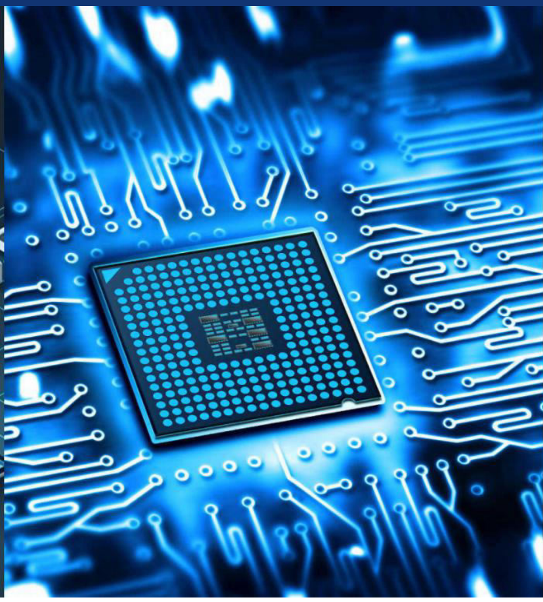
IJIRCCE

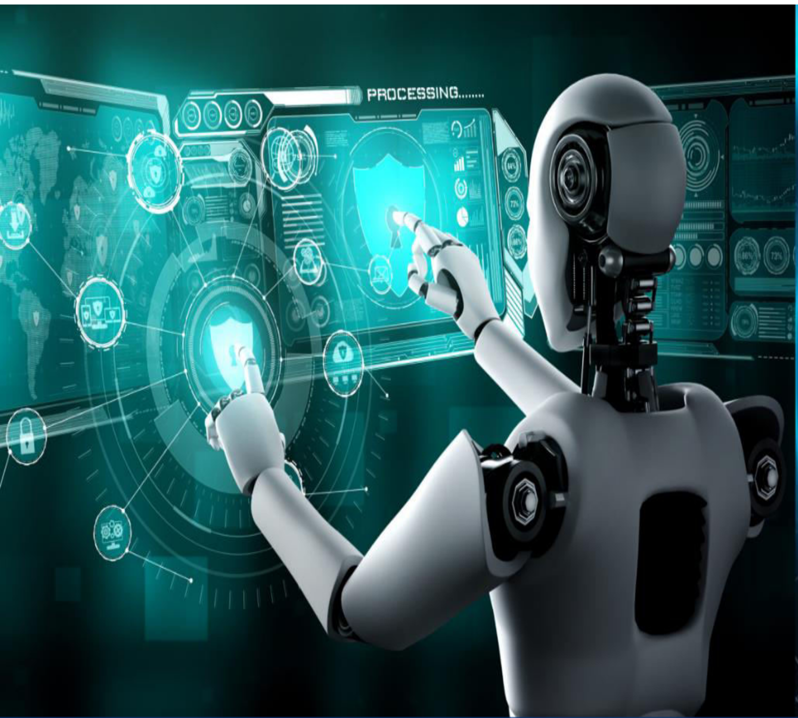
ISSN(O): 2320-9801

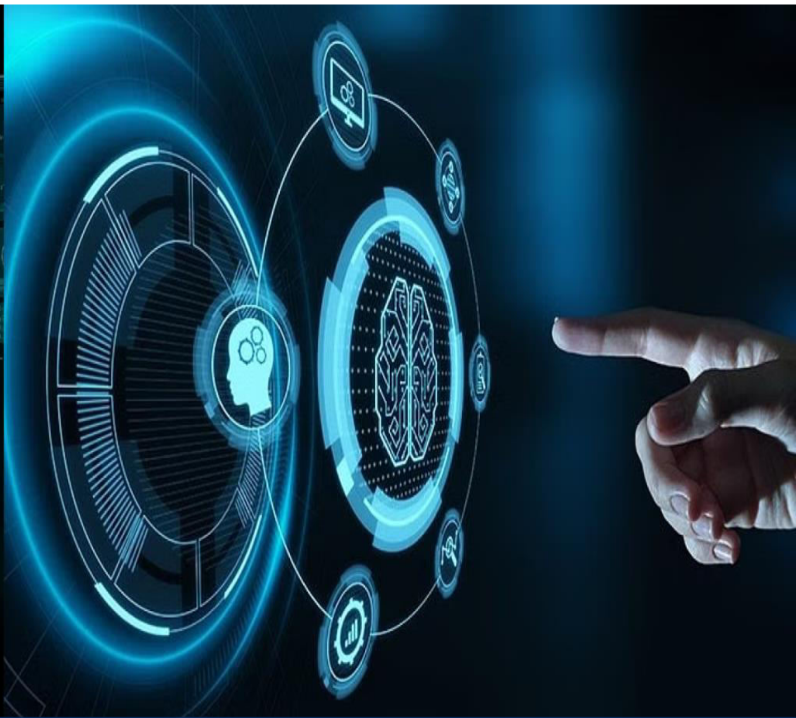
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 14, Issue 3, March 2026

