# *Profit and Loss Prediction*

*Md Shariful Islam*
*6352222*

*Hiruni Piyumika Bellanthuda Achchige*
*6352751*

*Eunice Kakra Ama Ewudzie*
*6351805*

MSc Artificial Intelligence and Intelligent Systems

Faculty 3 - Mathematics and Computer Science

University of Bremen

February 2025

Table of Contents

# 1. Problem & Challenges

Profit and loss prediction involves analysing historical profit and loss statements of a company so as to estimate the future revenues, costs and ultimately net profits of the company. In most companies, the profit and loss statement is a summary document of revenues, costs, operating expenses and other financial items for every fixed period. These documents help managers to understand how well a company has performed in the past so that they can make informed decisions about investments, budgets, and risk management. However, different things influence such financial outcomes. These include market conditions, internal operational efficiency, and external economic variables. It makes economic forecasting for the future very hard. This requires not only an understanding of trends from the past but also some forecasting of variation and changes through unseen events or seasonal trends.

The future profit and loss performance of a company may not only be predicted but rather is going to be subjected to a few important challenges. Firstly, the past data sets may be noisy and incomplete due to differing accounting methods, reporting delays, or clerical errors. Financial statements may come in different formatting style that would require high-level data preprocessing and normalization steps to ensure consistency. Secondly, being quite volatile, profit and loss is directly affected by the market performance, regulatory changes, or downturns in the economy. Capturing the impact of such dynamic and sometimes unpredictable influences demands a forecasting model that is both flexible and robust.

Such effects might be regarded as messing with seasonality and non-linearity, which would therefore not be captured by any simple linear approach but might introduce more complications in financial patterning. Expense structures, revenue channels, and operational costs tend to be quite variable, making forecasting that much more challenging, for such factors might be inconstant themselves over time. Uncertainty quantification, nonetheless, is vital, with decision-makers needing to appreciate the probable future 'most-likely' outcomes and the possible variations. This translates to requiring methodologies that can provide not just relative forecasts but also confidence intervals or probability distributions. Finally, the broader question of continual incorporation of up-to-date information and ensuring that the model learns without adapting to noise is another major challenge of the design of such a system.

# 2. Research Related Works

*1. Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison*

U. M. Sirisha, M. C. Belavagi and G. Attigeri, "Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison," in IEEE Access, vol. 10, pp. 124715-124727, 2022, doi: 10.1109/ACCESS.2022.3224938.

Link to research paper https://ieeexplore.ieee.org/document/9964190

*2. Profit Prediction for Businesses using Machine Learning Algorithms*

A. Gaikwad, T. Ghodke, A. Jadhav, M. Pande and S. Mirchandani, "Profit Prediction for Businesses using Machine Learning Algorithms," 2023 8th International Conference on Communication and

Electronics Systems (ICCES), Coimbatore, India, 2023, pp. 1039-1044, doi: 10.1109/ICCES57224.2023.10192797

Link to research paper https://ieeexplore.ieee.org/document/10192797

*3. Profit prediction optimization using financial accounting information system by optimized DLSTM*

Th Wei Tang, Shuili Yang, Mohammad Khishe, Profit prediction optimization using financial accounting information system by optimized DLSTM, Heliyon, Volume 9, Issue 9, 2023, e19431, ISSN 2405-8440,

Link to research paper https://doi.org/10.1016/j.heliyon.2023.e19431

# 3. Choice of Technology

Based on the reviewed literature and the nature of the problem led to the selection of a deep learning technology that employs Long Short-Term Memory (LSTM) networks to solve the problem of profit and loss prediction. Since LSTM networks can remember long-term memory in sequential data, they are particularly well suited for time-series forecasting. Financial statements including P&L reports are indexed in time, and considerations about past trends and seasonal patterns in time are necessary for an accurate forecast. Uncertainty estimation techniques such as Bayesian neural network approaches or Monte Carlo dropout will be evaluated along with LSTM, which will also enable the model to give prediction intervals instead of mere point estimates.

The selected LSTM model first layer that contains 50 units and uses the tanh activation function. Its return_sequences=True indicates that the output sequence should be actually forwarded to the next recurrent layer. After this, there is a dropout layer at a 20% of so that overfitting is avoided. The process continues with an additional LSTM layer of 50 units with tanh activation function. This, therefore, captures deeper dependencies in the data. A dropout layer has been added (20%) further to reduce the effect of overfitting. Finally, a fully connected Dense layer is introduced, comprising a single neuron, for outputting the predicted stock price. The model was compiled with the Adam optimizer to optimize the mean squared error loss. It has undergone training for about 50 epochs with a batch size of 8 and had validation data for assessing generalization. This architecture thus easily captures these temporal patterns with trade-off complexity against regularization.

