LEVERAGING ARTIFICIAL INTELLIGENCE MODELS FOR FINANCIAL FORECASTING: A DETAILED ANALYSIS OF PREDICTIVE PERFORMANCE AND BENCHMARKS

Mfon Akpan *

* Department of Accounting & Financial Economics, Methodist University, Fayetteville, USA

OPEN ACCESS

How to cite: Akpan, M. (2025). Leveraging artificial intelligence models for financial forecasting: A detailed analysis of predictive performance and benchmarks. In M. Pazarskis, A. Kostyuk, V. Santolamazza, & P. Capuano (Eds.), *Corporate governance: Scholarly research and practice* (pp. 96–100). Virtus Interpress. https://doi.org/10.22495/cgsrapp18

Received: 08.10.2024
Accepted: 23.10.2024
Keywords: Artificial Intelligence, Financial Forecasting, Predictive Analytics, Benchmarking, Revenue Prediction, Net Income Prediction
JEL Classification: C45, C53, G17, M41
DOI: 10.22495/cgsrapp18

Abstract

The increasing integration of artificial intelligence (AI) into financial forecasting has garnered significant interest within finance and accounting domains. AI systems, proficient in mathematical and logical computations, can enhance the precision of financial predictions, making them invaluable for decision-making and risk management. This research investigates the predictive power of various AI models in forecasting critical financial metrics — specifically revenue and net income — while correlating their benchmark scores with predictive effectiveness. Previous literature underscores AI's advantages over conventional statistical methods, with deep learning and ensemble approaches frequently cited for their accuracy in forecasting financial outcomes. Benchmark assessments such as the Multi-task Language Understanding (MMLU) and Grade School Math 8K (GSM8K) are integral in evaluating a model's mathematical capabilities while problem-solving benchmarks like the ARC-Challenge and Graduate-level Problem Solving Questions (GPAQ) test their reasoning abilities.

Despite these advancements, a comprehensive evaluation of AI models’ predictive accuracy across a variety of financial indicators still needs to be more extensive. This study addresses this research gap by analyzing historical data from ten publicly traded companies spanning 2020 to 2022 and predicting their 2023 financial performance. A zero-shot prompt-based approach is employed, and the predictive outputs are compared against actual financial results, assessing model accuracy in relation to benchmark scores. The findings of this study enhance the understanding of how AI can be leveraged for financial forecasting and provide practical insights for implementation in accounting practices. Emphasizing the importance of data quality, model transparency, and bias management, this research contributes to the growing body of knowledge on the application of AI in financial analysis.

1. INTRODUCTION

The advent of artificial intelligence (AI) in finance and accounting represents a transformative shift in how these sectors approach complex problem-solving. AI’s strengths in processing large datasets, identifying patterns, and generating predictive insights offer a compelling alternative to traditional financial forecasting methods. Accurate financial forecasts are essential for effective risk management, strategic planning, and decision-making, underpinning the long-term success of organizations (Sezer et al., 2020).

This study evaluates AI models’ predictive accuracy in forecasting revenue and net income by leveraging historical financial data and assessing correlations with benchmark scores. This approach provides an empirical basis for understanding the capabilities and limitations of current AI models in financial forecasting.

2. LITERATURE REVIEW

2.1. AI in financial forecasting

AI’s adoption in financial forecasting has evolved due to its ability to handle complex data-driven analyses more effectively than traditional statistical methods. Deep learning and ensemble models are particularly noted for their capacity to capture non-linear relationships and generate more accurate predictions (Ranaldi et al., 2022). Such models can assimilate extensive historical data, enabling nuanced predictions that can adapt to market shifts and trends (Khattak et al., 2023).

2.2. Benchmarking AI models

Evaluating AI models often involves standardized benchmarks that assess their proficiency in specific tasks. For example, the MMLU test measures a model’s reasoning and problem-solving abilities across disciplines (Hendrycks et al., 2021). Similarly, the GSM8K benchmark gauges mathematical reasoning essential for accurate financial modeling (Clark et al., 2018).

2.3. Current gaps in the literature

While benchmarks provide a preliminary indication of an AI model's potential, limited research correlates benchmark scores with real-world predictive accuracy in financial forecasting. Addressing this gap, this study explores the connection between benchmark performance and the predictive accuracy of AI models when applied to real financial data.

3. METHODOLOGY

3.1. Data collection and preparation

The study draws on financial data from ten publicly traded companies from 2020 to 2022. This period was chosen to capture financial fluctuations due to market variability, including post-pandemic economic conditions. Data was extracted from the companies' annual 10-K filings, ensuring accuracy and consistency. Key metrics analyzed include revenue and net income, structured into a dataset for comparative analysis.

