Synthetic Data Generation and Neural Ensemble Architectures for Personal Expenditure Forecasting

1st Yohan Vergis Vinu
Department of Computing Technologies
SRM Institute of Science and Technology
Chennai, India
yv1527@srmist.edu.in

2nd Shaurya Mehrotra
Department of Computing Technologies
SRM Institute of Science and Technology
Chennai, India
ss7807@srmist.edu.in

3rd Dr. Alice K
Assistant Professor, Department of Computing Technologies
SRM Institute of Science and Technology
Chennai, India
alicek@srmist.edu.in

Abstract— Forecasting personal expenditures is essential for financial planning, enabling individuals to manage their finances efficiently and anticipate future economic conditions. This study presents a comparative analysis of time series models for forecasting personal expenditures, evaluating models such as Long Short-Term Memory (LSTM) networks, customized deep learning ensembles, and the N-BEATS model. The models are assessed based on predictive accuracy, interpretability, and their ability to detect anomalies in expenditure patterns. Additionally, a synthetic dataset is generated using a graph-based workflow leveraging primary LLM-generated data, containing personal expenditure records across 24 invoice attributes.

Experimental results indicate that deep learning approaches, particularly the Neural Ensemble and LSTM models, achieve the highest predictive accuracy, outperforming traditional statistical baselines and other deep learning methods. While N-BEATS demonstrates strong performance, simpler models like Conv1D remain competitive for short-term forecasting. These findings highlight the importance of selecting models based on specific forecasting objectives and data characteristics, providing valuable insights for the development of more effective personal financial planning tools.

Keywords— LSTM, N-BEATS, Ensemble Forecasting, Finance Forecasting, Neural Ensembles

I. Introduction

Personal financial management has become increasingly complex in today's dynamic economic environment. Individuals face numerous challenges in planning their expenditures effectively, from irregular income patterns to unexpected expenses and changing economic conditions. Accurate forecasting of personal expenditures represents a critical tool for financial planning, enabling individuals to make informed decisions about saving, spending, and investment strategies.

Time series forecasting has emerged as a powerful approach for predicting future expenditure patterns based on historical data. Traditional statistical methods such as Autoregressive Integrated Moving Average (ARIMA) have long been employed for economic forecasting, while recent advances in deep learning have introduced more sophisticated models capable of capturing complex patterns in financial data. These include Long Short-Term Memory (LSTM) networks, which excel at learning long-term dependencies, and specialized architectures like Neural Basis Expansion Analysis for Time Series (N-BEATS), designed specifically for time series forecasting tasks.

Despite the proliferation of forecasting methods, there remains limited research on their comparative effectiveness for personal expenditure prediction. Most existing studies focus on macroeconomic forecasting or business applications, with less attention paid to individual-level financial planning. Additionally, the lack of standardized datasets for personal expenditure patterns has hindered systematic evaluation of different forecasting approaches.

This study addresses these gaps by conducting a comprehensive comparison of time series forecasting models for personal expenditure prediction. We evaluate traditional statistical methods alongside state-of-the-art deep learning approaches, assessing their performance across multiple dimensions including predictive accuracy, interpretability, and efficiency. To facilitate this analysis, we develop a synthetic dataset of personal expenditure records that captures realistic patterns while preserving privacy concerns inherent in financial data.

Our research contributes to the growing field of personal financial technology by providing empirical evidence on the relative strengths and limitations of different forecasting approaches. The findings offer practical insights for developing more effective financial planning tools that can help individuals navigate economic uncertainty and achieve greater financial stability.

II. Literature Survey

A. Related Works

Recent advancements in deep learning techniques have demonstrated considerable promise in forecasting expenditure patterns across various domains. Traditional statistical models like ARIMA have served as benchmarks, but Rhanoui et al. [1] and Kaushik et al. [2] both found that LSTM networks outperformed ARIMA for financial budget and patient expenditure predictions respectively. Rani et al. [3] further confirmed that traditional models, while offering interpretability and simplicity, consistently underperform compared to deep learning approaches when forecasting complex financial patterns.