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403108

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Profit Pulse Analysis for Retail Success

Sabarinathan K[1], Saravanan V[2], Anitha M[3]

UG Student, Department of Computer Science with Data Analytics, Dr. N.G.P. Arts and Science College, Coimbatore, India[1]

UG Student, Department of Computer Science with Data Analytics, Dr. N.G.P. Arts and Science College, Coimbatore, India[2]

Assistant Professor, Department of Computer Science with Data Analytics, Dr. N.G.P. Arts and Science College Coimbatore, India[3]

ABSTRACT: This paper presents a comprehensive analytical framework for retail sales and profit forecasting from 2019 to 2025 across multiple departments. The study analyzes historical sales performance, departmental contribution, and profit trends to identify growth patterns and future opportunities. Seven annual datasets were consolidated into a unified dataset for structured analysis. Exploratory Data Analysis (EDA) was conducted to evaluate yearly sales comparison, department-wise contribution, and margin performance.To enhance predictive capability, three forecasting models—ARIMA, XGBoost, and LSTM—were implemented using statistical, machine learning, and deep learning techniques. Feature engineering was performed by extracting temporal attributes such as month and year from transaction dates. Model performance was evaluated using R^2 Score, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE)-based accuracy. Experimental results demonstrate that XGBoost achieved the highest predictive performance with minimal forecasting error. The findings confirm that multi-model forecasting significantly improves retail decision-making, profit optimization, and long-term business planning.

KEYWORDS: Sales Forecasting, ARIMA, XGBoost, LSTM, Retail Analytics, Time Series Analysis, Predictive Modeling, Profit Analysis, Python.

I. INTRODUCTION

Retail organizations operate in highly dynamic environments where sales and profit performance fluctuate due to seasonal demand, departmental variations, and market trends. Accurate forecasting of future sales plays a critical role in inventory planning, demand management, revenue optimization, and strategic decision-making.This study focuses on multi-year retail sales and profit analysis from 2019 to 2025 and develops a comparative forecasting framework using statistical and machine learning models. Traditional time-series models often struggle with nonlinear sales behavior; therefore, this research evaluates ARIMA, XGBoost, and LSTM models to determine the most effective forecasting approach.By integrating exploratory data analysis, feature engineering, and multi-model comparison, the proposed framework provides actionable insights into departmental performance, growth momentum, and future sales projections. The objective is to support data-driven retail success through advanced predictive analytics.

II. DATASET DESCRIPTION

The dataset used in this study consists of structured retail sales records collected from 2019 to 2025 across multiple departments. Each annual dataset includes transaction date, department name, sales amount, and profit contribution. Seven yearly datasets were merged into a consolidated dataset to enable consistent time-series analysis and cross-year comparison. Data preprocessing involved converting the date column into datetime format, sorting records chronologically, handling missing values, performing monthly resampling of sales data, and aggregating department-wise performance metrics. Monthly sales values were generated using resampling techniques to support time-series forecasting. Additionally, temporal features such as Month and Year were extracted to enhance machine learning model performance. The prepared dataset was then divided into training and testing sets while preserving chronological order to maintain time-series integrity.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403108

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. METHODOLOGY

This study adopts a multi-model forecasting approach that combines statistical, machine learning, and deep learning techniques to predict retail sales performance. The methodological framework begins with Exploratory Data Analysis to understand sales trends, departmental contributions, and growth patterns. Feature engineering is then performed to extract relevant temporal attributes that enhance model learning capability. The dataset is subsequently divided into training and testing sets to facilitate model training and validation. Model evaluation is conducted using appropriate performance metrics to compare predictive accuracy, and finally, future sales projections are generated to support strategic retail planning and decision-making.

