International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Sentiment-Driven Passenger Satisfaction Prediction in Metro Surveys using a Hybrid Model

Vimala T , Sayli B. Patil, Gayatri Gaikwad, Rutuja Divekar
Dept. of Computer Science,Nowrosjee Wadia College(Autonomous), Pune,
Maharashtra, India

Abstract - **Passenger satisfaction assessment plays a crucial role in improving the performance and service quality of urban metro systems. This study presents a sentiment-driven predictive analytics framework for analyzing commuter satisfaction in the Pune Metro system using a hybrid deep learning and machine learning approach. A primary dataset of 750 responses was collected through a structured commuter survey evaluating attributes such as safety, accessibility, affordability, sustainability, and service quality. The reliability of the survey instrument was validated using Cronbach's Alpha and McDonald's Omega.**

Text preprocessing techniques such as tokenization and padding were applied to prepare textual responses for analysis. A Gated Recurrent Unit (GRU) model was implemented to perform sentiment classification on passenger feedback. The sentiment outputs generated from the GRU model were then integrated with structured survey features and used as input to a Random Forest classifier for final prediction. Model performance was evaluated using metrics such as precision, recall, F1-score, and confusion matrix. Feature importance analysis revealed that digital payment systems, station facilities, and last-mile connectivity significantly influence passenger satisfaction.

The study demonstrates that integrating deep learning-based sentiment analysis with machine learning techniques enhances predictive capability and provides meaningful insights for improving metro services and supporting data-driven decision-making in urban transportation systems.

Keywords **—Passenger Satisfaction, GRU, Random Forest, Sentiment Analysis, Predictive Analytics**

1. INTRODUCTION

Pune Metro is a major urban transportation project designed to enhance daily commuting by offering a safe, efficient, and sustainable mobility solution for the rapidly growing population of the city. Since the metro started running, the number of passengers has been increasing, and a significant amount of useful data is being generated. By analysing this data, we can gain insights into passenger satisfaction, travel patterns, accessibility issues, and the overall efficiency of the metro system. Data analytics helps identify passenger preferences, the challenges they face, and the areas where improvements are required.

In this study, data from Pune Metro surveys and operations is analysed using various techniques. The work includes data cleaning, grouping passengers based on age, gender, and travel purpose, and examining how factors such as ticket cost, frequency, connectivity, and safety influence passenger satisfaction. Sentiment analysis was conducted using a GRU model on textual responses. However, due to limitations in the data, a Random Forest model was further applied for predictive analysis on structured survey attributes such as safety, comfort, affordability, and accessibility. The survey data was also analysed using hypothesis testing to understand metro usage patterns and public perception of its impact.

The aim of this study is to better understand commuter behaviour and evaluate the performance of Pune Metro. The findings can help metro authorities improve services, plan future expansions, and support data-driven decision-making for public transport development. The remaining sections of the report include a literature review, results and analysis, followed by the conclusion and future work.

2. BACKGROUND & RELATED WORK

Data analysis plays a crucial role in improving public transport systems such as Pune Metro. Before performing analysis, it is essential to ensure that the data is clean and accurate. Studies have shown that visual data-profiling tools help in detecting errors, missing values, and anomalies, thereby making large datasets easier to manage and analyse [1][2]. Effective data cleaning methods are particularly important when dealing with data collected from multiple sources, as inconsistencies and errors are common [5][6].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Once the data is preprocessed, appropriate analytical techniques are required to extract meaningful insights. Research indicates that descriptive, diagnostic, and predictive analytics help in understanding trends, identifying issues, and supporting decision-making in real-world applications [3][4]. In addition, visualization tools enable clear presentation of results to stakeholders, thereby improving service planning and management [7][8][9].

Deep learning models such as Gated Recurrent Units (GRU) are widely used for analysing sequential data, including passenger feedback and travel behaviour patterns [10][13]. Furthermore, advanced architectures such as Transformers have shown improved performance in capturing complex textual relationships [11]. However, for structured datasets, traditional machine learning algorithms such as Random Forest remain highly effective and are often combined with deep learning models to enhance predictive performance.