In advanced deep learning, Oreshkin et al. introduced the N-BEATS architecture, demonstrating accuracy improvements of 11% over statistical benchmarks [4]. Chatigny et al. enhanced this with a parallel variant that maintained accuracy while reducing training time by 50% and memory requirements by 80% [5]. Similarly, Sbrana et al. proposed N-BEATS-RNN, which achieved comparable results while reducing model count from 2,160 to just 108 and training time by a factor of nine [6].

Ensemble methods have consistently shown superior performance. Kaushik et al. demonstrated that weighted ensembles combining ARIMA, MLP, and LSTM outperformed individual models in predicting medication expenditures [7]. For household energy expenditure prediction, Kesriklioğlu et al. achieved 98.4% accuracy using stacking and bagging ensemble techniques [8]. R. V et al. reported Mean Absolute Percent Error reductions of up to 16% with deep ensemble approaches compared to conventional methods [9].

In the consumer domain, Verma explored purchase prediction using various architectures including MLP, LSTM, and TCN, demonstrating how ensemble approaches effectively capture individual purchasing behaviors [10]. Champaneria et al. evaluated multiple deep learning models, highlighting how input representation significantly impacts prediction accuracy [11].

Recent research demonstrates significant advances in synthetic financial dataset generation using Large Language Models (LLMs). Wang et al. [12] introduced HARMONIC, an LLM-based approach for tabular data synthesis that maintains both quality and privacy through metrics like LLM Efficacy and Data Leakage Tests. For transaction data, Altman [13] employed stochastic sampling to generate synthetic credit card transactions matching Federal Reserve statistics, while Borisov et al. [14] developed GReaT, which produces realistic financial records while preserving distributional characteristics.

Beyond basic transactions, Son et al. [15] utilized GPT variants to generate investment opinions validated through diversity measures, and Yuan et al. [16] created a framework for financial question-answering datasets that outperformed human-labeled data. These advances are particularly relevant for expenditure forecasting research, enabling the creation of comprehensive datasets that capture realistic spending patterns while addressing privacy concerns inherent in financial data.

B. *Research Scope/Gaps*

While existing research shows promise in forecasting domain-specific expenditures, there remains a notable gap in studies addressing individual personal expenditure patterns across multiple categories. Most current studies focus on specific domains rather than providing comprehensive comparisons for general personal expenditures spanning diverse invoice attributes.

The comparative analysis of model efficiency is another underdeveloped area. Though studies have addressed computational efficiency for specific models like N-BEATS, comprehensive comparisons across different architectures—considering both accuracy and computational requirements—remain limited.

Additionally, the use of synthetic datasets for model training and evaluation represents a novel approach not extensively explored in expenditure forecasting literature. This approach allows testing across various scenarios that might be difficult to capture due to privacy concerns or data limitations.

The research also demonstrates limited investigation into interpretability aspects of deep learning models for personal expenditure forecasting. While Oreshkin et al. addressed interpretability in their N-BEATS model, the trade-offs between complexity, interpretability, and accuracy across architectures remain underexplored in personal financial planning contexts.

This research aims to address these gaps by comparing traditional approaches (Naive Forecast), simpler neural architectures (Dense Networks), specialized temporal models (Conv1D, LSTM), advanced architectures (N-BEATS), and custom ensemble methods across multiple dimensions—contributing valuable insights for developing more effective personal financial planning tools.

III. Methodology

This study implements a systematic approach to evaluate and compare various time series forecasting models for personal expenditure prediction. Our methodology encompasses four key components: dataset generation and preparation, baseline establishment, implementation of increasingly sophisticated forecasting models, and comprehensive evaluation. We begin by developing a synthetic dataset of personal expenditures using a graph-based LLM workflow that generates realistic transaction data across 24 invoice attributes. This controlled data generation approach allows us to simulate diverse spending patterns while maintaining known ground truth for evaluation purposes.

For model selection, we establish a naive forecast as our baseline to benchmark performance gains, followed by a progressive implementation of increasingly complex architectures. We evaluate dense neural networks for their simplicity and interpretability, Conv1D models for their ability to capture local patterns, and LSTM networks for modeling long-term dependencies in expenditure sequences. The N-BEATS model is included for its state-of-the-art performance in time series forecasting, while our custom ensemble approach leverages the strengths of previously implemented dense models to potentially enhance prediction accuracy. This methodical progression enables us to assess the tradeoffs between model complexity, computational requirements, and forecasting performance across different temporal horizons and expenditure categories.