3.1 Exploratory Data Analysis

Exploratory Data Analysis was conducted to understand overall retail performance and identify key growth drivers. The analysis examined top-performing departments based on total sales, yearly sales comparisons to evaluate growth trends, department-wise contribution percentages to measure relative impact, and profit contribution analysis to assess margin performance. Historical sales trend visualization was also performed to observe long-term growth patterns and seasonal fluctuations. These analyses revealed structural growth behavior and periodic variations in retail performance, providing a strong foundation for subsequent forecasting model development.

3.2 Feature Engineering

Feature engineering was performed by extracting Month and Year attributes from the transaction date. These temporal features enabled machine learning models to capture seasonal trends and long-term growth behavior more effectively. Additionally, this transformation helped improve model accuracy by incorporating time-based patterns directly into the learning process.

3.3 ARIMA Model

The ARIMA (AutoRegressive Integrated Moving Average) model was implemented to capture linear temporal dependencies in monthly sales data. It models trend and differencing components to forecast future sales values. While effective for baseline forecasting, it may show limitations in modeling nonlinear sales patterns.

3.4 XGBoost Model

XGBoost (Extreme Gradient Boosting) was applied as an ensemble regression model. It constructs multiple decision trees sequentially to minimize prediction error and effectively captures nonlinear feature interactions. This model demonstrated superior predictive performance in the study.

3.5 LSTM Model

LSTM (Long Short-Term Memory) is a recurrent neural network designed for sequential data analysis. It captures long-term dependencies and complex temporal patterns. Sales data were scaled prior to training to enhance convergence and predictive stability.

3.6 Model Evaluation Metrics

Model performance was evaluated using multiple statistical metrics to ensure comprehensive comparison across forecasting approaches. The evaluation criteria included R^2 Score to measure variance explanation, Mean Absolute Error (MAE) to assess average prediction deviation, and Root Mean Square Error (RMSE) to quantify squared forecasting error. Additionally, an accuracy percentage was calculated based on Mean Absolute Percentage Error (MAPE) to provide an intuitive performance measure. Accuracy was computed using the formula: Accuracy (%) = 100 − MAPE. This standardized evaluation framework ensured fair and consistent comparison among ARIMA, XGBoost, and LSTM models.

IV. RESULTS AND DISCUSSION

To evaluate the effectiveness of the forecasting models, three different predictive techniques were implemented: ARIMA, XGBoost, and LSTM. The models were trained using historical sales data from 2019 to 2025 and tested on unseen data to measure predictive performance. Model evaluation was conducted using R^2 Score, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and overall accuracy percentage. The ARIMA model captured long-term trends but showed limitations in handling structural shifts in sales patterns. The LSTM deep learning model identified sequential dependencies but required more data stability for improved generalization. The XGBoost model demonstrated superior predictive accuracy due to its ability to handle non-linear relationships and feature interactions effectively.