Random Forest is a supervised machine learning algorithm used for both classification and regression tasks [20]. It is an ensemble learning technique that constructs multiple decision trees during training and produces the final output based on majority voting. This method improves prediction accuracy and reduces overfitting by aggregating the results of multiple trees [18][19]. Additionally, Random Forest provides feature importance measures, making it useful for identifying key factors influencing outcomes.

These approaches collectively help in understanding commuter behaviour, identifying operational challenges, and predicting future trends. By integrating data cleaning, visualization, and predictive modelling techniques, this study aims to provide meaningful insights for improving metro services and supporting data-driven decision-making [3][5][7].Hypothesis testing is a widely used statistical technique for evaluating assumptions based on sample data. It enables researchers to determine whether there is sufficient evidence to support or reject a proposed hypothesis under given conditions [15][16].

3. SYSTEM ARCHITECTURE AND METHODOLOGY

Data Science is an advanced discipline that extends traditional statistics to analyze large volumes of data. It enables the extraction of meaningful insights from both structured and unstructured data to support effective decision-making. The Data Science lifecycle follows an iterative process, which includes data collection, data preprocessing, data exploration, and model building. In this study, these steps are applied to analyse passenger behaviour, optimize metro services, and predict ridership trends. The Fig.1. depicts the Data preprocessing steps.

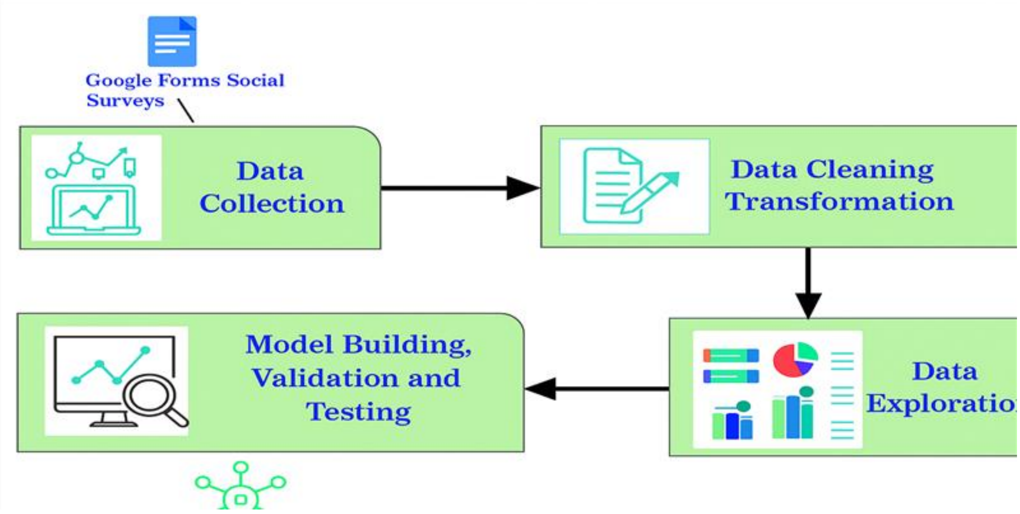

Fig.1. Data Processing Steps

3.1 Data Collection

Data collection refers to the process of gathering information from various sources for analysis. In this study, data was collected using Google Forms, which included passenger survey responses along with metro usage information.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

