Research Project Plan – Academic Track

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

1. Research Participants

Nome completo
Student - Cristiane de Andrade Coutinho
Co-supervisor - Adriana Vieira
Supervisor - Geraldo Magela Severino Vasconcelos

2. Motivation and Justification

Research Topic: Forecasting asset returns with deep learning algorithms

When financial institutions lend resources, a trading environment is created with the objective of obtaining financial returns from investments. In this context, the rewards derived from these negotiations are referred to as “securities.” When acquiring a debt security, the investor also assumes the risk of such return, i.e., the possibility of this investment being devalued.

With the aim of maximizing results, a whole technological structure has been developed to support and establish the relationship of time and causality in asset gains. According to Guedes (2022), the development of algorithms in the financial sector was gradually incorporated into trading markets and evolved as more technical analyses of mathematics, probability, and statistics were embedded in so-called “trading robots.”

With these technological advances, the development of machine learning and deep learning algorithms has become increasingly viable and attractive, particularly for asset selection and portfolio optimization. Shimabukuro (2024) attributes these

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

advances to the growing increase in computational processing power, reducing processing costs and expanding capacity, making the implementation of optimization and prediction algorithms more affordable.

Soares (2023) highlights that machine learning algorithms are widely applied to regression and classification problems for asset pricing prediction and asset classification. On the other hand, Shimabukuro (2024) emphasizes the limiting factors associated with applying machine learning models to prediction tasks, such as the need for large data volumes, low sensitivity to market noise, and the risk of overfitting.

In light of this, this study proposes a qualitative approach to analyzing different types of neural network architectures, applying deep learning techniques to capture market noise and generalizing the problem in such a way that these architectures can be applied to different types of financial assets.

3. Research Problem

The prediction of a financial asset's return is extremely complex due to its noisy nature and the difficulty algorithms face in capturing its nuances and generalizing the problem. Thus, the research problem lies in identifying which deep learning architectures can best generalize the prediction of financial asset returns.

4. Objetivos

- **General Objective:** Analyze the predictive quality of financial return forecasting using different neural network architectures with deep learning techniques on time series data.
- **Specific Objectives:**
    1. Evaluate the performance of trained architectures by comparing predicted values with actual values.

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

2. Measure the generalization capacity of the architectures by testing them on different datasets.

5. Scope

This work aims to study different deep learning architectures for predicting financial asset returns. In addition, it seeks to understand how these architectures can generalize the problem across different assets.

Datasets will be searched on online platforms providing time series data so that the behavior over time can be observed. After collecting the datasets, exploratory data analysis will be conducted to identify patterns, anomalies, and their main characteristics. If anomalies are detected, data cleaning, normalization, and standardization will be performed.

Alongside this step, different neural network architectures will be studied to find an appropriate solution to the problem. A series of tests with the datasets will then be carried out.

In this way, the study is expected to contribute with new perspectives on the subject for the scientific community, delivering consistent results.

6. Research Methodology

- **Type of research**: Exploratory.
- **Approach**: Quantitative.
- **Data collection methods:** Literature review and analysis of public datasets.
- **Analysis techniques**: Exploratory data analysis.

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

7. Work Schedule

- **SPRINT 1** 04/08 - 15/08

  Deliverable 1: Project Submission Form

  Provide relevant project information, clearly describing the research problem, objectives, and hypotheses, along with a high-level scope of what will be developed in each sprint.

  Deliverable 2: Project Plan Document

  Outline the necessary steps to conduct the research, specifying the sprint-based work schedule, detailing each deliverable, and including the research justification, problem description, scope, methodology, and expected results.

- **SPRINT 2** 18/08 - 29/08

  Deliverable 1: Introduction to the research topic with literature review

  - Writing of the introduction, highlighting key points regarding its relevance, literature review, current context of the problem, and intended scientific contributions.

  Deliverable 2: Justification of the chosen topic

  - Writing of the project justification, emphasizing the motivations that led to the choice of the topic.

  Deliverable 3: Defined research objectives

  - Definition and writing of the general and specific research objectives.

  Deliverable 4: Research hypotheses

  - Definition and writing of the research hypotheses.

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

Deliverable 5: Research problem

- Delimitation of the research problem, considering the chosen neural network architectures.

Deliverable 6: Description of the dataset used

- Description of the data source.
- Description of the dataset variables.
- Description of the type of each variable.

Deliverable 7: Definition of neural network architectures to be used

- Description of the architectures used.
- Justification for their use.

- **SPRINT 3** 01/09 - 12/09

Deliverable 1: Exploratory data analysis

- Number of examples for each variable.
- Verification of missing or empty data.
- Descriptive statistics (mean) for numerical variables.
- Descriptive statistics (mode) for categorical variables.
- Detection of outliers.
- Distribution of each variable.
- Correlation matrix.

Deliverable 2: Data cleaning

- Remove or impute missing values
- Eliminate duplicates
- Handle outliers
- Normalization / standardization

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

- Encoding of categorical variables
- Encoding de variáveis categóricas

Deliverable 3: Materials and methods to be used in the research

- Description of the methods applied for data cleaning
- Description of exploratory data analysis

Deliverable 4: Implementation of the chosen architectures

- Implementation of the selected architectures
- Evaluation of training and results obtained

- **SPRINT 4** 15/09 - 26/09

  Deliverable 1: Implementation of the chosen architectures

  - Refinement of the architectures used
  - Re-evaluation of the architectures

  Deliverable 2: Materials and methods

  - Description of the implemented architectures
  - Description of the methodological processes used

  Deliverable 3: Peer review

  - Submission of the project to another professor for evaluation of scope and development

- **SPRINT 5 -** 29/09 - 10/10

  Deliverable 1: Refinement of the chosen architectures

  - After evaluation, submission of the architectures to a refinement process

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br

Deliverable 2: Study conclusion

- Compilation of the results obtained for writing the study's conclusion

Deliverable 3: Final presentation of the study

- Preparation of a presentation slide deck in English

8. Expected Results

This study is expected to provide new insights into return prediction using deep learning, as well as propose new approaches and improved results compared to the current literature.

9. General Remarks on the Research to be Developed

10. References

[1] GUEDES, Anderson Cerqueira et al. *Computational evolution for portfolio optimization of trading strategies in the financial market*. 2022.

[2] SOARES, Taís Rigor. *Forecasting Asset Returns in the Financial Market Through Optimization Methods*. Master's Dissertation, 2023.

[3] SHIMABUKURO, Camilo Ilzo. *Deep Learning Applied to Return Prediction in the Ibovespa: An Analysis of LSTM Neural Network Performance Using Log-Returns and Fractional Differentiation*. PhD Thesis, University of São Paulo, 2024.

Inteli - Instituto de Tecnologia e Liderança
Avenida Professor Almeida Prado, 520, Butantã, São Paulo, SP.
CEP: 05508-070
www.inteli.edu.br