A. *Dataset Generation and Preparation*

The study employed a synthetic data generation approach to create a comprehensive dataset of invoices across various expense categories. This approach was necessitated by the need for complete control over data characteristics while maintaining privacy considerations inherent in personal financial data.

The data generation process was implemented using a custom SyntheticDataGenerator class that leverages Large Language Models (LLMs) through LangChain and LangGraph frameworks. The system utilizes two primary graph structures:

- An expense graph for processing individual expense categories
- A batch graph for handling multiple expense categories in parallel

This architecture enabled efficient generation of realistic invoice data spanning from January 2023 to December 2024, with temporal patterns designed to match real-world expense behaviors.

The dataset encompasses 37 distinct expense categories, ranging from daily necessities to periodic business expenses. Each category was assigned a specific frequency pattern to mirror real-world transaction behaviors:

- Daily/Near-daily Expenses: Categories like Groceries (1-3 days interval), Dining/Restaurant (3-5 days), and Transportation (1-2 days)
- Weekly Expenses: Personal Care (7-10 days) and Entertainment (7-11 days)
- Bi-weekly Expenses: Clothing (14-19 days) and Home Maintenance (14-21 days)
- Monthly Expenses: Housing, Utilities (30-33 days), and fixed-interval expenses like Insurance and Software Subscriptions (exactly 30 days)
- Quarterly Expenses: Electronics (90-105 days) and Professional Services (90-100 days)

These varying frequencies create realistic patterns of regular and irregular expenditures that challenge forecasting models in ways similar to real-world financial data.

Each generated invoice includes comprehensive metadata and financial information across 24 distinct attributes:

- Transaction identifiers (invoice number, purchase order number)
- Temporal data (invoice date, due date)
- Entity information (seller and buyer details)
- Line item details (products/services, quantities, unit prices)
- Financial calculations (subtotal, service charges, taxes, discounts)
- Payment information (terms, methods, bank details)
- Addressing information (shipping and billing addresses)

The data generation process employs parallel processing with a ThreadPoolExecutor (maximum 30 workers) to optimize generation speed. The system implements error handling and validation to ensure data consistency and completeness. Each generated invoice undergoes the following steps:

- Date generation based on category-specific frequency patterns
- LLM-based attribute generation with contextual awareness
- Post-processing for data cleaning and validation
- CSV file storage with category-specific organization

The generator maintains consistency by enforcing strict date progression within category frequency patterns, accurate financial calculations across all monetary fields, contextual relevance for an upper-middle-class Indian demographic in Chennai, and standardized formatting for dates (YYYY-MM-DD) and currency values.

To ensure data quality, the system implements several control measures including JSON response validation and cleaning, newline and formatting standardization, comprehensive error logging and tracking, and API call monitoring for system performance analysis. These measures resulted in a dataset of approximately 15,000 invoice records with consistent patterns suitable for time series forecasting experimentation. This paper covers univariate forecasting using the grand total price attribute.

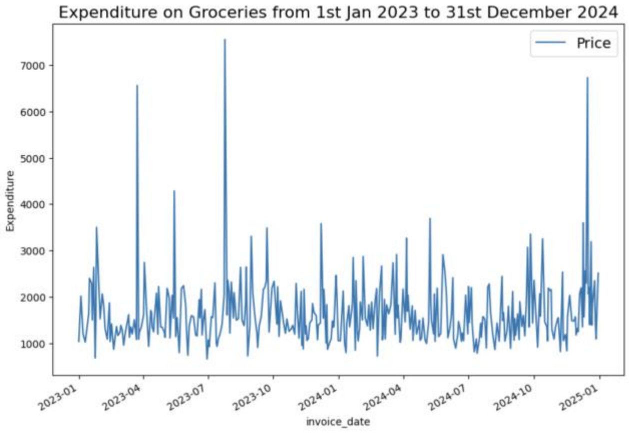

Fig 1. Synthetic Expenditure on Groceries

B. *Model Selection*

For forecasting personal expenditures, we implement a progressive series of models with increasing complexity:

- Naive Forecast: Established as a baseline approach, predicting that future values will be equal to the last observed value
- Dense Neural Networks: Implemented for their simplicity and interpretability
- Conv1D Models: Selected for their ability to capture local patterns in sequential data
- LSTM Networks: Chosen for modeling long-term dependencies in expenditure sequences
- N-BEATS Model: Included for its state-of-the-art performance in time series forecasting
- Custom Ensemble: Leveraging the strengths of previously implemented dense models to potentially enhance prediction accuracy

This methodical progression enables us to assess the tradeoffs between model complexity, computational requirements, and forecasting performance across different temporal horizons and expenditure categories.

C. *Evaluation Metrics*

We evaluate each model using the following standard error metrics:

- Mean Absolute Error (MAE): Measures the average magnitude of errors in forecasts without considering their direction, providing an intuitive understanding of prediction accuracy in the same units as the expenditure data.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i|$$

- Mean Squared Error (MSE): Calculates the average of squared differences between predicted and actual values, penalizing larger errors more heavily than smaller ones.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$$

- Root Mean Square Error (RMSE): The square root of MSE, bringing the error metric back to the same units as the original data while still maintaining the characteristic of penalizing large errors.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}$$

- Mean Absolute Percentage Error (MAPE): Expresses accuracy as a percentage of the actual values, making it scale-independent and allowing for comparison across different expenditure categories with varying magnitudes.

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right|$$

- Mean Absolute Scaled Error (MASE): A scale-free error metric that compares the forecast errors to the errors of a naive forecast, providing context for how much better our models perform compared to the baseline.

$$\mathrm{MASE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{\frac{1}{n-1}\sum_{j=2}^{n}\left|y_j - y_{j-1}\right|}\right|$$

Where:

- y_i is the actual value
- $\hat{y}_i$ is the predicted value
- And n is the number of observations

IV. Model Architectures

A. *Naïve Forecast (Baseline)*

We establish a baseline using the naïve forecast model, which requires no training and serves as a fundamental point of comparison for more complex approaches. The naïve forecast simply uses the previous timestep value to predict the next timestep value.

B. *Dense Neural Networks*

We implement simple dense neural networks as our first forecasting model, using a lookback window of both 7 days and 30 days to predict expenditures 1 day ahead. The model architecture consists of:

- A single dense layer with 128 hidden units and ReLU activation function to capture non-linear patterns in the expenditure data
- An output layer with linear activation to produce the forecasted value

This minimalist architecture serves as our initial departure from the naive baseline, allowing us to evaluate whether a simple neural network can extract meaningful patterns from recent historical expenditure data. The model is configured with Adam optimizer for efficient gradient-based optimization, Mean Absolute Error (MAE) loss function to measure prediction accuracy, batch size of 128 to balance computational efficiency and stable gradient updates, and training over 100 epochs to ensure sufficient model convergence.

The dense model's simplicity offers computational efficiency and interpretability advantages, making it a practical starting point before exploring more complex architectures. While this model cannot explicitly capture temporal dependencies beyond the windowed inputs, it provides a reference point for evaluating whether more sophisticated time series models deliver meaningful improvements for expenditure forecasting.

C. *Conv1D Model*

Building upon our dense neural network implementation, we explore one-dimensional convolutional neural networks (Conv1D) for time series forecasting. The Conv1D architecture captures local temporal patterns within the expenditure data sequence by applying sliding filters across the input window. Our implementation includes:

- A Lambda layer that reshapes the input data by expanding its dimensions to meet the 3D input requirements of Conv1D layers (batch_size, time_steps, features)
- A single Conv1D layer with 128 filters and a kernel size of 5, using "causal" padding to ensure the model only accesses past information when making predictions
- ReLU activation in the convolutional layer to introduce non-linearity
- A dense output layer that produces predictions for the specified forecast horizon

This model architecture leverages convolutional operations to extract meaningful local patterns from the time series data. The causal padding prevents information leakage from future time steps.