# 4. System Description (End-to-End)

The Advanced Profit & Loss Prognostication Solution developed here is a integrated and moulded system aimed at transforming raw financial information into actionable insights utilizing deep learning approaches and uncertainty measurement strategies. The system structure includes the following modules:

## 4.1 Data Collection and Preparations

Data Loading: The process starts with the system trying to ingest historical profit and loss information from a CSV dataset labeled (historical_pl.csv). If not available, dummy monthly datasets are created in order to mimic possible financial records from January 2018 to December 2023. The dataset produced includes necessary features such as Date, Revenue, and Expenses.

Data preprocessing involves orderly arrangement of existing data in chronological order, in addition to filling in missing values with a forward-filling method. The system next computes Profit based on the difference between Revenue and Expenses. In addition, key temporal features (e.g., Month, Year, and Quarter) and lag features (e.g., Revenue in preceding months) are formed in order to identify intrinsic temporal trends.

## 4.2 Transformation of Data for LSTM Forecasting:

Given the wide fluctuation in revenue amounts, the system uses a MinMaxScaler to convert the revenue information into a normalized range (0–1). Normalization is necessary in order to successfully train neural networks.

In order to convert time series data into a structure compatible with deep learning models, a windowing method is utilized. The sequences are formed based on a specified look-back horizon (e.g., the last 12 months), and these are taken as the input features, with the revenue in the next month taken as the target variable. The method helps the model identify the temporal and long-range dependencies in financial data.

## 4.3 LSTM-Based Forecasting Module:

The core forecasting engine is built on an LSTM network, which is well-suited for handling sequential data. The architecture comprises two LSTM layers interleaved with dropout layers to reduce overfitting, followed by a Dense output layer that predicts the next month's revenue.

The model is trained with the use of past data with mean squared error (MSE) serving as the loss function designated with the optimizer, Adam. The training process is monitored carefully in order to converge and avoid overfitting with the curves comparing loss on the train and validation.

Iterative Forecasting: After finishing the train phase, the model moves on to produce systematic predictions of future revenue in a step-by-step process. Beginning with the recent sequence of observed results, the system creates monthly predictions, next updating the sequence with every following prediction. The forecasted values are next back-transformed in order to fit the original revenue measurement.

## 4.4 Assessing Uncertainty with Monte Carlo Simulation:

The residuals, hereinafter called residuals, are examined in order to evaluate how uncertain the predictions made with the test dataset are. The residuals' standard deviation is a measurement of the forecasting error. To improve risk evaluation, a Monte Carlo simulation is run on the forecast month.

By taking repeated draws from a normal distribution with the forecast value and the residual standard deviation, the system generates a distribution of potential results. The simulation produces a 90% prediction interval, and consequently, a quantitative measurement of the uncertainty in the forecast.

### 4.5 Visualization and Reporting:

Data and Visualization of Predictions: The tool uses matplotlib for generating in-depth visualization. These visuals include graphical presentations of Revenue and Profit curves in the past, training loss curves representing how models converge, and predictive visuals superimposing forecast revenue over already collected information.

Portrayal of Uncertainty: The simulation results are presented in a histogram representing the 5th and 95th percentile. In addition, a shading within the forecasting graph marks the prediction interval in the following month, thus making uncertainty understandable to decision-makers

### 4.6 The integration of users:

Holistic Workflow: The system covers all stages, from data ingestion right up to the final visualization, as an isolated workflow in a Jupyter Notebook. This platform allows for not just efficient prototyping and verification but also includes an easy-to-use interface that allows stakeholders to monitor and participate in the forecasting process.

Prototype Accessibility: The entire implementation which includes the LSTM model, Monte Carlo simulation, and visualization features is incorporated within a Python-developed prototype and made available on GitHub. The prototype supports execution, adjustment, and augmentation, and in doing so ensures openness and reproducibility. You can access the prototype at: https://github.com/hirupiyumika/Profit-Loss-Prediction-System

## 5. Limitations of the System

The proposed system offers a good solution for profit and loss prediction, but it has some limitations. A main challenge is the depend on quality and consistency of historical data. If input P&L statements are incomplete or inconsistently formatted, model accuracy may decline. In addition, it does not offer a complete response to unexpected events which affect economic activities, for example market dips, regulatory changes, etc. These events tend to make management performance vary quite significantly. Although uncertainty quantification does offer prediction intervals, it still cannot fully account for rare events of very high impact.

Lack of interpretability is another limitation of deep-learning models. Despite their great power of prediction, LSTM networks will hardly relieve the financial manager by making decisions transparent since they will somehow operate as “black boxes.” This lack of transparency in decision-making might impact the trust and acceptance by the stakeholders who expect clear decision-making tools. Moreover, retraining and tuning the parameters continuously are needed to maintain the performance of the model. A change in business model or market environment could also necessitate fairly frequent updates that are resource-intensive.