Expense Insight (2024)

1. ***Akshat Kaushik*** *Department of Computer Science and Engineering, SRMIST Delhi NCR campus, Modinagar, Ghaziabad. Email-id:* [*an0846@srmist.edu.in*](mailto:an0846@srmist.edu.in)*.*
2. ***Aditya Gutpa*** *Department of Computer Science and Engineering, SRMIST Delhi NCR campus, Modinagar, Ghaziabad. Email-id:* [*ak6319@srmist.edu.in*](mailto:ak6319@srmist.edu.in)*.*
3. ***Nikhil Mittal*** *Department of Computer Science and Engineering, SRMIST Delhi NCR campus, Modinagar, Ghaziabad. Email-id:* [*nd0258@srmist.edu.in*](mailto:nd0258@srmist.edu.in)

***Rishab Ghatge*** *Department of Computer Science and Engineering, SRMIST Delhi NCR campus, Modinagar, Ghaziabad. Email-id:* [*rh7663@srmist.edu.in*](mailto:rh7663@srmist.edu.in)*.*

1. ***Asst. Prof. Mr. Mayank Gupta*** *Department of Computer Science and Engineering, SRMIST Delhi NCR campus, Modinagar, Ghaziabad.*
2. ***Asst. Prof. Mr. Gajender Kumar*** *Department of Computer Science and Engineering, SRMIST Delhi NCR campus, Modinagar, Ghaziabad*

[[1]](#footnote-1)

***Abstract*—** **This research focuses on expense forecasting using time series analysis, a critical component of effective financial management. We employed the Prophet model alongside traditional regression algorithms, including Decision Tree, Support Vector Regression (SVR), and Linear Regression. The primary objective is to predict future expenses based on historical data while ensuring the data's integrity through preprocessing and exploratory data analysis (EDA). Our findings contribute to understanding the performance of various forecasting methods, offering insights for practical applications in financial planning. The study highlights the importance of model selection based on the characteristics of the dataset and presents avenues for future research that could further enhance forecasting accuracy.**

# I. INTRODUCTION

e

XPENSE forecasting is a crucial element of effective financial management in both personal and organizational contexts. Accurate predictions can significantly enhance budgeting, resource allocation, and strategic planning processes. As organizations strive for financial sustainability, the ability to forecast expenses accurately becomes indispensable. In this increasingly data-driven landscape, leveraging advanced analytical techniques is essential for developing robust forecasting models.

Traditionally, expense forecasting has relied on simple statistical methods and linear regression techniques. However, these approaches often struggle to account for the complexities and non-linear relationships inherent in financial data. Consequently, there is a growing interest in more sophisticated modeling techniques that can better capture the nuances of financial behaviors over time.

This paper explores various modeling techniques, particularly focusing on time series analysis, to enhance forecasting accuracy. By comparing the performance of *Prophet, Decision Tree, SVR, and Linear Regression*, we aim to analyze their strengths and weaknesses in the context of expense forecasting. Our analysis not only evaluates the predictive capabilities of these models but also provides insights into their applicability in real-world financial scenarios.

The significance of this research lies in its practical applications for financial decision-making. Accurate forecasting not only aids in effective budget management but also empowers organizations to anticipate changes in cash flow, allocate resources more effectively, and improve overall financial health. Through our analysis, we aim to provide valuable insights that can be utilized by finance professionals across various sectors, from small businesses to large corporations.

# II. Methodology

## A. Data Preprocessing

Data preprocessing is essential for preparing the dataset for analysis. We followed a systematic approach to ensure the integrity and reliability of our data. Our preprocessing steps included:

## 1. Standardization

*Standardization* is a critical step in preparing numerical features for modeling. By transforming features to have a mean of 0 and a standard deviation of 1, we mitigate the risk of larger values disproportionately influencing model performance. This normalization ensures that all features contribute equally to the model, enhancing convergence during training. The impact of this step was particularly crucial for models like *SVR*, which are sensitive to the scale of input features. Additionally, we employed *Min-Max scaling* for certain features where preserving the original range was important.

## 2. Data Preprocessing

Outliers can adversely affect model accuracy and distort predictions. We employed statistical methods such as the *Z-score* and *Interquartile Range (IQR)* to identify and remove outliers. By applying these techniques, we created a more robust dataset, ultimately improving the models' performance and reliability. Visualizations such as box plots and scatter plots aided in understanding the distribution of the data and identifying outliers effectively. We also analyzed the context of outliers to determine if they represented valid anomalies or errors in data collection.