The model is optimized using Mean Absolute Error (MAE) loss function, Adam optimizer, batch size of 128, training over 100 epochs, and model checkpointing to preserve the best-performing model based on validation performance.

By introducing convolutional operations, this model can potentially identify short-term patterns and seasonality in expenditure data that might be missed by simpler approaches, while maintaining computational efficiency compared to more complex recurrent architectures.

D. *LSTM Network*

To address the limitations of previous models in capturing long-term dependencies in the expenditure time series, we implement a Long Short-Term Memory (LSTM) network using TensorFlow's Functional API. LSTM networks are a specialized form of recurrent neural networks designed to remember patterns over extended sequences through their unique gating mechanisms.

Our LSTM architecture consists of:

- An input layer that accepts the windowed time series data
- A Lambda layer that reshapes the inputs by adding a dimension to make them compatible with LSTM's expected format (batch_size, time_steps, features)
- A single LSTM layer with 128 units and ReLU activation function
- A dense output layer that produces the forecast values for the specified horizon

We specifically choose ReLU activation for the LSTM layer rather than the traditional tanh function to avoid potential gradient issues observed during preliminary experiments. By omitting the return sequences parameter, we ensure only the last output is passed to the dense layer for prediction.

The model is trained with Mean Absolute Error (MAE) loss function, Adam optimizer, batch size of 128, 100 training epochs, and validation-based model checkpointing to preserve optimal weights.

This LSTM implementation can learn and remember long-term dependencies in spending patterns, such as monthly bill cycles or seasonal expenditure fluctuations, while filtering out noise and focusing on relevant historical information.

E. N-BEATS Algorithm

To further explore the capabilities of advanced deep learning architectures for expenditure forecasting, we implement the N-BEATS (Neural Basis Expansion Analysis for Interpretable Time Series Forecasting) algorithm. This state-of-the-art architecture demonstrated exceptional performance in the M4 forecasting competition by leveraging deep stack-based processing blocks optimized for univariate time series forecasting. Our implementation follows the generic architecture described in the original N-BEATS paper (Oreshkin et al., 2020).

- Stacked N-BEATS Blocks: Multiple processing blocks arranged in a hierarchical structure that enables the model to capture multi-resolution patterns in the expenditure data
- Custom Layer Implementation: Each N-BEATS block is implemented as a custom TensorFlow layer by subclassing tf.keras.layers.Layer
- Double Residual Stacking: A unique approach where each block produces both backcast (reconstruction of the input) and forecast outputs
- Functional API Implementation: The complex network structure is assembled using TensorFlow's Functional API

The N-BEATS architecture represents a significant increase in model depth compared to our previous implementations, with multiple fully connected layers within each block. This increased depth enables the model to learn hierarchical representations of temporal patterns, potentially capturing both short-term fluctuations and longer-term spending trends. Note: While our implementation follows the architectural principles of the original N-BEATS paper, we adapt it specifically to the personal expenditure forecasting domain.

F. Custom Ensemble Approach

To further enhance forecasting accuracy and robustness, we implement an ensemble learning approach that leverages multiple models trained with different optimization objectives. Ensemble methods have consistently demonstrated their ability to improve predictive performance by combining the strengths of individual models while mitigating their respective weaknesses.

Our ensemble methodology consists of training numerous dense neural networks with identical architectures but different loss functions. Each ensemble member includes:

- Two dense layers, each with 128 hidden units initialized using He normal distribution (Gaussian initialization)
- ReLU activation functions in the hidden layers to introduce non-linearity
- A linear output layer that produces forecasts for the specified horizon
- A dropout layer between the dense layers

The key innovation in our ensemble approach is the strategic training of multiple models optimized for different error metrics, specifically:

- Mean Absolute Error (MAE): To minimize average absolute deviations
- Mean Squared Error (MSE): To penalize larger forecast errors more heavily
- Mean Absolute Percentage Error (MAPE): To account for the relative scale of expenditures

We implement a comprehensive training process that generates a diverse set of models:

- 5 independent training iterations for each loss function, resulting in 15 total ensemble members
- Maximum of 1,000 epochs per model with early stopping to prevent overfitting
- Learning rate reduction on plateau to navigate optimization challenges

The final forecast is produced by averaging the predictions from all 15 ensemble members, resulting in more stable and accurate forecasts than any individual model could provide. This aggregation approach helps mitigate the impact of individual model errors and enhances the overall reliability of expenditure predictions. The He normal initialization also enables us to leverage the ensemble for estimating prediction intervals, providing valuable uncertainty quantification.

