BU3001
Finance

Assignment 2
Predicting Quarter 2 Earnings of Snowman Logistics

By
Arya Lamba
Dhirain Vij
Manjree Kothari
Nimrat Kaur
Samarth Anand

Data Collection & Exploration

This study aims to forecast the quarterly earnings of **Snowman Logistics**, with data gathered from **Refinitiv Eikon** in the **Jefferies Lab**. The data collection process was designed to capture a holistic view of the company's financial performance as well as relevant macroeconomic indicators, allowing for a more comprehensive modeling approach. The data for Snowman Logistics includes several financial statements and key ratios, each serving a unique purpose in the analysis:

1. **Income Statement (Standardized)**:
    - This statement provides a snapshot of Snowman Logistics' revenues, costs, and profits over each quarter. Standardization ensures that the figures conform to a consistent format, making it easier to analyze trends and perform comparative analysis. Metrics such as **net income**, **operating expenses**, and **gross profit** from the income statement serve as core indicators of the company's financial health and earnings potential.
2. **Balance Sheet (Standardized)**:
    - The balance sheet includes the company's assets, liabilities, and shareholders' equity for each quarter. Key items such as **total assets**, **total liabilities**, and **shareholder equity** help assess the company's financial stability and liquidity.
3. **Cash Flow Statement (Standardized)**:
    - This statement tracks cash inflows and outflows, focusing on three main areas: **operating activities**, **investing activities**, and **financing activities**.

Cash flow data highlights Snowman Logistics' ability to generate cash from its core operations, which is crucial for sustaining earnings.

4. **Key Ratios**:

    - A set of key financial ratios, including **profitability ratios**, **liquidity ratios**, **leverage ratios**, and **efficiency ratios** were also collected. These ratios provide a quick assessment of the company's financial condition and performance. For example, profitability ratios, such as **return on equity (ROE)** and **net profit margin**, offer insight into how effectively Snowman Logistics is converting revenues into profit.

Macroeconomic Indicators

To account for broader economic influences on Snowman Logistics' earnings, we also collected macroeconomic data that could have a significant impact on performance. These indicators include:

1. **Quarterly Gross Domestic Product (GDP) and GDP Growth Rate**:

    - GDP represents the total economic output, and quarterly GDP growth rates indicate the pace at which the economy is expanding or contracting. Since Snowman Logistics operates within an economic environment, changes in GDP and its growth rate can influence demand for logistics services. For example, during periods of economic growth, logistics companies often see higher demand due to increased production and consumption.

2. **Quarterly Repo Rates**:

    - The repo rate, set by the central bank, affects the cost of borrowing in the economy. A higher repo rate can lead to increased financing costs for businesses, potentially reducing profit margins. On the other hand, a lower repo rate may encourage borrowing and investment, which can support business growth. By including repo rates, the model can better capture the influence of changing interest rates on Snowman Logistics' financial performance.

Feature Selection & Engineering

For predicting the quarterly earnings of our company, we needed features that would help us do so. We followed these steps for feature selection:

1. **Categorization of Features Based on Domain Knowledge**

- Historical Financial Data: Revenue, gross profit, and previous net income are strong predictors of a company's financial performance due to their direct relationship.

- Company-Specific Characteristics: Industry sector, company size, and geographic revenue distribution provide context, helping the model differentiate between companies based on their structural characteristics.
- Macroeconomic Indicators: GDP growth, inflation, and other broader economic conditions indicate the economic environment and its potential impact on corporate profitability.
- Market Data and Sentiment: Stock price trends, trading volumes, and other variables capture investor perception and market sentiment, which can influence financial outcomes indirectly.
- Company Announcements and Competitor Performance: Major company events, such as new product launches, and shifts in competitor performance often have an incremental influence on revenue and market position.
- Dividing these features into domains would allow us to use variables which are directly or indirectly related to Net Income.

2. **Removing redundant features**

- Features that are highly correlated with each other can be removed to reduce redundancy, keeping only the feature that is most closely related to Net Income. This makes models less redundant and more computationally efficient.

3. **Empirical Validation through Baseline Model Testing**

- Iterative Refinement: If the model doesn't perform well (like high MSE), we know we need to make adjustments. We add or remove features based on this feedback to get the best feature set for our final model.
- Baseline testing lets us take an iterative approach, refining the feature set until we find a good balance between simplicity and predictive power.

