

Public Transport Delay Prediction Using Machine Learning Techniques

Y.NagaMalleswarao¹, K.Pavani², T.Swapna³

#1 Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada.

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

Abstract: Delays in public transit have a substantial impact on urban mobility, resulting in passenger annoyance and decreased system reliability. The majority of current delay estimate techniques rely on set timetables and past averages, which are unable to adjust to changing traffic circumstances. In order to successfully understand delay patterns, this research suggests a machine learning-based public transportation delay prediction system that makes use of historical and traffic-related data. For precise delay prediction, a Random Forest model is trained using the preprocessed data. Users can enter travel-related data and get information about anticipated delays as well as explanations. The suggested approach increases forecast accuracy, facilitates better trip planning, and raises the general effectiveness and dependability of public transportation systems by adjusting to shifting traffic patterns.

Index terms - — *Public Transport, Delay Prediction, Machine Learning, Traffic Data, Random Forest, Travel Time Estimation*

1. INTRODUCTION

Public transportation systems are essential for efficient urban mobility, offering cost-effective and environmentally friendly travel options while reducing traffic congestion and fuel consumption. Rapid urbanization and increasing population density have significantly increased the demand for reliable public transport services. However, one of the major challenges faced by these systems is the occurrence of unpredictable delays caused by traffic congestion, peak-hour demand, road conditions, weather variations, and operational inefficiencies. These delays negatively impact passenger satisfaction, increase travel uncertainty, and reduce the overall efficiency of transportation networks. Traditional public transport systems primarily rely on fixed schedules and historical averages to estimate arrival and departure times. While these approaches are simple to implement, they fail to adapt to dynamic real-time conditions such as sudden traffic congestion, accidents, or route changes. As a result, the delay information provided to passengers is often inaccurate and unreliable. With the advancement of

technologies such as GPS tracking, IoT devices, and smart city infrastructure, large volumes of real-time and historical transport data are now available. This creates an opportunity to apply data-driven techniques for more accurate and intelligent delay prediction.

Machine Learning (ML) has emerged as a powerful tool for analyzing complex and high-dimensional datasets. It can capture non-linear relationships between various factors such as speed, travel time, traffic density, route information, and temporal attributes. In this project, a Random Forest algorithm is used due to its ability to handle large datasets, reduce overfitting, and provide high prediction accuracy. The model is trained using preprocessed transport and traffic data to learn delay patterns effectively. Feature engineering techniques such as extraction of time-based attributes (hour and minute) further enhance the model's predictive capability.

In addition to prediction accuracy, interpretability is an important aspect of modern intelligent systems. Therefore, this project integrates Explainable Artificial Intelligence (XAI) techniques using SHAP (SHapley Additive exPlanations) to provide transparency in model predictions. SHAP values help identify the contribution of each feature, such as speed or travel time, in determining the predicted delay. This improves user trust and allows passengers as well as transport authorities to understand the reasons behind delays.

The proposed system is implemented as a web-based application that includes modules for data administration, model training, and prediction. Users can input travel-related parameters and receive delay predictions along with explanatory insights. By

combining machine learning with real-time data analysis and explainability, the system enhances travel planning, reduces uncertainty, and improves the reliability and efficiency of public transportation services.

2. LITERATURE SURVEY

a) Real-Time Transit Delay Prediction Using Gps Data:

By examining both historical and current transit data gathered from metropolitan transportation systems, this study aims to forecast transit delays. The authors stress how crucial it is to manage massive transportation information and identify intricate delay patterns using machine learning approaches. The article examines many predictive models and emphasizes how ensemble learning techniques might increase prediction accuracy. The suggested approach offers accurate delay forecasts and constantly adjusts to shifting traffic circumstances. The findings show that data-driven prediction models greatly outperform conventional schedule-based systems and contribute to increased operational effectiveness and passenger satisfaction.

b) MACHINE LEARNING FOR URBAN TRAFFIC ANALYTICS:

An extensive examination of machine learning applications in urban traffic analytics is presented in this research. The authors look into a number of traffic-related issues, including delay prediction, trip time estimation, and congestion detection. Supervised learning techniques are used to preprocess and analyze large amounts of traffic data. The study demonstrates how non-linear correlations between traffic characteristics and delays may be found using machine learning algorithms. According to

experimental findings, sophisticated models outperform conventional statistical techniques in terms of accuracy and resilience. The study comes to the conclusion that developing intelligent transportation systems requires machine learning.

c) PUBLIC TRANSPORT DELAY FORECASTING USING RANDOM FOREST:

A Random Forest-based method for predicting delays in public transportation systems is proposed in this article. To train the model and examine delay behavior, the authors use past transportation and traffic-related data. Through ensemble learning, the Random Forest method minimizes overfitting and efficiently manages high-dimensional data. The study shows that, in a variety of traffic scenarios, the suggested model achieves good forecast accuracy. The findings demonstrate that Random Forest's stability, resilience, and capacity to handle complicated datasets make it an excellent choice for delay prediction problems.

d) SMART CITY TRANSPORT SYSTEMS USING AI

The incorporation of artificial intelligence into smart city transportation systems is examined in this research. The authors go into how AI-driven models might enhance delay prediction, public transportation dependability, and traffic management. The report emphasizes how data analytics may be used to improve transportation operations and lessen traffic. The usefulness of intelligent systems in urban settings is illustrated by looking at a number of case examples. The study highlights how AI might improve decision-making procedures and raise the general effectiveness of public transportation systems.

e) TRAFFIC CONGESTION ANALYSIS USING GPS TRAJECTORIES:

This study uses extensive trajectory data gathered from metropolitan transport networks to investigate patterns of traffic congestion. To find congestion hotspots and comprehend traffic behavior, the authors use data mining and machine learning approaches. The study demonstrates how road network performance may be understood and trip delays can be estimated using trajectory-based analysis. The experiment's findings show that data-driven congestion analysis increases the precision of traffic forecasts. The results are in favor of using sophisticated analytics for efficient traffic management and monitoring.

3. METHODOLOGY

i) Proposed Work:

By efficiently leveraging historical and traffic-related data, the suggested approach presents an intelligent machine learning-based framework for forecasting delays in public transportation. The system continually learns delay patterns from real-world data gathered from GPS records, traffic conditions, and temporal elements like peak and off-peak hours, in contrast to conventional schedule-based techniques. To guarantee excellent data quality, the gathered data is preprocessed using feature encoding, cleaning, and normalization. The processed information is used to train a Random Forest machine learning model, which enables precise and reliable delay prediction by capturing intricate and non-linear correlations between traffic factors and delay occurrences.

Users engage with the system during the operating phase by entering trip-related input parameters, such

as route information, journey time, and traffic conditions. After analyzing these inputs, the trained model produces anticipated delay values and explanatory insights that emphasize the main variables affecting the forecast. Transparency and user confidence in the system are enhanced by this explanation feature. The suggested method improves trip planning, lowers passenger uncertainty, and raises the general dependability and effectiveness of public transportation systems by adjusting to shifting traffic patterns and learning from past trends.

ii) System Architecture:

The suggested system architecture is a web-based intelligent framework that combines prediction, machine learning, and data management modules into a single workflow. A Front-End Web System, which acts as the primary layer of user interface, is where the architecture starts. Authorized administrators may safely log in, upload datasets pertaining to traffic and public transportation, and examine registered user information using the Data Administration module. To ensure restricted access and data integrity, these datasets are saved and sent to the Machine Learning Model module. While preserving system security and scalability, this modular architecture makes dataset maintenance easier.

Preprocessing the provided data and utilizing the Random Forest technique to create the prediction model are the responsibilities of the Traffic Analytics module. The Prediction Engine allows registered users to log in and add real-time or test trip data once the trained model has been preprocessed. In order to forecast transportation delays and produce explanations that emphasize significant variables like traffic density and trip duration, the system runs the input through the trained model. This end-to-end

design guarantees smooth data flow from data intake to prediction, facilitates real-time decision-making, and improves the public transportation delay prediction system's transparency, dependability, and usability.

