COMP7705 Project

Detailed Project Proposal

Project Title: A Machine Learning-Based Model for Predicting Fund Transactions

Mentor: Dr. P.S. Vivien Chan

Student 1 (Leader) Chen Yuanhong

Student 2 Chen Peng

Student 3 Zhang Wenhao

Student 4

Student 5

Aim

The primary objective of this project is to develop a robust predictive model for forecasting daily fund subscriptions and redemptions. By accurately forecasting daily fund flows, we will help optimize inventory management, reduce investment risks for fund managers, and maximize market demand satisfaction. The system will incorporate financial market insights, product characteristics, and user behavior patterns to predict subscription and redemption amounts during both normal and abnormal market conditions.

Brief Literature Review

The prediction of financial time-series, the application of behavioral finance, and the optimization of inventory management have all emerged as crucial areas in fund flow forecasting and risk management. Financial time-series prediction traditionally relied on classical models like ARIMA and GARCH [1], but with advancements in machine learning, methods such as LSTM networks and hybrid models have been increasingly used for more accurate and robust forecasts, particularly under volatile market conditions [2]. These predictive models allow for the identification of complex patterns in fund flows, aiding in the management of liquidity and market risks. Simultaneously, behavioral finance introduces psychological factors, such as investor sentiment, overconfidence, and loss aversion, which often drive market anomalies [3][4]. Recently, inventory optimization models, including stochastic models and dynamic programming, have been adapted to the financial sector to predict future fund flows and determine optimal inventory management strategies [6]. The combination of these models, especially those incorporating machine learning, not only improves predictive accuracy but also enhances the flexibility of fund managers to respond to changing market conditions in real-time. The integration of these domains thus offers a comprehensive approach to fund flow forecasting, balancing market demand with risk management strategies to optimize liquidity and maximize market satisfaction.

References:

1. **Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.**
2. **Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669.**
3. **Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127).**
4. **Barberis, N., & Thaler, R. H. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, 1, 1053-1128.**
5. **Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.**
6. **Silver, E. A., Pyke, D. F., & Peterson, R. (1998). *Inventory Management and Production Planning and Scheduling*(3rd ed.). John Wiley & Sons.**

Proposed Methodology

Data Collection and Preprocessing

- Develop an ensemble framework to combine predictions from multiple models
- Implement dynamic weighting based on recent performance and market conditions

1. **Anomaly Detection Component**:

- Design a specialized module to detect abnormal market conditions
- Implement regime-switching capabilities to adapt predictions during market volatility

2. **Multi-task Learning**:

- Joint prediction of subscription and redemption **Historical Fund Data**: Collect historical subscription and redemption data for funds on the Ant Fortune platform.
- **Fund Characteristics**: Gather fund-specific features including fund type, size, historical performance, fee structure, and risk metrics.
- **Market Indicators**: Incorporate market indices, interest rates, volatility measures, and other macroeconomic indicators.
- **Temporal Features**: Extract time-based features including day of week, month, holidays, and seasonal patterns.
- **Data Cleaning**: Handle missing values, outliers, and normalize features appropriately.

Feature Engineering

1. **Lag Features**: Create lagged variables of subscription/redemption amounts to capture temporal patterns.
2. **Technical Indicators**: Develop financial technical indicators specific to fund performance.
3. **Market Sentiment Analysis**: Extract sentiment features from financial news and social media.
4. **Cross-Product Relationships**: Model substitution and competition effects between different funds.
5. **Volatility Measures**: Develop indicators to capture market volatility and abnormal conditions.

Exploratory Data Analysis (EDA)

1. Perform statistical analysis to understand the distribution and trends in the data. Visualize relationships between different features and their impact on fund flows.

Model Development

We propose a hybrid modeling approach combining multiple techniques:

3. **Base Models**:
    - Time series models (ARIMA, Prophet)
    - Machine learning models (XGBoost, LightGBM)
    - Deep learning models (LSTM, GRU networks)
    - Transformer-based architectures
4. **Ensemble Framework**:

- amounts
- Incorporate auxiliary tasks such as predicting market volatility

Evaluation Framework

1. **Metrics**: RMSE, MAE, MAPE for quantitative evaluation
2. **Cross-validation**: Time-based cross-validation to ensure robustness
3. **Backtesting**: Simulate real-world application through historical backtesting
4. **Anomaly Period Analysis**: Specific evaluation during periods of market volatility

Milestones

	Tasks	*Estimated completion time*	*Estimated number of learning hours*
1	Project Initialization	Mar 10	30
2	Data Collection and Exploration	Mar 31	40
3	Feature Engineering	April 15	40
4	Base Model Development	April 30	40
5	Advanced Model Development	May 15	40
6	System Integration and Optimization	May 31	40
7	Evaluation and Refinement	June 15	40
8	Documentation	June 30	30
9			
10			
			Total: 300

Deliverables

Items	
1	Prediction System
2	Technical Document
3	Research Report
4	Deployment Package
5	Website
6	
7	
8	
9	
10	