4. **Final feature selection brings together data-driven insights and industry expertise**

- External Events: Although features like GDP growth and competitor earnings may not always show strong statistical significance, they offer valuable context, especially in cyclical industries.
- Sector-Specific Variables: In certain sectors, seasonality or competitor metrics can provide subtle insights that enhance model understanding and accuracy.

Model Development

We started the model development with an extensive literature review to understand how the industry works on this problem. Initially, we identified 4 models that could be used: Long Short-Term Memory (LSTM) networks, Transformer models, ARIMA (Autoregressive Integrated Moving Average), and MIDAS (Mixed Data Sampling). Each model has its own advantages in handling time series data and financial forecasting applications.

From the 4 selected models though, we eliminated Transformers as it requires substantial amounts of high-frequency data which our dataset lacked and does not perform well on sparse data. In our case, we had several features but only about 40 observations for the model.

We tried an LSTM model with a feature set but that led to the vanishing gradient problem and could not predict losses. So, we simplified the model to use raw data inputs. This resulted in the model performing effectively.

ARIMA is expected to be successful in time series forecasting but, in our case, it gave a high Mean Squared Error (MSE). The intrinsic linearity of ARIMA, which was unable to adequately capture the intricate, non-linear correlations found in Snowman Logistics' earnings patterns, is what led to that result.

Additionally, we explored more research papers that emphasized the superior performance of ensemble methods in earning predictions. The models were mostly regression ones. We made a final ensemble with GRU (Gated Recurrent Unit), LSTM, CNN-LSTM hybrid, traditional RNN (Recurrent Neural Network), and MIDAS.

The GRU offers efficient processing of temporal sequences with its simplified architecture, making it particularly effective for capturing shorter-term patterns in the financial data. The LSTM, with its more complex memory structure, excels at identifying and utilizing long-term dependencies in the quarterly earnings patterns. The CNN-LSTM hybrid architecture provides a powerful combination where the CNN component extracts relevant features from the input data, while the LSTM processes these features temporally, capturing both spatial and temporal patterns in the financial indicators. The traditional RNN serves as a complementary model, offering basic sequential pattern recognition that can capture immediate temporal dependencies in the data.

The ensemble approach offers several key advantages over single-model implementations. By combining multiple models, we reduce the impact of individual model biases and limitations. The diversity in modelling approaches allows us to capture different aspects of the earnings patterns, from immediate market reactions to long-term trends. The weighted combination of predictions from each model provides a more robust and reliable forecast, minimizing the risk of significant prediction errors that might occur with any single model.

Methodology & Analysis

Data Preparation:

- Used historical quarterly financial data including Revenue, Net Income, Gross Profit, EBITDA, Income Tax, and Operating Expenses
- Reversed data chronologically to ensure recent data has more influence
- Applied MinMaxScaler for normalization to scale all features between 0 and 1

- Used sequence length of 4 (one year of quarterly data) for predictions
- Split data into sequences for training using sliding window approach

Feature Selection:

- While multiple financial metrics were available, focused on Net Income as the target variable
- This choice was made as net income is the most direct measure of earnings
- Maintained chronological order of data to preserve temporal patterns
- Used historical sequence of 4 quarters to predict the next quarter

Model Used and its Strengths:

The final model used was an ensemble model - a machine learning approach that combines multiple individual models to create a more robust and accurate prediction system. It was a combination of 4 different models: LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), CNN-LSTM (A combination of Convolutional Neural Network and LSTM), and RNN (Recurrent Neural Network). Equal weightage was not given to every model, this is because when we ran each model independently, we got the best results from LSTM based models in comparison to all other models, that is why LSTM based models have been given the highest weightage.

The ensemble approach is particularly effective because it leverages the strengths of each individual model while minimizing their weaknesses, leading to more stable and reliable predictions than any single model could provide on its own. By combining multiple models, we also reduce the risk of overfitting and increase the model's ability to generalize across different market conditions.