## 3. Missing Value Handling

Handling missing values is crucial for maintaining dataset integrity. We systematically checked for missing values and applied appropriate imputation techniques, such as *mean or median imputation* for numerical features and *mode imputation* for categorical features. To enhance our approach, we explored advanced methods like *k-nearest neighbors (KNN)* imputation and multiple imputation, ensuring that the dataset remained comprehensive and usable for modeling. We also conducted sensitivity analysis to assess how different imputation techniques affected the model performance.

## B. Exploratory Data Analysis (EDA)

EDA was conducted to derive insights into the dataset, guiding our modeling decisions. Key steps included:

* 1. Trend Analysis: We utilized visualizations such as line graphs and moving averages to identify overall trends in expense data over time. This process allowed us to observe patterns and shifts in spending behavior, revealing long-term trends that could inform our modeling approach. We also calculated rolling averages to smooth out short-term fluctuations and highlight longer-term trends.
  2. Seasonality Checks: Seasonal decomposition of the time series data helped analyze periodic fluctuations. By employing techniques such *as STL (Seasonal-Trend decomposition using Loess)*, we identified recurring patterns that could enhance model accuracy. This step was particularly important for the *Prophet model*, which relies on seasonal patterns to make accurate forecasts. We also visualized seasonal plots to better understand how expenses varied across different periods (monthly, quarterly, etc.).
  3. Correlation Analysis: Using heatmaps and correlation coefficients, we assessed relationships between different features and expenses, identifying key predictors for our forecasting models. Understanding these relationships enabled us to select relevant features and improve the predictive power of our models. We also employed scatter plots to visualize relationships and identify potential non-linear patterns that may not be captured by traditional correlation measures.

## C. Forecasting Models

The following models were implemented for expense forecasting:

## 1. Prophet Models

Developed by *Facebook*, *Prophet* is specifically designed for *forecasting time series data*. It excels at capturing *seasonal trends* and *managing missing values* effectively. Prophet utilizes an additive model that combines seasonal, trend, and holiday effects, making it suitable for a variety of forecasting tasks. The model’s flexibility allows for easy adjustments based on domain knowledge, providing a user-friendly interface for practitioners. One of the advantages of Prophet is its capability to handle *irregular time series data*, which is common in financial datasets.

## 2. Decision Models

*Decision Trees* are non-parametric models that can capture non-linear relationships within data. By recursively partitioning the dataset, Decision Trees identify complex interactions among features. They are easy to interpret and visualize, making them a popular choice for many predictive modeling tasks. Their ability to handle both numerical and categorical data further enhances their applicability in diverse scenarios. However, Decision Trees are prone to *overfitting*, especially when the tree depth is not properly controlled. To mitigate this, we employed techniques such as *pruning* and set maximum depth constraints.

## 3. Support Vector Regression (SVR)

*SVR* extends the *Support Vector Machine (SVM)* framework for regression tasks. It is particularly effective in high-dimensional spaces and can model complex relationships. By defining a margin of tolerance around the predicted values, SVR provides a robust approach to handling noise in data. The choice of *kernel function (e.g., linear, polynomial, RBF)* allows for additional flexibility, enabling SVR to adapt to various data distributions. We conducted hyperparameter tuning using techniques such as *Grid Search* and *Cross-Validation* to optimize the SVR performance.

## 4. Linear Regression

Linear Regression represents a traditional approach that model’s relationships between variables linearly. It serves as a useful baseline for comparison due to its simplicity and interpretability, though it may struggle with complex or non-linear data patterns. Despite its limitations, Linear Regression remains a foundational technique in statistical modeling and offers insights into the relationships between features and expenses. We employed *regularization techniques*, such as *Lasso and Ridge regression*, to enhance model performance and prevent overfitting.

# III. Literature Review

Expense tracker applications have gained significant attention in recent years due to their potential to enhance personal finance management. This literature review synthesizes recent findings and methodologies from various studies related to the development and application of these tools.

Market Overview and Trends

According to the "Expense Tracker Apps Global Market Report" by Zoho Corporation (2024), the global demand for expense tracker applications has surged, driven by an increasing focus on financial literacy and management. This trend is reflected in various studies exploring user preferences and functionalities of these applications (Dadhich, 2024; Wijaya, 2024).

Technical Developments and Machine Learning Integration

Several studies emphasize the integration of machine learning techniques to improve the functionality and predictive capabilities of expense trackers. For instance, Zhao and Tian (2020) present a hybrid model combining time series analysis and machine learning for financial forecasting, showcasing the potential for enhanced accuracy in expense predictions. Similarly, Park and Lee (2020) propose a personalized expense tracking system utilizing machine learning to tailor recommendations based on user behavior.