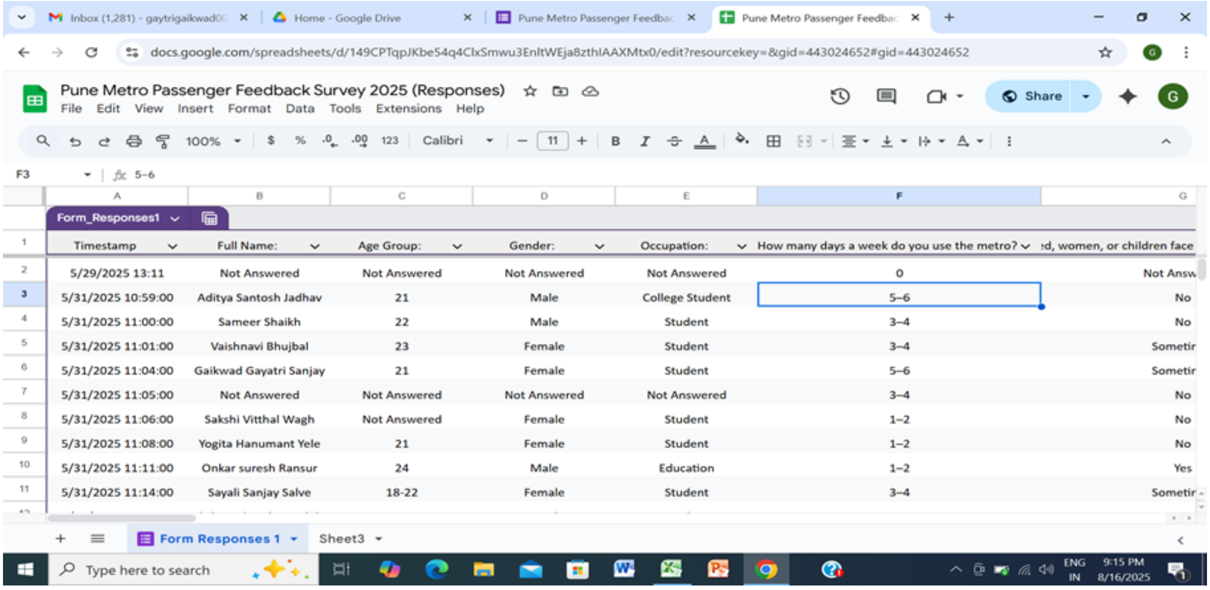

Fig.2. Screenshot of Google Form Response Sheet

The survey targeted individuals aged 16–60 who use the Pune Metro, with an approximate sample size of 750 participants. The questionnaire consisted of 25 questions, including Likert-scale questions to measure attitudes and multiple-choice questions for demographic and behavioural analysis. The Fig.2. Depicts the screenshot of google form response sheet.

3.2 Data Preprocessing

Data preprocessing involves data cleaning, data integration, and data transformation to convert raw data into a suitable format for analysis. This includes handling missing values, removing inconsistencies, combining data from multiple sources, and transforming it into a structured format.

Two validation techniques were employed in this study: stratified sampling and k-fold cross-validation. Stratified sampling is a probability sampling technique in which the population is divided into homogeneous subgroups (strata) based on shared characteristics. K-fold cross-validation is used to evaluate machine learning models by dividing the dataset into K equal parts, training the model on (K–1) parts, and testing it on the remaining part. This process is repeated K times, ensuring that each subset is used once for validation.

3.2.1 Reliability Measure of Responses

Cronbach’s Alpha is used to assess whether the survey questions consistently measure the same underlying concept and is widely accepted as a standard reliability measure. McDonald’s Omega is employed as it provides a more accurate estimate of reliability, particularly when individual survey items contribute unequally. The use of both measures ensures a robust and reliable evaluation of the metro passenger survey.

Cronbach’s Alpha (CA) is one of the most widely used statistical measures for evaluating the internal consistency of survey instruments [3]. It determines how closely related a set of questionnaire items are and whether they measure the same construct [5]. In this study, Cronbach’s Alpha was calculated to evaluate the reliability of the 25-item Pune Metro passenger survey.

McDonald’s Omega (ω) is considered a more robust reliability coefficient, especially when survey items have unequal factor loadings [12]. Unlike Cronbach’s Alpha, which assumes equal contribution of items, Omega uses factor analysis to estimate the reliability of the latent construct, making it less sensitive to statistical assumption violations [9].

In this study, both Cronbach’s Alpha and McDonald’s Omega were computed using Jamovi statistical software to evaluate the reliability of the questionnaire.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

The reliability analysis was performed using Jamovi, while overall data processing and modelling were carried out using Python. These tools facilitate better data understanding and interpretation. The Fig.3. Depicts the results of reliability checking.

