AN EMPRICAL STUDY ON MPCE PREDICTION USING REGRESSION AND MACHINE LEARNING TECHNIQUES

Dr. M. Navaneetha Krishnan, Professor & Head of Department
Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu
Email Id – mnksjce@gmail.com

Mr. Praveen Shylack A, Student
Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu
Email Id – apraveenvincy1@gmail.com

Mr. Joe Henry J, Student
Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu
Email Id – joemaths2004@gmail.com

Abstract - **Monthly Per Capita Expenditure (MPCE) is a key indicator used to measure household consumption and economic well-being. Traditional estimation techniques rely on basic statistical methods that often fail to capture complex socio-economic relationships. This project proposes a data-driven framework using regression and machine learning models for accurate MPCE prediction. Algorithms such as Linear Regression, Decision Trees, Random Forest, and Gradient Boosting are applied to structured household survey data. Feature engineering and preprocessing techniques enhance model performance and reliability. The models are evaluated using RMSE, MAE, and R^2 metrics to ensure predictive accuracy. The proposed system supports poverty analysis, welfare planning, and evidence-based policy formulation through improved expenditure forecasting.**

I. Introduction

Monthly Per Capita Expenditure (MPCE) is a crucial indicator for measuring household consumption patterns and overall economic well-being. Accurate estimation of MPCE helps governments and policymakers assess poverty levels and design welfare programs effectively. Traditional estimation methods mainly rely on periodic surveys and basic statistical techniques. However, these conventional approaches often fail to capture complex relationships among socio-economic variables. With the growth of large-scale household datasets, there is a need for more advanced analytical techniques. Machine learning and regression models provide better predictive capabilities by identifying hidden patterns in data.

This project proposes a data-driven framework for MPCE prediction using statistical and machine learning algorithms. The system analyzes socio-economic attributes such as income, education, occupation, and family size. By applying multiple regression and ensemble learning methods, the model improves prediction accuracy. Overall, the proposed approach enhances scalability, reliability, and evidence-based socio-economic decision making.

II. Background and Motivation

A. Overview

This project focuses on predicting Monthly Per Capita Expenditure (MPCE) using regression and machine learning techniques. MPCE is an important economic indicator that reflects household consumption and living standards. Traditional estimation methods rely on survey averages and basic statistical analysis, which may not capture complex socio-economic relationships. The proposed system introduces a data-driven predictive framework for improved accuracy. It utilizes structured household survey datasets containing income, education, occupation, family size, and location attributes. The system performs data preprocessing to clean, normalize, and prepare the dataset for analysis. Feature engineering and selection techniques identify the most influential socio-economic variables affecting MPCE. Multiple models such as Linear Regression, Decision Trees, Random Forest, and Gradient Boosting are implemented. These algorithms learn both linear and non-linear relationships within the data. Model performance is evaluated using RMSE, MAE, and R^2 metrics. The best-performing model is selected for accurate MPCE prediction. The system supports both household-level and regional-level expenditure estimation. Visualization dashboards present actual vs predicted values for better interpretation. The framework enhances scalability when handling large datasets. Overall, the project provides reliable insights for poverty assessment, welfare planning, and socio-economic policy formulation.

B. Importance of the project

The prediction of Monthly Per Capita Expenditure (MPCE) is crucial for understanding household consumption behavior and economic conditions. Accurate MPCE estimation helps governments identify poverty levels and income disparities. Traditional statistical approaches often fail to capture complex and non-linear socio-economic relationships.

This project introduces machine learning techniques to enhance predictive accuracy. The system improves scalability when handling large household datasets. It reduces dependency on manual statistical estimation methods. Feature selection techniques help identify key socio-economic factors influencing expenditure. The use of ensemble learning models improves reliability and robustness of predictions.

Evaluation metrics such as RMSE, MAE, and R^2 ensure performance transparency. The framework supports both household-level and regional-level analysis. It enables policymakers to design targeted welfare and subsidy programs. The system can assist in monitoring economic development trends. Predictive modelling helps forecast future expenditure patterns effectively. Data visualization dashboards make results understandable for non-technical users. The approach strengthens evidence-based decision-making in socio-economic planning.

Overall, the project contributes to efficient, scalable, and data-driven economic policy formulation.