Raj Thakare et al. (2023) explore the application of the Naive Bayes algorithm in expense tracker applications, demonstrating its effectiveness in categorizing expenses and predicting future spending patterns. This approach highlights the growing role of artificial intelligence in automating financial management processes.

User-Centric Design and Functionality

The importance of user experience in expense tracker applications is a recurrent theme. The study by Anitha et al. (2016) focuses on the development of an easy-to-use Android application for expense management, emphasizing the need for intuitive interfaces to enhance user engagement. Similarly, S. Chandini et al. (2019) discuss the benefits of an online income and expense tracker, which offers users greater flexibility and accessibility.

Research by Yurochkin et al. (2021) delves into the development of applications specifically designed for expense accounting, reinforcing the need for tailored solutions that cater to diverse user needs.

Behavioral Insights and Predictive Modeling

Understanding spending behavior is crucial for the effectiveness of expense trackers. V. Singh et al. (2020) investigate the role of socio-mobile features in predicting spending behavior, suggesting that integrating social factors could enhance the predictive capabilities of expense management systems. This aligns with findings from Dadhich (2024), who highlights the importance of user data in shaping application features and functionalities.

Challenges and Future Directions

Despite the advancements, challenges remain in the field of expense tracking. The article by Nestor Maslej (2024) points out issues such as data privacy and the need for secure financial management tools. As highlighted by u and Chau (2019), there are opportunities for AI-enabled solutions, but they also bring challenges related to user trust and data security.

In summary, the literature on expense tracker applications indicates a dynamic intersection of technology, user behavior, and market demand. Future research should continue to explore innovative methodologies, user-centric designs, and address challenges related to data security and privacy to enhance the effectiveness of these applications. The continued integration of machine learning and AI will likely play a pivotal role in shaping the future landscape of personal finance management.

# IV. Guidelines for Graphics Preparation and Submission

## A. Types of Graphics

The following list outlines the different types of graphics published in IEEE journals. They are categorized based on their construction, and use of color / shades of gray:

* 1. **Color/Grayscale Figures**  
     Figures that are meant to appear in color, or shades of black/gray. Such figures may include photographs,   
     illustrations, multicolor graphs, and flowcharts.
  2. **Line Art Figures**  
     Figures that are composed of only black lines and shapes. These figures should have no shades or half-tones of gray, only black and white.
  3. **Tables**  
     Data charts which are typically black and white, but sometimes include color.

## B. Multipart Figures

These are figures compiled of more than one sub-figure presented side-by-side or stacked. If a multipart figure is made up of multiple figure types (one part is line art, and another is grayscale or color), the figure should meet the stricter guidelines.

## C. File Formats for Graphics

Format and save your graphics using a suitable graphics processing program that will allow you to create the images as PostScript (PS), Encapsulated PostScript (.EPS), Tagged Image File Format (.TIFF), Portable Document Format (.PDF), JPEG, or Portable Network Graphics (.PNG). These programs can re-size them and adjust the resolution settings. If you created your source files in one of the following programs you will be able to submit the graphics without converting to a PS, EPS, TIFF, PDF, or PNG file: Microsoft Word, Microsoft PowerPoint, or Microsoft Excel. Though it is not required, it is strongly recommended that these files be saved in PDF format rather than DOC, XLS, or PPT. Doing so will protect your figures from common font and arrow stroke issues that occur when working on the files across multiple platforms. When submitting your final files, your graphics should all be submitted individually in one of these formats along with the manuscript.

## D. Sizing of Graphics

Most charts, graphs, and tables are one column wide (3.5 inches / 88 mm / 21 picas) or page wide (7.16 inches / 181 millimeters / 43 picas). The maximum depth a graphic can be is 8.5 inches (216 millimeters / 54 picas). When choosing the depth of a graphic, please allow space for a caption. Figures can be sized between column and page widths if the author chooses, however, it is recommended that figures not be sized less than column width unless when necessary.

The final printed size of author photographs is exactly   
1 in wide by 1.25 in tall (25.4 mm x 31.75 mm / 6 picas x 7.5 picas). Author photos printed in editorials measure 1.59 in wide by 2 in tall (40 mm x 50 mm / 9.5 picas x 12 picas).

## E. Resolution

The proper resolution of your figures will depend on the type of figure it is as defined in the “Types of Figures” section. Author photographs, color, and grayscale figures should be at least 300dpi. Line art, including tables should be a minimum of 600dpi.

## F. Vector Art

In order to preserve the figures’ integrity across multiple computer platforms, we accept files in the following formats: .EPS/.PDF/.PS. All fonts must be embedded or text converted to outlines in order to achieve the best-quality results.