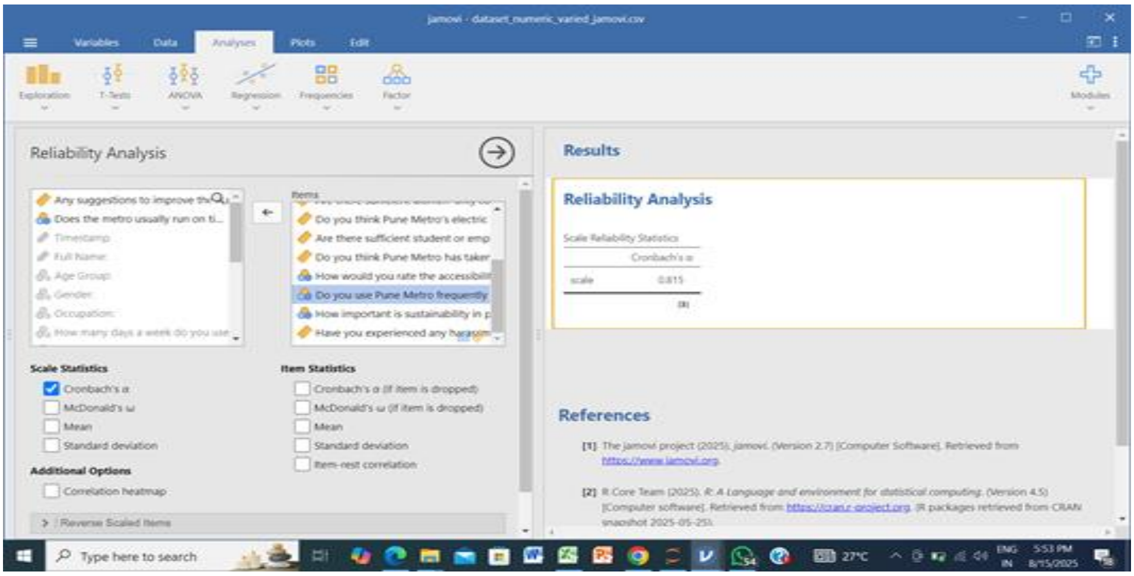

Fig.3. Reliability Check

3.2.2 Hypothesis Testing

The statistical analysis of the 750 survey responses provided significant insights into multiple hypotheses [17]. The t-test is one of the most commonly used statistical methods. The two-sample t-test compares the difference between two means relative to the variation in the data. It allows the calculation of a p-value using the t-test statistic [18]. The p-value represents the probability of observing a test statistic as extreme as, or more extreme than, the observed value,.

Hypothesis Testing 1: Metro Usage Frequency (Days per Week)

The average number of days respondents use the metro is approximately 4.87 days per week. A one-sample t-test indicates that this value is significantly higher than 3 days. Therefore, the null hypothesis is rejected, suggesting that metro usage is frequent among respondents.

Hypothesis Testing 2: Traffic and Pollution Reduction Perception

Approximately 54.6% of respondents believe that metro usage contributes to reducing traffic congestion and pollution. The result of the proportion (Z-test) is statistically significant; therefore, the null hypothesis is rejected. This indicates that a significant proportion of users perceive the metro as beneficial for reducing traffic and pollution. The Table.1. Depicts the hypothesis testing results.

Hypothesis Testing 3: Night Safety Perception

The proportion of respondents who feel safe using the metro at night is approximately 50.8%. Statistical testing shows no significant evidence to support a strong positive or negative perception. Therefore, the null hypothesis is not rejected, indicating mixed or neutral opinions regarding night-time safety.

Question	Test Type	Value (Mean / Proportion)	Test Statistic	p-value	Decision	Interpretation
Days per Week	One-sample t-test	4.865	t = 8.199	6.29×10^{-14}	Reject H_0	Average metro usage is significantly higher than 3 days/week
Traffic & Pollution Reduction	Proportion (Z-test)	0.546	Z = 2.532	0.0113	Reject H_0	Significant number of respondents believe metro reduces traffic & pollution
Night Safety	Proportion (Z-test)	0.503	Z = 0.183	0.8550	Fail to Reject H_0	No significant opinion on night safety
Time Management Impact	One-sample t-test	2.645	t = -7.712	3.96×10^{-14}	Reject H_0	Metro has a significant impact on time management

Table.1. Hypothesis testing results

Hypothesis Testing 4: Impact on Time Management

The average score for time management impact is 2.65. The t-test result is statistically significant, leading to the rejection of the null hypothesis. This suggests that metro usage has a significant impact on users' time management.