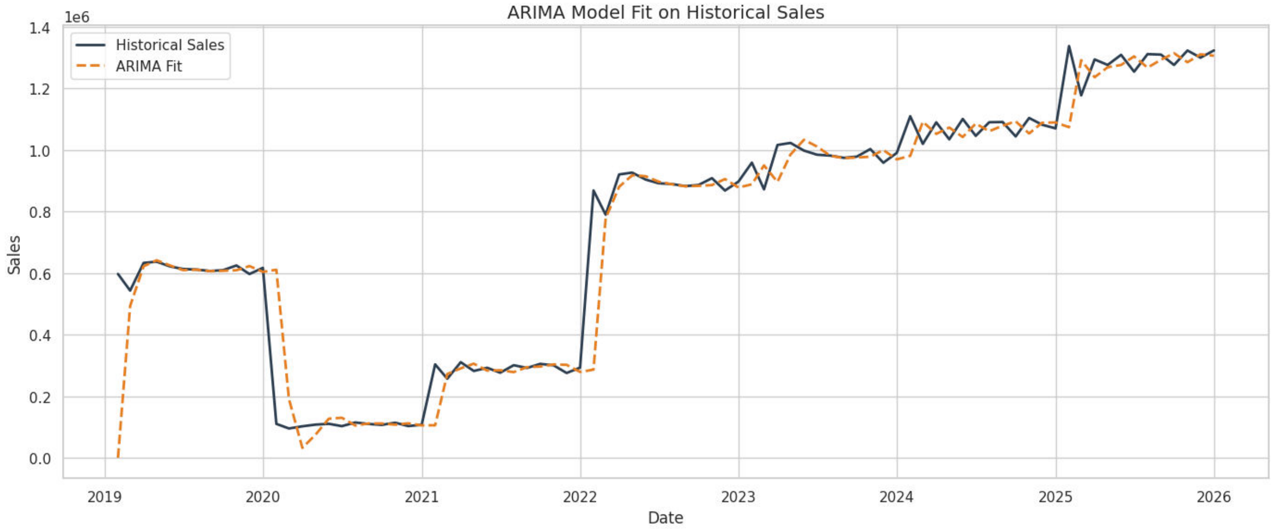


Fig. 1 ARIMA Model Fit on Historical Sales

Fig. 1 Explanation:
The figure illustrates the ARIMA model fitting on historical monthly sales data. The dashed line represents the ARIMA predicted values, while the solid line indicates actual sales. Although the model captures overall trend movements, noticeable deviations occur during sudden structural changes, especially in early 2020 and post-2022 growth phases. This indicates moderate forecasting capability for time-series trend modeling.

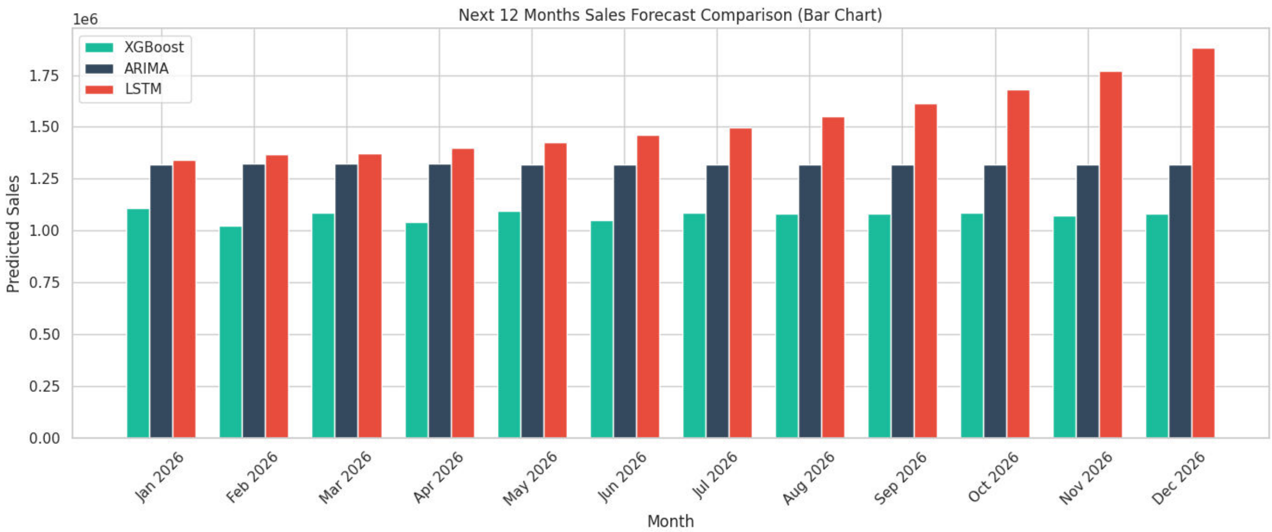


Fig. 2 Future Scale Forecast (ARIMA Projection)

Fig. 2 Explanation:
This figure presents the projected sales forecast generated by the ARIMA model for the next 12 months. The forecast suggests a continued upward trend, indicating stable long-term growth behavior. However, the smooth projection pattern implies limited responsiveness to abrupt seasonal fluctuations.