3.2. Benchmark assessment

The selected AI models were assessed using benchmark scores from well-recognized tests:

- MMLU: Evaluates multi-disciplinary problem-solving skills.
- GSM8K: Measures mathematical proficiency at the grade school level.
- ARC-Challenge: Tests reasoning through complex, real-world problems.

These benchmarks were chosen because they are relevant to financial forecasting tasks that require both mathematical acumen and logical reasoning.

3.3. Model implementation

Three AI models — Claude 3.5, ChatGPT-4, and Gemini — were employed for predictions. A zero-shot prompting technique was used to simulate real-world conditions where models might not be fine-tuned for specific tasks. The following input format was utilized:

Prompt: "Given the historical data from 2020 to 2022, predict the revenue and net income for 2023 for [Company Name]. The data includes: [list revenue and net income]".

3.4. Evaluation metrics

Model performance was evaluated based on the variance between predicted and actual 2023 values. The absolute variance percentage was used to measure prediction accuracy, with lower percentages indicating more accurate predictions.

4. RESULTS AND ANALYSIS

4.1. Comparative performance of AI models

The findings revealed that ChatGPT-4 had the lowest average absolute variance (90.40%), followed by Claude 3.5 (97.19%) and Gemini (98.02%). ChatGPT-4 demonstrated superior accuracy across multiple metrics, aligning with its benchmark scores in mathematical reasoning and knowledge application (Hendrycks et al., 2021; Clark et al., 2018).

4.2. Benchmark scores vs. predictive accuracy

The analysis showed a strong negative correlation between high benchmark scores and lower variance percentages. For instance, Claude 3.5's high GSM8K score (95.0%) was reflected in its consistent revenue prediction performance. However, ChatGPT-4 outperformed in overall financial predictions despite slightly lower benchmark scores, suggesting additional contributing factors such as training data diversity or model architecture.

4.3. Observations on model strengths

Claude 3.5 excelled in predictions for companies with complex financial profiles, such as Tesla and Walmart, indicating its proficiency in handling varied financial data. ChatGPT-4's balanced performance across all metrics indicates its robustness in tasks requiring reasoning and numerical processing. The Gemini model, while competitive, needed to be more consistent in complex forecasting scenarios, potentially due to limitations in handling advanced mathematical reasoning.

5. DISCUSSION

5.1. Practical implications for financial professionals

The results underscore the potential for AI models to significantly enhance financial forecasting. Financial professionals can leverage benchmark scores as initial indicators when selecting AI tools, focusing on models with proven mathematical and reasoning capabilities. This can improve the accuracy of projections, aiding in strategic decision-making and risk management.

5.2. Challenges and recommendations

Despite the advantages, challenges such as data quality and model transparency persist. High-quality, diverse data sets are critical for achieving reliable predictions. Additionally, model interpretability remains a concern, especially for high-stakes decisions in finance.

6. CONCLUSION

This research demonstrates that benchmark scores are valuable indicators of an AI model's potential for accurate financial forecasting. However, factors beyond benchmark results can also influence real-world performance, such as training data and model architecture. The study highlights the importance of selecting appropriate models and ensuring high data quality and transparency for effective AI integration in finance.

REFERENCES

Balakrishnan, K. P., Prakash, L., & Ramesh, L. (2020). Impact of AI technology in accounting and finance. *The International Journal of Analytical and Experimental Modal Analysis, 12*(5), 629–636.

Cao, L. (2021). *AI in finance: Challenges, techniques and opportunities*. https://doi.org/10.2139/ssrn.3869625

Clark, P., Cowhey, I., Etzioni, O., Khot, T., Sabharwal, A., Schoenick, C., & Tafjord, O. (2018). Think you have solved the question answering? Try ARC, the AI2 reasoning challenge. *arXiv*. https://doi.org/10.48550/arXiv.1803.05457

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021). Measuring massive multitask language understanding. In *Proceedings of ICLR 2021 — 9th International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.2009.03300

Khattak, B. H. A., Shafi, I., Khan, A. S., Flores, E. S., Lara, R. G., Samad, M. A., & Ashraf, I. (2023). A systematic survey of AI models in financial market forecasting for profitability analysis. *IEEE Access, 11*, 125359–125380. https://doi.org/10.1109/ACCESS.2023.3330156

Ranaldi, L., Gerardi, M., & Fallucchi, F. (2022). CryptoNet: Using auto-regressive multi-layer artificial neural networks to predict financial time series. *Information, 13*(11), Article 524. https://doi.org/10.3390/info13110524

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing Journal, 90*, Article 106181. https://doi.org/10.1016/j.asoc.2020.106181