## G. Color Space

The term “color space” refers to the entire sum of colors that can be represented within the said medium. For our purposes, the three main color spaces are grayscale, RGB (red/green/blue), and CMYK (cyan/magenta/yellow/black). RGB is generally used with on-screen graphics, whereas CMYK is used for printing purposes.

All color figures should be generated in RGB or CMYK color space. Grayscale images should be submitted in grayscale color space. Line art may be provided in grayscale OR bitmap colorspace. Note that “bitmap colorspace” and “bitmap file format” are not the same thing. When bitmap color space is selected, .TIF/.TIFF/.PNG are the recommended file formats.

## H. Accepted Fonts Within Figures

When preparing your graphics, IEEE suggests that you use one of the following Open Type fonts: Times New Roman, Helvetica, Arial, Cambria, or Symbol. If you are supplying EPS, PS, or PDF files, all fonts must be embedded. Some fonts may only be native to your operating system; without the fonts embedded, parts of the graphic may be distorted or missing.

A safe option when finalizing your figures is to strip out the fonts before you save the files, creating “outline” type. This converts fonts to artwork which will appear uniformly on any screen.

## I. Using Labels Within Figures

1. **Figure Axis Labels**
   1. Figure axis labels are often a source of confusion. Use words rather than symbols. As an example, write the quantity “Magnetization” or “Magnetization *M*,” not just “*M*.” Put units in parentheses. Do not label axes only with units. For example, write “Magnetization (A/m)” or “Magnetization (Am−1),” not just “A/m.” Do not label axes with a ratio of quantities and units. For example, write “Temperature (K),” not “Temperature/K.”
   2. Multipliers can be especially confusing. Write “Magnetization (kA/m)” or “Magnetization (103 A/m).” Do not write “Magnetization (A/m) × 1000” because the reader would not know whether the top axis label means 16000 A/m or 0.016 A/m. Figure labels should be legible, approximately 8- to 10-point type.
2. **Subfigure Labels in Multipart Figures and Tables**

Multipart figures should be combined and labeled before final submission. Labels should appear centered below each subfigure in 8-point Times New Roman font in the format of (a) (b) (c).

## J. Referencing a Figure or Table Within Your Article

When referencing your figures and tables within your article, use the abbreviation “Fig.” even at the beginning of a sentence. Do not abbreviate “Table.” Tables should be numbered with Roman numerals.

## K. Submitting Your Graphics

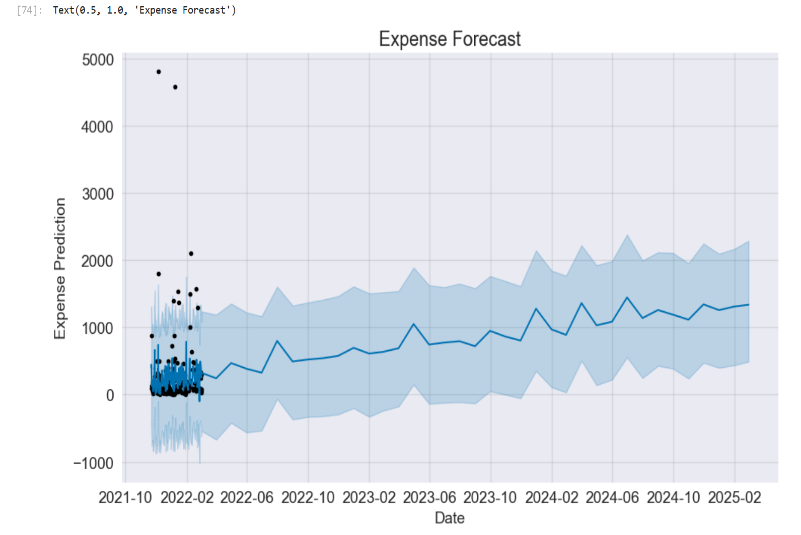
Because IEEE will do the final formatting of your article, all figures, figure captions, and tables can be placed at the end of your article. However, if you do place your figures within the article, they should be placed at the top of the page, closest to the first mention in the text. Figures should be submitted as individual files, separate from the manuscript in one of the file formats listed above. Place figure captions below the figures; place table headings above the tables. Do not include captions as part of the figures, or put them in “text boxes” linked to the figures. Also, do not place borders around the outside of your figures.

## L. Color Processing / Printing in IEEE Transactions, Journals, and Letters

All IEEE Transactions, Journals, and Letters allow an author to publish color figures on IEEE *Xplore* at no charge, and automatically convert them to grayscale for print versions. In most journals, figures and tables may alternatively be printed in color if an author chooses to do so. Please note that this service comes at an extra expense to the author. If you intend to have print color graphics, you will have the opportunity to indicate this in the Author Gateway and will be contacted by PubOps to confirm the charges.