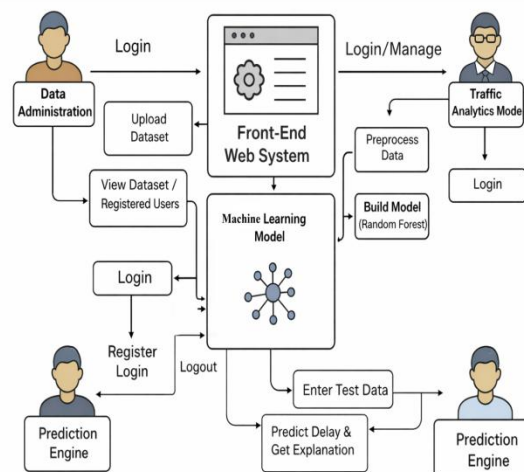


Fig1 proposed architecture

iii) Modules:

a) Data Administration Module

This module manages system access and dataset handling. It allows authorized administrators to log in securely, upload public transport and traffic-related datasets, and view registered user information. It ensures data integrity, controlled access, and proper management of system resources.

b) Traffic Analytics Module

This module is responsible for data preprocessing and model development. It performs operations such as cleaning missing values, feature encoding, and normalization before training the Random Forest machine learning model. The trained model learns

delay patterns from historical and traffic data and is prepared for prediction tasks.

c) **Machine Learning Model Module**

This module acts as the core analytical component of the system. It stores the trained Random Forest model and handles model evaluation and validation. The module ensures accurate and reliable delay prediction by analyzing learned patterns and updating the model when new data is available.

d) **Prediction Engine Module**

This module enables end users to interact with the system for delay prediction. Users can register, log in, and enter travel-related input data. The system processes the input through the trained model to predict delays and provide explanations, helping users understand the influencing factors behind the prediction.

iv) **Algorithms:**

a. **Random Forest Algorithm**

Random Forest is the primary machine learning algorithm used in the proposed system for public transport delay prediction. It is an ensemble learning technique that constructs multiple decision trees using different subsets of training data and features, and combines their outputs to produce accurate predictions. This approach reduces overfitting and improves generalization performance. In this system, the Random Forest model is trained using historical public transport and traffic-related data to learn complex and non-linear delay patterns. During prediction, the algorithm analyzes input parameters such as traffic density, route information, and time of

travel to estimate delays reliably and provide robust prediction results.

4. **EXPERIMENTAL RESULTS**

Real-world traffic and public transportation datasets supplied via the Data Administration module were used for the experimental assessment of the suggested public transportation delay prediction system. The screenshots show how to successfully upload datasets in CSV format, perform preprocessing, and integrate models into the web-based framework. In order to properly learn delay patterns, the Random Forest model was trained using historical transit records and traffic-related variables. The system interface verifies seamless module interactions, such as administrator login, dataset administration, and user registration, confirming the functional behavior of the established design.

The trained model produced delay estimates together with explanatory insights when users entered travel-related input variables through the Prediction Engine during testing. The findings demonstrate that by taking temporal and traffic variables into account, the algorithm generates accurate and consistent delay estimations. The suggested solution showed better prediction reliability and user transparency when compared to conventional static schedule-based methods. The machine learning-based approach's efficacy in improving public transportation delay prediction and assisting with well-informed trip planning is validated by the successful completion of all modules and precise prediction outputs.

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$\text{Accuracy} = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

$$\text{Precision} = \frac{\text{True positives}}{(\text{True positives} + \text{False positives})} = \frac{TP}{(TP + FP)}$$

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

$$\text{Recall} = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k$$

AP_k = the AP of class k
 n = the number of classes

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(\text{Recall} \cdot \text{Precision})}{(\text{Recall} + \text{Precision})}$$