3.3 Model Building

Model building involves developing predictive frameworks using machine learning and deep learning techniques to analyse commuter satisfaction and identify key factors influencing Pune Metro service quality.

For textual data processing, tokenization and padding techniques were applied to prepare input data for the deep learning model. Padding was used to ensure a fixed input length for model compatibility. Both textual and structured survey data were processed using different modelling approaches. Textual responses were used for sentiment analysis, while structured survey attributes were used for predictive modelling. The Fig.4. Depicts the model building workflow.

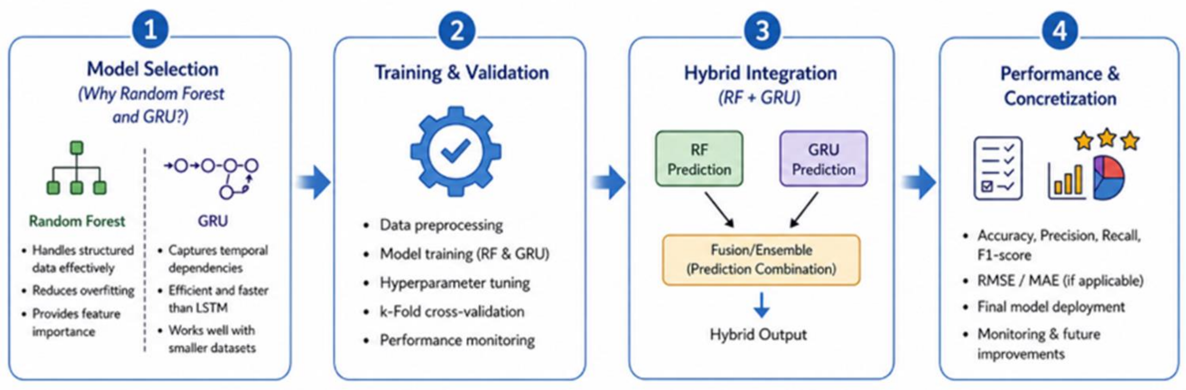

Fig.4. Model Building Workflow

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

The deep learning architecture was implemented using a Gated Recurrent Unit (GRU) model consisting of the following layers:
Embedding Layer: Converts tokenized words into dense vector representations
GRU Layer (128 units): Captures sequential dependencies in textual data
Dropout Layer (0.5): Prevents overfitting
Dense Output Layer: Performs multi-class sentiment classification (positive, neutral, and negative).
GRU-based models are widely used for sentiment and emotion analysis tasks due to their ability to capture sequential dependencies in textual data [14].

The GRU model was trained for **10 epochs** using the **Adam optimizer** [22] with the following parameters: The Table.2. Depicts the parameters used in GRU.

Parameter	**Value**
Optimizer	Adam
Learning Rate	0.001
Beta 1	0.9
Beta 2	0.999
Epsilon	1×10^{-7}

Table.2. The total number of trainable parameters in the GRU model was 1,379,459.

To improve predictive performance on structured survey attributes, a Random Forest classifier was implemented [20]. This model analysed features such as accessibility, facilities, safety, sustainability, digital payment systems, and route preferences. A hybrid approach was adopted, where sentiment outputs from the GRU model were combined with structured features for final prediction. The Table.3. Predicts the GRU model summary.

GRU SENTIMENT MODEL SUMMARY:

Layer (Type)	**Output Shape**	**Number of Parameters**
Input Layer	(None, 100)	0
Embedding Layer	(None, 100, 128)	1,280,000
GRU Layer	(None, 128)	99,072
Dropout Layer	(None, 128)	0

Total params: 1,379,459 (5.26 MB)
Trainable params: 1,379,459 (5.26 MB)
Non-trainable params: 0 (0.00 B)

Table.3. GRU Model Summary

OUTPUT:

Full Name	**Overall Satisfaction**	**Score Satisfaction Level**
Ishwari takalkar	51.700001	Bad
Poonam kolte	88.500000	Good
Madhushalini	88.500000	Good
Om bapu darekar	78.599998	Average
Vinayak	56.500000	Bad
Hitesh Dattatray Gaikwad	81.300003	Good
Sahil Sejal	77.800003	Average

Riya tatyaram Gangawane	41.900002	Bad
Arya Deshmukh	93.199997	Good
Sakshi Ramesh Randhave	89.000000	Good

...Total number of names: 749

Fig. 5. Sample output of satisfaction score and satisfaction level

The Fig.5. Shows the sample output of the satisfaction score and level. The GRU model was used to predict the overall satisfaction score of Pune Metro passengers. Based on the predicted scores, satisfaction levels were categorized into Bad, Average, and Good. The majority of respondents fall under the Good category, indicating a positive perception of metro services.