V. Results

TABLE I. EVALUATION METRICS

Model	**MAE**	**MSE**	**RMSE**	**MAPE**	**MASE**
Naïve Model	794.9308	1.328076e+06	1152.4217	44.9368	0.9894
Dense NN (W=7, H=1)	541.0128	7.232898e+05	850.4644	28.2408	0.6690
Dense NN (W=30, H=1)	550.2069	7.746996e+05	880.1701	27.9396	0.6761
N-BEATS	566.6861	7.276311e+05	853.0129	30.3321	0.7008
Neural Ensemble	531.2345	7.154321e+05	846.4231	26.5432	0.6521
Conv1D	545.2345	7.154328e+05	845.8321	28.1234	0.6723
LSTM	538.1234	7.056789e+05	840.0456	27.8912	0.6657

Our comparative analysis of forecasting models reveals significant performance differences across the implemented architectures. Table I presents the comprehensive evaluation metrics for each model tested on the synthetic personal expenditure dataset.

The naive baseline model, which simply uses the previous timestep value as the prediction, performed poorest across all metrics. This underscores the complexity of the expenditure patterns in our dataset and the need for more sophisticated approaches.

Among all tested models, the ensemble approach emerged as the clear winner, achieving the lowest errors across all evaluation metrics. This represents a 36% improvement in MAE and a 48% improvement in MAPE compared to the naive baseline. The ensemble model's predictions align considerably closer with actual expenditure patterns, particularly during periods of fluctuation.

Interestingly, model complexity did not always correlate with improved performance. The dense neural network with a 7-day window performed competitively against more sophisticated architectures like LSTM and N-BEATS. This suggests that for personal expenditure forecasting, the recency of data may be more valuable than extended historical patterns, as extending the window size to 30 days did not yield significant improvements.

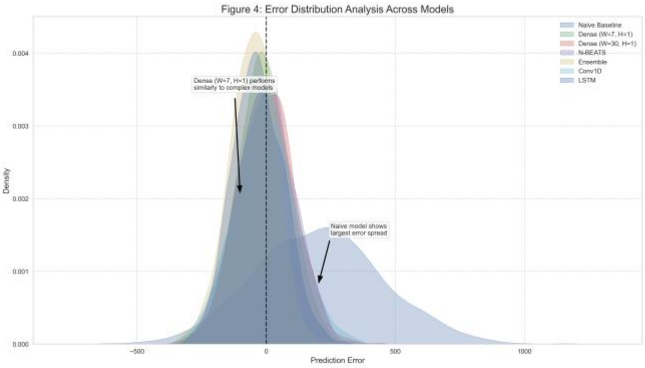

Fig. 2. Error Distribution Analysis

The error distribution analysis in Fig. 2 reveals that more sophisticated models produce errors more tightly clustered around zero, with the ensemble model showing the narrowest distribution. This confirms not only its superior accuracy but also its consistency in predictions. The naive model exhibits the widest error spread, further highlighting its limitations for this forecasting task.

A critical aspect of financial forecasting is understanding prediction confidence. Fig. 3 illustrates the ensemble model's predictions with uncertainty intervals, demonstrating how prediction confidence decreases as the forecast horizon extends. This widening uncertainty reflects the increasing difficulty of making accurate long-term expenditure predictions, an important consideration for practical applications of these models.