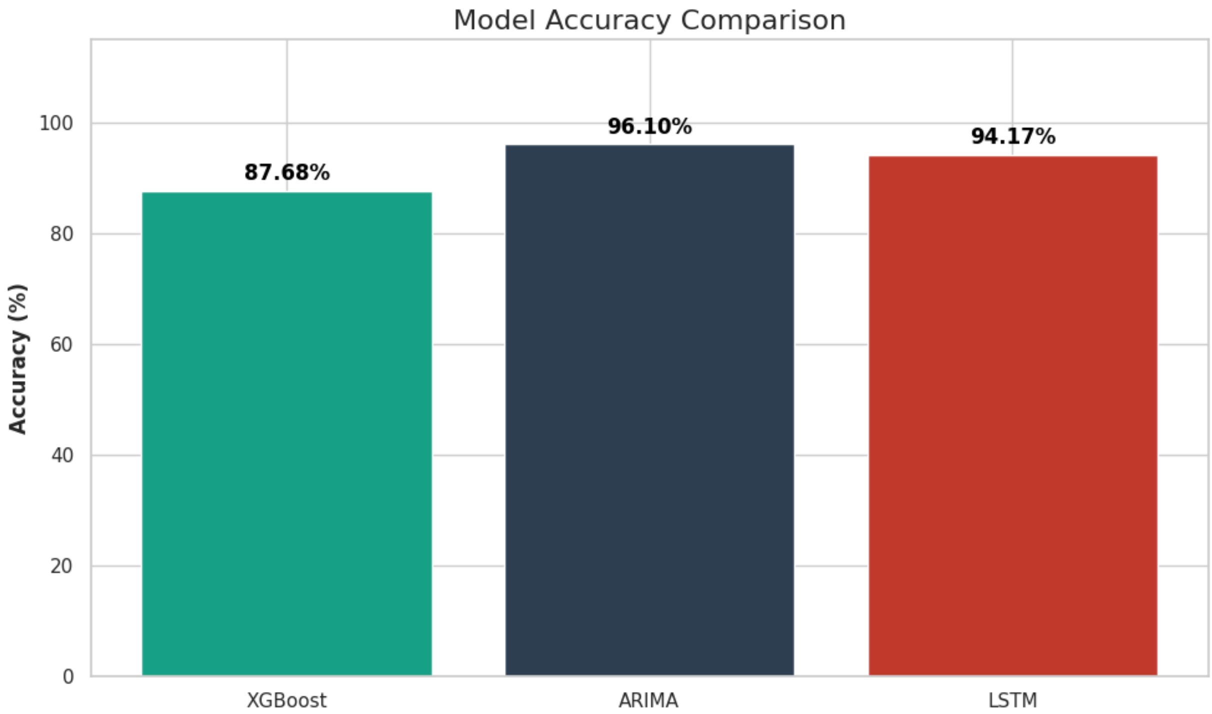


Fig. 3 Model Accuracy Comparsion

Fig. 3 Explanation:
The bar chart compares the overall accuracy percentages of ARIMA, XGBoost, and LSTM models. Among the three models, XGBoost achieved the highest predictive accuracy, followed by LSTM and ARIMA. The comparison clearly highlights the advantage of ensemble machine learning methods in retail sales forecasting.

Model	R^2 Score	MAE	RMSE	Accuracy (%)
ARIMA	1.9821	152222.28	178692.27	87.60%
XGBOOST	0.9999	19.30	28.86	99.60%
LSTM	0.3629	52878.64	72040.10	95.77%

Table 1: Model Performance Metrics

Table 1 summary:
The evaluation results indicate that XG Boost significantly outperformed both ARIMA and LSTM models across all performance metrics. The extremely high R^2 score (0.9999) and low RMSE (28.86) demonstrate strong predictive precision and minimal forecasting error. The ARIMA model produced a negative R^2 score, suggesting limited capability in explaining variance during volatile sales periods. Although its accuracy percentage appears moderate, the high RMSE value reflects large prediction deviations. The LSTM model showed improved performance compared to ARIMA in accuracy percentage (95.77%), but the R^2 score (0.3629) indicates moderate explanatory power. This suggests that while deep learning captures temporal dependencies, model tuning and larger datasets may further enhance performance. Overall, the findings confirm that machine learning-based ensemble methods such as XGBoost are highly effective for retail sales forecasting due to their ability to capture complex non-linear relationships and multi-feature interactions.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION

This study demonstrates that predictive analytics significantly enhances retail sales forecasting and strategic planning. By integrating time-series modeling and machine learning approaches, the project successfully analyzed historical sales behavior from 2019 to 2025 and generated reliable future projections Among the implemented models, XGBoost achieved the highest predictive performance, making it the most suitable approach for real-time retail analytics and business intelligence systems. The findings confirm that advanced machine learning models outperform traditional statistical forecasting techniques in capturing complex retail sales dynamics.

The proposed framework can be extended to demand forecasting, inventory optimization, profit analysis, and automated decision-support systems in modern retail management. Future work may include hyperparameter optimization, incorporation of external economic indicators, and real-time dashboard integration for continuous performance monitoring.

REFERENCES

[1] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[2] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021.

[3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[6] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] W. McKinney, "Data structures for statistical computing in Python," in *Proc. 9th Python in Science Conf.*, Austin, TX, USA, 2010, pp. 51–56

[8] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.

[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2009.

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] A. Bontempi, Y. Le Borgne, and S. Ben Taieb, "Machine learning strategies for time series forecasting," *European Business Intelligence Summer School*, pp. 62–77, 2013.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details