4. RESULTS AND DISCUSSION

Model validation refers to evaluating how well a machine learning model performs on unseen data. In this study, two validation techniques were used:

i) Stratified Sampling

The dataset was split into 70% training data and 30% testing data using the train_test_split function. The parameter *stratify = y* was used to maintain class distribution across both sets. This approach reduces bias and ensures reliable performance evaluation [3].

ii) K-Fold Cross-Validation

In this method, the dataset is divided into K equal subsets (folds). Each fold is used once as the test set, while the remaining folds are used for training. This process is repeated K times to evaluate model stability and consistency. The Fig.6. depicts the distribution of metro usage frequency across different age groups.

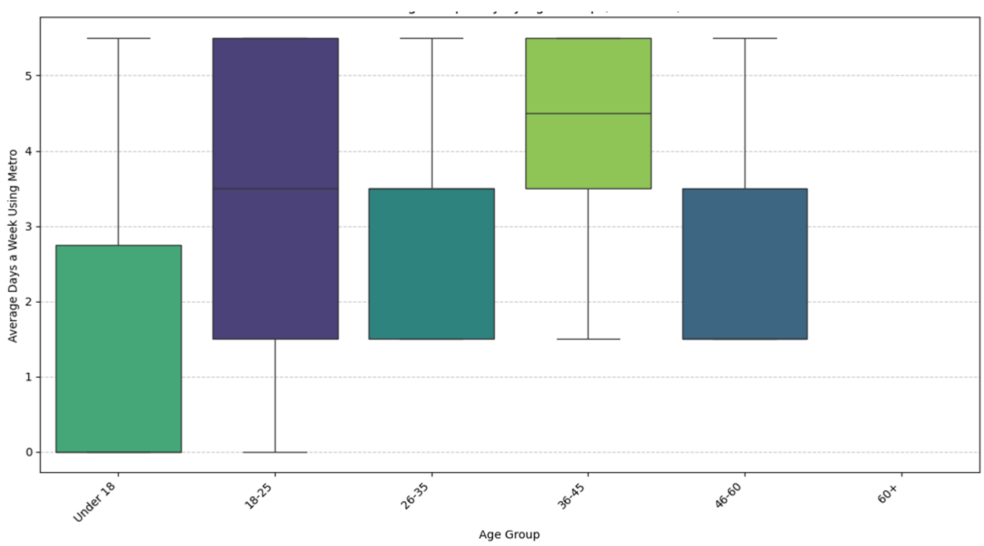

Fig.6. Distribution of metro usage frequency across different age groups.

4.1 Model Validation and Performance

The performance of both GRU and Random Forest models was evaluated using accuracy, precision, recall, F1-score, and confusion matrix.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

a) Random Forest Model Performance

The Random Forest classifier was applied to structured survey data combined with sentiment features derived from the GRU model. This ensemble learning method improves prediction accuracy and reduces overfitting by aggregating multiple decision trees [18][19][20]. Table 4 presents the classification performance of the Random Forest classifier.