C. Motivation for This Research

Traditional MPCE (Monthly Per Capita Expenditure) estimation methods have long relied on basic statistical averages, which often fail to capture the true complexity of household economic behaviour. Socio-economic variables such as income, education, occupation, and family size interact in highly non-linear ways, making simplistic approaches inadequate for accurate prediction. As a result, traditional models risk oversimplifying the dynamics of expenditure, leading to limited insights for policymakers and researchers.

Conventional regression models, while widely used, struggle to capture hidden patterns in high-dimensional datasets. Household expenditure data is inherently complex, with multiple interacting variables that cannot be fully explained through linear relationships. This limitation reduces the effectiveness of regression-based approaches in identifying subtle but influential socio-economic factors. Consequently, there is a growing need for more advanced analytical frameworks that can uncover deeper insights from large-scale survey data.

Large-scale household surveys generate vast amounts of information, requiring scalable and automated analytical approaches. Manual statistical analysis of such datasets is not only time-consuming but also prone to human error, which undermines the reliability of results. Automated machine learning techniques offer a promising alternative by handling large datasets efficiently, reducing human intervention, and improving accuracy. This shift toward automation is essential for modern socio-economic research, where timely and reliable insights are critical.

Accurate MPCE prediction plays a vital role in poverty assessment and welfare policy design. Reliable expenditure estimates help governments and organizations identify vulnerable populations, allocate resources effectively, and design targeted welfare programs. Without precise predictions, policy interventions risk being misaligned with actual needs, reducing their impact. Thus, improving MPCE estimation is not just a technical challenge but a social imperative for equitable development.

Machine learning provides a powerful toolkit for addressing these challenges. Feature engineering can significantly improve prediction accuracy by identifying influential factors that traditional models often overlook. Advanced algorithms such as Random Forest and Gradient Boosting are particularly effective in handling complex, non-linear relationships within socio-economic data. These models can capture intricate dependencies between variables, offering more robust and reliable predictions compared to conventional approaches.

Equally important is the use of reliable evaluation metrics to ensure transparency and model validation. Metrics such as RMSE, MAE, and R^2 provide objective measures of predictive performance, allowing researchers to compare models and select the most effective ones. Transparent evaluation builds trust in AI-driven systems and ensures that predictions can be confidently used for policy-making.

Looking ahead, data-driven forecasting has the potential to support proactive economic planning and budgeting. Regional-level expenditure prediction can enable targeted policy implementation, ensuring that resources are directed where they are most needed. The integration of predictive analytics into governance enhances decision-making efficiency and supports evidence-based policy design. Modern governance increasingly demands intelligent systems for socio-economic monitoring, and this research aims to bridge the gap between traditional estimation methods and AI-driven prediction frameworks, creating a more accurate, scalable, and socially impactful approach to MPCE estimation.

III. Novel Applications of the project

The proposed MPCE prediction framework introduces several innovative applications in socio-economic planning and governance. It can be deployed as a real-time poverty monitoring system to identify economically vulnerable households. The model enables targeted welfare scheme distribution based on predicted expenditure levels. It supports district-level and state-level expenditure analysis for efficient resource allocation.

Governments can use the system for dynamic budget forecasting and financial planning. The framework enhances national census and survey accuracy through data-driven estimation methods. Financial institutions can apply MPCE prediction for credit risk assessment and loan eligibility evaluation. The system assists in rural and urban development planning by identifying low-consumption regions. It can be used to monitor food and essential commodity expenditure trends. Policymakers can analyse consumption inequality and income disparity more effectively.

The platform supports economic impact analysis during inflation or financial crises. It enables predictive modelling of future household expenditure trends. Integration with digital governance portals ensures automated socio-economic monitoring. The model can assist NGOs in identifying target beneficiaries for welfare programs. Researchers can use the framework for advanced empirical economic studies. It supports comparative analysis between demographic groups and regions.

The system can aid in subsidy optimization and public distribution planning. Interactive dashboards make insights accessible to non-technical decision makers. The scalable architecture allows nationwide deployment on large datasets. Overall, the project establishes an intelligent, data-driven ecosystem for socio-economic forecasting and policy formulation.

IV. Role and Potential

1. Role in Enhancing Socio-Economic Analysis

The project plays a significant role in improving the accuracy of MPCE estimation using data-driven methods. It captures complex and non-linear relationships among socio-economic variables.The framework replaces traditional average-based estimation with predictive modelling techniques. It enables detailed household-level and regional-level economic analysis. Advanced machine learning models improve reliability compared to conventional regression methods. Thus, the system strengthens empirical research in economic and poverty analysis.

