CRISTAIL

AVOIDING THE PITFALLS ON STOCK MARKET: CHALLENGES AND SOLUTIONS IN DEVELOPING QUANTITATIVE STRATEGIES

Marco Bergianti*, Nicola Cioffo*, Francesco Del Buono*, Matteo Paganelli*, and Angelo Porrello**
*name.surname@cristail.com **name.surname@unimore.com

Ital-IA
ITALIA INTELLIGENZA ARTIFICIALE
cini National Lab AIIS

CONTENT

INTRODUCTION

In the field of investment management, **stock selection plays a critical role and has experienced a significant transformation in recent years**. With the abundance of financial data, traditional statistical techniques are being replaced by state-of-the-art **Machine Learning** (ML) and **Deep Learning** (DL) methodologies. These modern approaches excel in **uncovering hidden patterns and possess high predictive power, revolutionizing the way stock selection is conducted**.

Surveying the main critical issues and dangers that afflict fintech practitioners in the development of an ML/DL pipeline, and present **some possible solutions** to mitigate them.

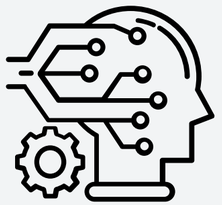
PIPELINE

Preparing the data for analysis by cleaning, organizing, and transforming it into a suitable format

DATA PREPARATION

Extracting meaningful features from the prepared data to capture relevant information for the modeling phase

FEATURIZATION



Developing and training machine learning or statistical models using the featurized data to make predictions or gain insights

MODELING

Assessing the performance and effectiveness of the models through various metrics and techniques to ensure their accuracy and reliability

EVALUATION

DATA PREPARATION

CHALLANGES

1. **Outliers & Missing values**: Financial data frequently contains stocks that trade intermittently and outliers.
2. **Survival bias**: Overestimation of the performance of the strategies by ignoring the stocks that have gone bankrupt or delisted.
3. **Look-ahead bias**: Provide information that are unavailable at inference time.

APPROACHES

1. **Exclude stocks** that have been traded on less than a certain percentage of trading days and **handle outlier values appropriately**.
2. **Include delisted securities** in the analysis or adjust the returns of surviving securities to account for the returns of the delisted securities.
3. Keep track to the **publication process for fundamental data**, and maintain proper **tracking of adjusted price updates** or transform them into a yield series.

FEATURIZATION

CHALLANGES

1. **Transforming raw price time-series**: Raw stock price series may exhibit **irregular frequency** and **varying statistical properties over time.**
2. **Transforming the label space**: The label space should be transformed consistently with the task at hand, considering factors such as **time horizon** and **class distribution.**

APPROACHES

1. **Explore alternative sampling methods** like volume bars and dollar bars to resample price series.
2. **Compute daily yields** or **fractionally differentiated features** to transform raw stock price series.
3. **Adjust prediction horizon dynamically** based on significant price changes relative to average volatility.

MODELING

CHALLANGES & APPROACHES

1. **Manage the stochastic nature of data** by implementing ad-hoc loss functions, adversarial training procedures, and intrinsically probabilistic models.
2. **Exploit correlations between stocks** through joint prediction of multiple stocks using techniques like GNN.
3. **Define the model optimization function** by adopting a ranking approach to select the most profitable stocks.

Stocks	S1	S2	S3	S4	S5	Performance		Profit Top-1 Stock
Returns	+30	-10	+20	+5	-30			
R1	+20	-5	+25	+10	-20	7	MAE	+20
R2	+20	-15	+10	+15	-10	11		**+30**
C1	↑0.60	↓0.60	↑0.70	↓0.55	↓0.55	**80%**	Acc.	+20
C2	↑0.70	↑0.55	↓0.55	↑0.60	↓0.60	60%		**+30**

Toy example derived showing that accurate regressors/classifiers (e.g., R1, C1) may be less profitable than other under-optimized methods (e.g., R2, C2)

EVALUATION

CHALLANGES

1. **Scenario Selection**: Defining representative and diverse scenarios for evaluation.
2. **Metric/KPI Design:** Choosing appropriate metrics and key performance indicators (KPIs) to reflect the underlying investment objectives.

APPROACHES

1. **Walk Forward &Combinatorial Purged Cross-Validation**: Employ techniques to define diverse scenarios that simulate out-of-sample performance, effectively reducing the risk of overfitting.
2. **Portfolio & Model Metrics**: Select and utilize robust portfolio metrics (e.g., Sharpe Ratio, Maximum Drawdown) and model metrics (e.g., MAPE, MAE, Accuracy) to effectively evaluate strategy performance and model reliability, ensuring a comprehensive assessment.

WHAT TO DO?

The proper management of price-adjusted and fundamental data to prevent look-ahead bias, and inclusion of delisted stocks to mitigate survivorship bias

DATA PREPARATION

Managing the unique properties of stock series and defining the label space

FEATURIZATION

Graph-based models and probabilistic output could be exploited for capturing stock correlations and managing their chaotic nature

MODELING

Introduce advanced cross-validation and backtesting for unbiased evaluation strategies

EVALUATION

WHAT'S NEXT?

Validate proposed solutions through real-world implementation and effectiveness evaluation

EMPIRICAL VALIDATION AND TESTING

Investigate innovative techniques to enhance performance in stock trading (i.e. LLM)

EXPLORATION OF NEW APPROACHES

Investigate the integration with other steps, such as risk management and portfolio construction

INTEGRATION

THANK'S FOR WATCHING

francesco.delbuono@cristail.com

CRISTAIL.COM

CRISTAIL