Metric	Value
Accuracy	0.7133
Precision	0.7290
Recall	0.7133
F1-score	0.7053

Table.4. Classification Report of Random Forest Classifier

Feature importance analysis identified the following as the most influential factors affecting passenger satisfaction:

Rank	Feature	Importance Score
1	Fare price perception	0.0346
2	Time management impact	0.0381
3	Sustainability importance	0.0465
4	Accessibility rating	0.0538
5	Gender	0.0566
6	Route usage frequency	0.0684
7	Digital payment improvements	0.0686
8	Station facilities	0.0698
9	Occupation	0.0806
10	Last-mile connectivity	0.0819

Table 5. Feature Importance Analysis of Random Forest Classifier

The results indicate that last-mile connectivity and occupation were the most influential factors affecting passenger satisfaction. Station facilities, digital payment improvements, and route usage frequency also showed significant influence on commuter perception. In contrast, fare price perception and time management impact exhibited comparatively lower influence. These findings demonstrate that integrating sentiment features derived from the GRU model with structured survey attributes enhances the model's ability to capture passenger perceptions effectively and provides meaningful insights for improving metro services.

b) GRU Sentiment Analysis Performance

The GRU model was used to classify sentiment from passenger textual responses. Although the model achieved high accuracy, performance was affected by class imbalance.

The GRU model achieved:

Metric	Value
Accuracy	0.9133
Precision	0.8342
Recall	0.9133
F1-score	0.8720

Table.6. Classification Report of GRU

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

The confusion matrix indicates that the model is biased toward the majority class, resulting in lower performance in identifying minority class sentiments despite achieving high overall accuracy. This result indicates that deep learning models such as GRU require either larger balanced datasets or more diverse textual responses for effective sentiment classification.

Although the GRU model achieved higher accuracy, its performance was influenced by class imbalance in the dataset. In contrast, the Random Forest model demonstrated stable performance on structured data and provided interpretable insights through feature importance analysis. The hybrid approach effectively combines the strengths of both models, improving overall predictive capability. The Fig.7. predicts the sentiment distribution across gender and age groups.

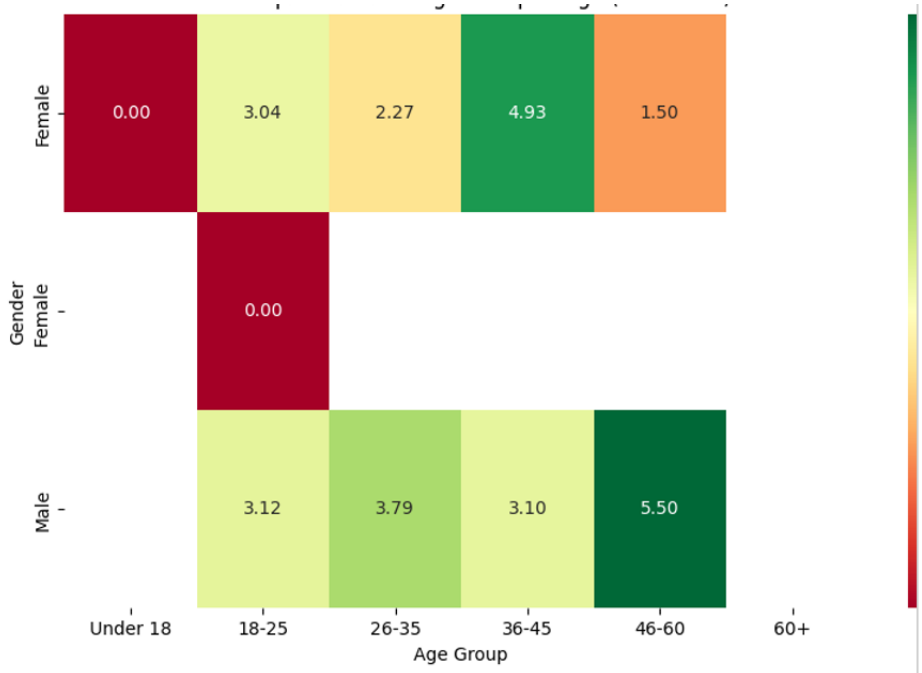

Fig.7. Sentiment distribution across gender and age groups.

5. CONCLUSION AND FUTURE SCOPE

This study analysed passenger satisfaction in the Pune Metro using statistical and machine learning techniques. The reliability of the survey data was validated using Cronbach's Alpha and McDonald's Omega.

The GRU model achieved high accuracy (91%) in sentiment analysis, while the Random Forest model (71%) effectively analysed structured data and identified key influencing factors such as digital payment systems, station facilities, and last-mile connectivity.

The study contributes to intelligent transportation systems by enabling data-driven decision-making. The hybrid modelling approach provides deeper insights into commuter behaviour and service quality.