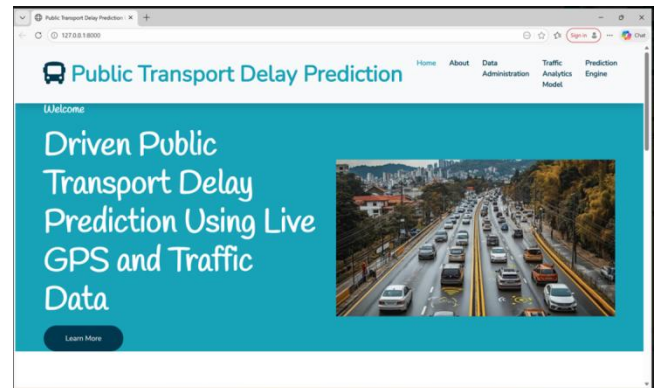


Figure 1 illustrates the Home Page of the System

This figure shows the home page of the Public Transport Delay Prediction web application. It presents the project title and highlights the objective of predicting public transport delays using live GPS and traffic data. The navigation bar provides access to different modules such as Data Administration, Traffic Analytics Model, and Prediction Engine, enabling smooth user interaction.

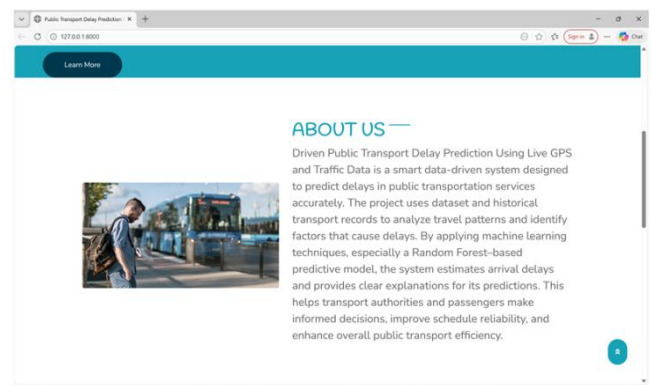


Figure 2 illustrates the About Us Page

This figure describes the purpose and functionality of the proposed system. It explains how historical transport records and traffic data are analyzed using machine learning techniques, particularly the Random Forest model, to predict delays accurately. The page emphasizes the system's role in improving decision-making for passengers and transport authorities.

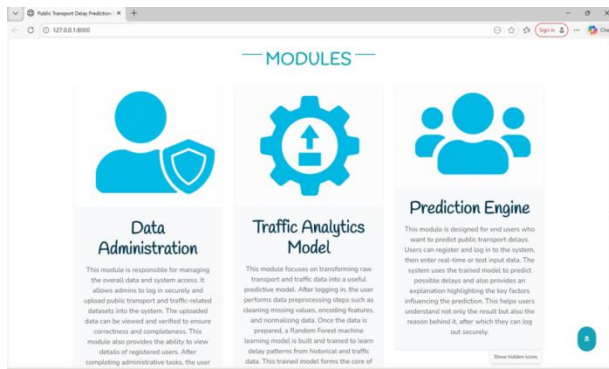


Figure 3 illustrates the Modules Overview Page

This figure displays the major functional modules of the system, namely Data Administration, Traffic Analytics Model, and Prediction Engine. Each module is visually represented with a brief description, providing users with an overall understanding of the system architecture and workflow.

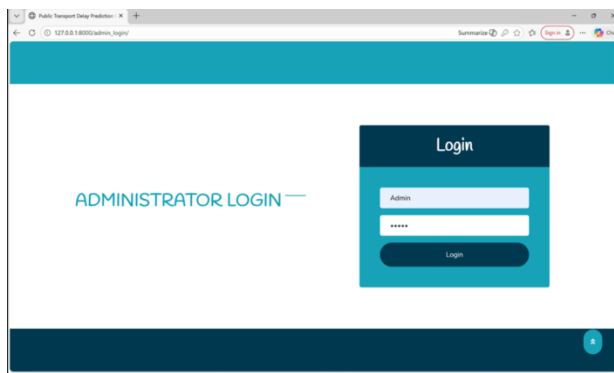


Figure 4 illustrates the Administrator Login Interface

This figure shows the secure login page for the administrator. Authorized administrators can access the system using valid credentials to manage datasets and registered users. This module ensures secure access control and system integrity.