2. Role in Supporting Policy Formulation

The system provides reliable expenditure predictions that assist policymakers in welfare planning. It helps identify economically weaker sections for targeted subsidy allocation. Regional MPCE comparison enables balanced resource distribution across districts and states. The predictive insights support poverty alleviation and development strategies. Accurate data improves transparency and evidence-based governance. Overall, the framework enhances strategic socio-economic decision making.

3. Potential for Scalable and Advanced Analytics

The architecture supports large-scale household survey datasets efficiently. It integrates multiple models such as Linear Regression, Random Forest, and Gradient Boosting. Feature engineering enhances model performance and reduces computational complexity. Cloud-based deployment can enable nationwide implementation. The system can be extended to incorporate deep learning and time-series forecasting. This makes the framework adaptable for future economic monitoring systems.

4. Potential for Cross-Domain Applications

The MPCE prediction model can be applied in banking and credit risk assessment. It supports financial planning and consumption trend forecasting. The framework can assist NGOs and development agencies in beneficiary identification. It can be integrated with digital governance portals for automated monitoring. Researchers can extend the system for broader macroeconomic modelling. Thus, the project demonstrates high adaptability and multi-sector potential.

V. Conclusion

This study presents a data-driven framework for predicting Monthly Per Capita Expenditure (MPCE) using regression and machine learning techniques, moving beyond traditional statistical estimation methods to capture complex and non-linear relationships among socio-economic factors. By implementing models such as Linear Regression, Decision Trees, Random Forest, and Gradient Boosting, the system achieves improved prediction accuracy, while evaluation metrics including RMSE, MAE, and R^2 ensure reliable performance comparison. The framework supports both household-level and regional-level expenditure estimation, offering valuable insights for poverty analysis, welfare planning, and policy formulation. Overall, the proposed approach enhances scalability, accuracy, and decision-making capability in socio-economic analysis, bridging the gap between conventional estimation practices and modern AI-driven prediction systems.

VI. Future Research Directions

- Integration of Deep Learning Models:
- Future work can incorporate Artificial Neural Networks (ANN) and Deep Learning techniques to improve MPCE prediction accuracy.
- Time-Series Forecasting Models:
- Applying ARIMA, LSTM, or other time-series models can help forecast future expenditure trends dynamically.
- Inclusion of Real-Time Economic Indicators:
- Integrating inflation rates, GDP growth, and employment statistics can enhance predictive reliability.
- Expansion to National-Level Datasets:
- Testing the model on large-scale national survey datasets can improve generalization and robustness.
- Hybrid Modeling Approaches:
- Combining statistical and ensemble machine learning methods may produce more accurate and stable predictions.
- Automated Feature Engineering:
- Using advanced feature selection and dimensionality reduction techniques like PCA can optimize model performance.
- Explainable AI (XAI) Integration:
- Implementing SHAP or LIME can improve model interpretability for policymakers.
- Cloud-Based Deployment Framework:
- Developing scalable cloud architecture for real-time prediction and monitoring.
- Interactive Web-Based Dashboard Development:
- Creating a user-friendly platform for policymakers to visualize predictions and trends.
- Integration with Government Data Portals:

- Linking the system with NSSO and other official databases for automated socio-economic monitoring and policy support.

Reference:

[1] Ramanela, V. T., Nooraeni, R., Sumarni, C., & Farhan, M., "Data Integration for Improving the Precision of Per Capita Expenditure Estimation Using Machine Learning Method," 2024.

[2] Chand, D. B., "Autoregressive Integrated Moving Average Predictive Modelling for Per Capita GDP of Nepal," 2024.

[3] World Bank, World Development Indicators: Household Consumption and Expenditure Data, 2023.

[4] National Sample Survey Office (NSSO), Household Consumer Expenditure Survey Report, Government of India, 2023.

[5] Breiman, L., "Random Forests," Machine Learning Journal, Vol. 45, No. 1, pp. 5–32, 2001.

[6] Friedman, J. H., "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, Vol. 29, No. 5, pp. 1189–1232, 2001.

[7] Pedregosa, F., Varoquaux, G., Gramfort, A., et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, Vol. 12, pp. 2825–2830, 2011.