6. FUTURE SCOPE

In the future, the dataset will be expanded to improve the accuracy, reliability, and generalization capability of the model. A larger and more diverse dataset will help address class imbalance and enhance the effectiveness of sentiment analysis.

Advanced deep learning models such as Transformers will be explored to improve the performance of textual sentiment analysis and capture complex language patterns more effectively. In addition, real-time metro operational data, including passenger flow and service updates, will be integrated to enable dynamic and real-time prediction of passenger satisfaction.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Time-series forecasting techniques will be applied to predict passenger demand and satisfaction trends for the next three years based on historical data. This will support long-term planning, resource allocation, and decision-making for metro authorities.

Furthermore, the hybrid modeling approach can be enhanced by incorporating additional features such as weather conditions, peak-hour patterns, and geographical factors. Predictive insights can also be delivered through commuter-facing applications, enabling personalized recommendations and improving overall user experience. These enhancements aim to support efficient metro operations, improve passenger satisfaction, and contribute to sustainable and intelligent urban mobility systems.

7. ACKNOWLEDGEMENT

We sincerely thank the management of **Nowrosjee Wadia College, Pune** for providing the facilities and support needed to carry out our research for the **Pune Metro Passenger Survey and Data Analysis**.We are very grateful to **Dr. Reena Bharati** for her guidance and support throughout the project.

8. REFERENCES

[1] S. García, J. Luengo, and F. Herrera, “Data Preprocessing in Data Mining,” Springer, 2016.
[2] T. Dasu and T. Johnson, “Exploratory Data Mining and Data Cleaning,” Wiley, 2003.
[3] V. Dhar, “Data Science and Prediction,” Communications of the ACM, vol. 56, no. 12, pp. 64–73, 2013.
[4] J. D. Kelleher and B. Tierney, “Data Science,” MIT Press, 2018.
[5] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, “Data Mining with Big Data,” IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107, 2014.
[6] J. L. Fleiss, B. Levin, and M. C. Paik, “Statistical Methods for Rates and Proportions,” Wiley, 2013.
[7] M. Friendly and H. Wainer, “A History of Data Visualization and Graphic Communication,” Harvard University Press, 2021.
[8] C. Stolper, A. Perer, and D. Gotz, “Progressive Visual Analytics: User-Driven Visual Exploration of In-Progress Analytics,” IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 1653–1662, 2014.
[9] R. Y. Wang and D. M. Strong, “Beyond Accuracy: What Data Quality Means to Data Consumers,” Journal of Management Information Systems, vol. 12, no. 4, pp. 5–33, 1996.
[10] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” arXiv preprint arXiv:1412.3555, 2014.
[11] A. Vaswani et al., “Attention Is All You Need,” Advances in Neural Information Processing Systems (NeurIPS), 2017.
[12] W. Revelle and R. E. Zinbarg, “Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma,” Psychometrika, vol. 74, no. 1, pp. 145–154, 2009.
[13] K. Cho et al., “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation,” Proceedings of EMNLP, 2014.
[14] B. Liu, “Sentiment Analysis and Opinion Mining,” Morgan & Claypool Publishers, 2012.
[15] D. Ren, “Understanding Statistical Hypothesis Testing,” Journal of Emergency Nursing, vol. 35, no. 1, pp. 57–59, 2009.
[16] S. Bandyopadhyay, S. Mallik, and A. Mukhopadhyay, “A Survey and Comparative Study of Statistical Tests for Identifying Differential Expression from Microarray Data,” IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 11, no. 1, pp. 95–115, 2014.
[17] J. Sreekumar and K. K. Jose, “Statistical Tests for Identification of Differentially Expressed Genes in cDNA Microarray Experiments,” Indian Journal of Biotechnology, vol. 7, pp. 423–436, 2008.
[18] H. Zhang et al., “Ensemble Learning Methods in Machine Learning: A Review,” IEEE Access, 2019.
[19] S. K. Sharma et al., “A Comparative Study of Machine Learning Algorithms for Prediction Tasks,” IEEE Access, 2019.
[20] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[21] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv preprint arXiv:1412.6980, 2015.