Created a system for model versioning, accuracy tracking, and comparison using a metrics database.
- Later incorporated **MLflow** for model tracking and reproducibility.
- Ensured automation boosted development speed and improved reliability.

Evaluation and Validation

- **Metrics:** Accuracy, Precision, Recall, F1-Score, Balanced Accuracy, MCC, and ROC-AUC
- **Validation:**
  - Rolling time-window testing
  - Out-of-sample forecasting to simulate live performance
- Built an automated evaluation framework to compare model stability and feature importance across iterations.
- Developed reporting scripts to summarise findings and visualise performance trends.

Documentation and Knowledge Sharing

- Authored comprehensive internal technical documentation covering:
  - Financial time-series modelling and statistical assumptions
  - Model interpretability and feature importance
  - Evaluation metrics and model comparison
  - Best practices in reproducible ML pipelines
- Created detailed methodology and results reports, with interpretations, recommendations, and model limitations.
- Contributed to process transparency and knowledge sharing within a small, fast-moving team.

Results and Impact

- Improved predictive accuracy and interpretability across multiple models.
- Achieved consistent performance across testing periods with robust feature selection.
- Delivered actionable insights for volatility trend analysis and risk forecasting.
- Strengthened internal technical infrastructure and documentation culture.
- Supported early-stage growth by establishing reusable codebases and technical foundations.

Reflection and Learning

- Gained deep understanding of volatility dynamics and time-series forecasting challenges.
- Strengthened ability to balance innovation, interpretability, and reliability in model design.
- Enhanced automation, data engineering, and model management skills through end-to-end delivery.
- Developed a strong analytical mindset, problem-solving ability, and independent ownership in a fast-paced environment.

Thank you.