# V. RESULTS

The results indicated varying levels of forecasting accuracy across the models. The Prophet model demonstrated superior performance in capturing seasonal trends, achieving the lowest *MSE* and *RMSE* values. Its effectiveness in dealing with missing values and seasonality made it a standout choice for this dataset. SVR and Decision Tree also showed promising results, particularly in capturing non-linear patterns, though their performance varied based on the complexity of the dataset. Linear Regression, while providing a reliable baseline, exhibited limitations in scenarios with complex interactions, as evidenced by higher error metrics.



# Appendix

Future research could explore hybrid models that combine the strengths of different algorithms or apply deep learning techniques for enhanced forecasting accuracy. Expanding the dataset to include more diverse features, such as macroeconomic indicators or categorical data (e.g., expense categories), may improve model performance further. Additionally, examining the impact of external factors such as market trends or economic fluctuations on expense forecasting could offer valuable insights for practitioners.

Additionally, implementing real-time forecasting systems that continuously update predictions based on new data would be a beneficial area of exploration, particularly in fast-changing financial environments. The integration of automation in forecasting processes can enhance decision-making and enable organizations to respond more agilely to changes.

Moreover, investigating the role of advanced time series techniques such as *ARIMA, LSTM (Long Short-Term Memory networks),* or *Bayesian forecasting* could provide additional depth to our understanding of forecasting performance. These advanced methods have shown promise in other domains and may yield valuable insights in the context of expense forecasting.

Finally, conducting case studies in different industries could further validate the findings and enhance the applicability of the proposed models. By applying these forecasting techniques to various contexts, we can gain a deeper understanding of their effectiveness and limitations, ultimately contributing to more informed financial decision-making.

References

1. Raj Thakare1, Ninad Thakare2, Raj Sangtani3, Shubham Bondre4, Amitkumar Maneka, “Expense Tracker Application using Naive Bayes” vol. 10 issue 3, June, 2023.
2. Sanjivani Patil, “Expense Tracker Report Using Python”, vol.12 , issue- 1(8), June 2024.
3. Rejwan Bin Sulaiman, G. Hariprasath, P. Dhinakaran , Utku Kose. “Time-series Forecasting of Web Traffic Using Prophet Machine Learning Model” , June 2023.
4. Draper, N.R., & Smith, H. (2014). Applied Regression Analysis. Wiley.
5. Y.Anitha, R.Ranjini, and S.Gomathi, “Easy App for Expense Manager Using Android,” International Journal of Trend in Research and Development, vol. 3, no. 2, pp. 65-67, 2016.
6. Hyndman, R.J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice.
7. Zoho Corporation Pvt. Ltd , “Expense Tracker Apps Global Market Report” , 2024.
8. Asst.Prof. Abhishek Dadhich, “Expense Tracker” , 2024.
9. International Journals of computer techniques, Vol: 3 Issue: 2, ISSN: 2394-2231
10. Cornellius Yudha Wijaya, KDnuggets Technical Content Specialist, 2024.
11. Manmohan Singh et al., “Email Spam Classification by Support Vector Machine,” International Conference on Computing, Power and Communication Technologies, IEEE, pp. 878-882, 2018.
12. Atiya Kazi et al., “Expense Tracker,” IRE Journals, vol. 4, no. 11, pp. 19-21, 2021.
13. Zhao, Y., & Tian, Y. “A hybrid model of time series analysis and machine learning for financial time series forecasting. Journal of Computational and Applied Mathematics”, 2020.
14. Denis E. Yurochkin, Anton A. Horoshiy, and Saveliy A. Karpukhin, "Development of an Application for Expense Accounting," IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, Moscow, Russian Federation, pp. 753-757, 2021.
15. S Gunabalan, “ Expense Tracker” , IJRASET52405 , 2023.
16. S. Chandini et al., “Online Income and Expense Tracker,” International Research Journal of Engineering and Technology, vol. 6, no. 3, pp. 4119-4124, 2019.
17. V. Singh, L. Freeman, B. Lepri, and A. Pentland, “Predicting spending behavior using socio-mobile features,”
18. u, Y., and Chau, M. (2019). AI-enabled personal finance management: Opportunities and challenges. Proceedings of the 40th International Conference on Information Systems, 1-15.
19. Park, S., and Lee, S. A Personalized expense tracking system based on machine learning. Journal of Internet Computing and Services, 21(3), 1-9, 2020.

1. [↑](#footnote-ref-1)