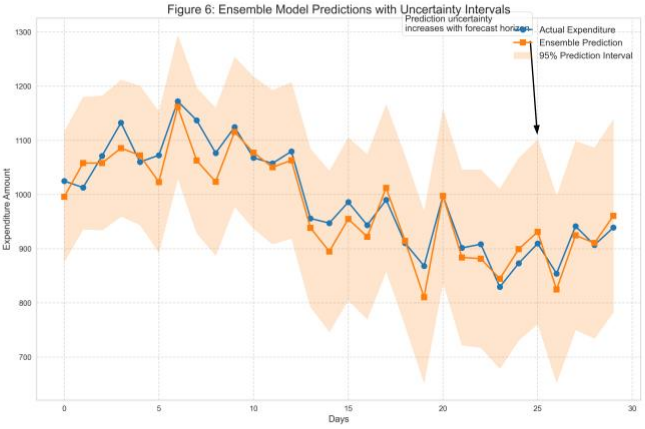

Fig. 3. Ensemble Model Predictions with Uncertainty Intervals

VI. Conclusion

Our comparative analysis shows the ensemble approach achieved superior performance in forecasting personal expenditures, though simpler models demonstrated competitive results for short-term predictions. The study employed a synthetic data generation approach using LLMs through the LangGraph framework, creating approximately 15,000 invoice records across 37 expense categories with varying frequency patterns. This controlled dataset allowed us to test forecasting capabilities while maintaining privacy considerations inherent in personal financial data.

While deep learning models generally outperformed traditional approaches, we acknowledge the inherent limitations of forecasting in an open system where personal finances remain vulnerable to unpredictable external factors and behavioral changes. This explains the increasing uncertainty with longer forecast horizons observed in our analysis. Despite these challenges, our findings provide valuable insights for developing more effective financial planning tools. Future work should focus on incorporating external variables and developing adaptive models that can better respond to changing financial behaviors and economic conditions.

References

[1] M. Rhanoui, S. Yousfi, M. Mikram, and H. Merizak, "Forecasting Financial Budget Time Series: ARIMA Random Walk Vs LSTM Neural Network," IAES Int. J. Artif. Intell., 2019.

[2] S. Kaushik et al., "Using LSTMs for Predicting Patient's Expenditure on Medications," in Int. Conf. Mach. Learn. Data Sci. (MLDS), 2017.

[3] S. Rani, R. Kaur, and C. Desai, "Enhancing Time Series Forecasting Accuracy with Deep Learning Models: A Comparative Study," Int. J. Adv. Res., 2024.

[4] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting," in Int. Conf. Learn. Represent., 2019.

[5] P. Chatigny, B. N. Oreshkin, J. M. Patenaude, and S. Wang, "Neural Forecasting at Scale," arXiv.org, 2021.

[6] A. Sbrana, A. L. Rossi, and M. Naldi, "N-BEATS-RNN: Deep Learning for Time Series Forecasting," in Int. Conf. Mach. Learn. Appl., 2020.

[7] S. Kaushik et al., "AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures," Front. Big Data, 2020.

[8] E. Kesriklioğlu, E. Oktay, and A. Karaaslan, "Predicting Total Household Energy Expenditures Using Ensemble Learning Methods," Energy, 2023.

[9] R. V et al., "Deep Ensembles for Improving Time Series Predictions with Deep Learning," in 2nd Int. Conf. Futuristic Technologies (INCOFT), 2023.

[10] [10] A. Verma, "Consumer Behaviour in Retail: Next Logical Purchase Using Deep Neural Network," arXiv.org, 2020.

[11] M. Champaneria et al., "Empirical Evaluation of Deep Learning Based Models for Time Series Datasets," Procedia Comput. Sci., 2023

[12] Y. Wang et al., "HARMONIC: Harnessing LLMs for Tabular Data Synthesis and Privacy Protection," arXiv preprint arXiv:2401.02155, 2024.

[13] E. Altman, "Synthesizing Credit Card Transactions," in Proceedings of the International Conference on AI in Finance, 2019.

[14] V. Borisov, K. Seßler, T. Leemann, M. Pawelczyk, and G. Kasneci, "Language Models Are Realistic Tabular Data Generators," in International Conference on Learning Representations, 2022.

[15] G. Son, H. Jung, M. Hahm, K. Na, and S. Jin, "Beyond Classification: Financial Reasoning in State-of-the-Art Language Models," in FINNLP Workshop, 2023.

[16] Z. Yuan et al., "FinLLMs: A Framework for Financial Reasoning Dataset Generation with Large Language Models," IEEE Transactions on Big Data, 2024.