1. LSTM (40% weight)
    - Best for quarterly patterns and long trends
    - Remembers important past financial events
    - Can spot relationships between distant quarters
    - Given highest weight for handling complex financial time series
2. GRU (20% weight)
    - Faster, lighter version of LSTM
    - Better for recent earnings trends
    - Less likely to overfit on limited financial data
    - Captures quick changes in earnings patterns
3. CNN-LSTM (25% weight)
    - CNN finds local patterns in earnings data
    - LSTM part tracks how these patterns evolve
    - Good at spotting sudden earnings changes
    - Helps filter out market noise
4. SimpleRNN (15% weight)

- Tracks basic quarter-to-quarter changes
- Simple but reliable for short-term trends
- Acts as stability check
- Helps prevent over-complication

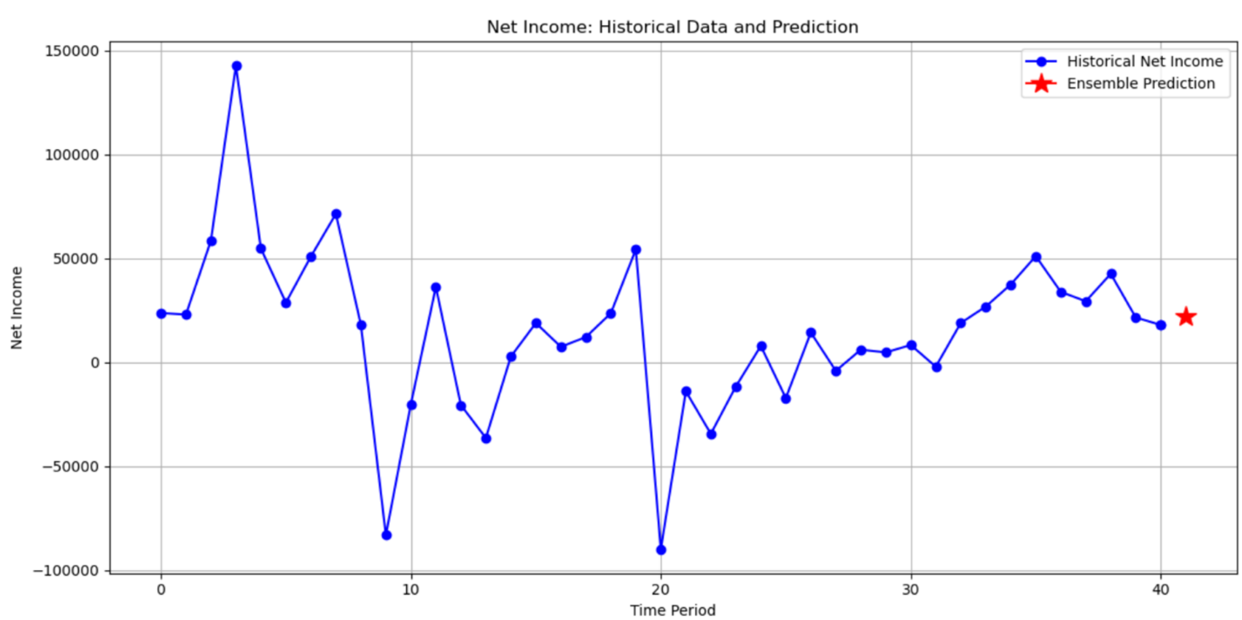
Why This Combination Works:

- Each model picks up different earnings patterns
- Mix of simple and complex approaches
- Balances short and long-term predictions
- More reliable than any single model

The ensemble combines LSTM's long-term memory, GRU's efficiency, CNN-LSTM's pattern detection, and RNN's simplicity to make better earnings forecasts. Weights are assigned based on each model's strengths in financial prediction.

Evaluation:

- Used Mean Squared Error (MSE) as training loss function
- Tracked average loss during training epochs
- Evaluated quarter-over-quarter percentage changes
- Compared model predictions against actual values
- Generated both weighted and equal-weighted ensemble predictions and evaluated effectiveness of both.



```
Individual Model Contributions:
LSTM: Prediction = 18364.56, Weight = 40.00%, Contribution = 7345.82
GRU: Prediction = 19182.21, Weight = 20.00%, Contribution = 3836.44
CNN-LSTM: Prediction = 26859.39, Weight = 25.00%, Contribution = 6714.85
RNN: Prediction = 18095.37, Weight = 15.00%, Contribution = 2714.31
```

```
Ensemble Model (Weighted Prediction):
                  Period  Net Income  Quarter-over-Quarter Change %
0            Last Quarter   17898.000                            NaN
1  Predicted Next Quarter   20611.419                      15.160459
```

Data Limitations:

- Limited to quarterly financial data
- Small dataset size for deep learning models

Model Limitations:
- Complex ensemble architecture requires significant computational resources
- Fixed weights in ensemble rather than adaptive weights
- Single-step (one quarter ahead) prediction only

Conclusion

Our final prediction for Snowman Logitics' Q2 **Net Income** is **20,611**.