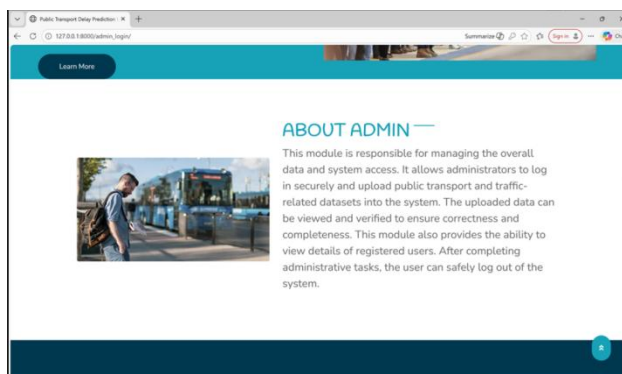


Figure 5 illustrates the About Admin Page

This figure explains the responsibilities of the Data Administration module. It highlights the administrator's role in uploading transport and traffic-related datasets, verifying data correctness, managing users, and securely logging out after completing administrative tasks.

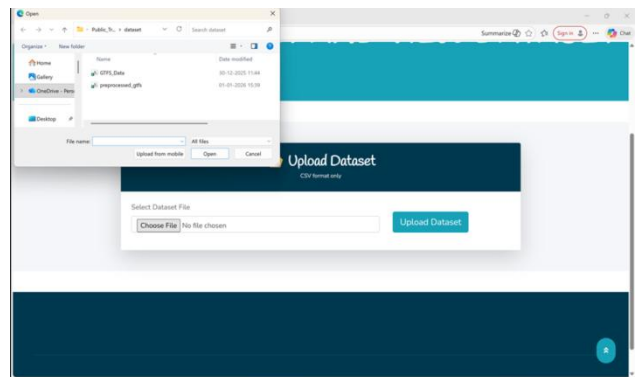


Figure 6 illustrates the Dataset Upload Interface

This figure shows the dataset upload functionality of the system. The administrator can upload CSV files containing public transport and traffic data. This interface ensures that only valid datasets are accepted for further processing.

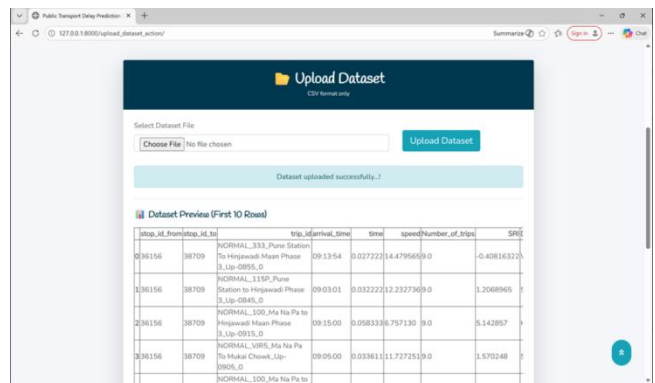


Figure 7 illustrates the Dataset Upload Confirmation and Preview

This figure displays a successful dataset upload message along with a preview of the first few rows of the uploaded dataset. It allows administrators to verify the correctness and structure of the data before it is used for preprocessing and model training.

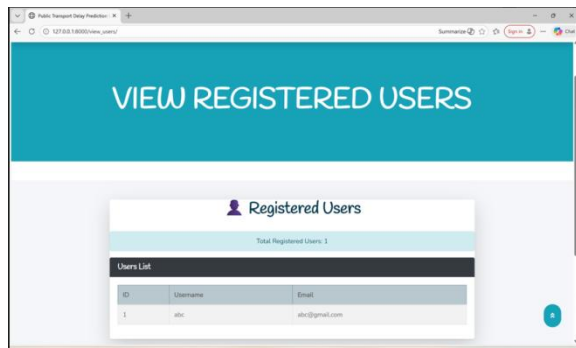


Figure 8 illustrates the View Registered Users Page

This figure shows the list of registered users in the system along with details such as user ID, username, and email address. It enables administrators to monitor system usage and manage user accounts effectively.

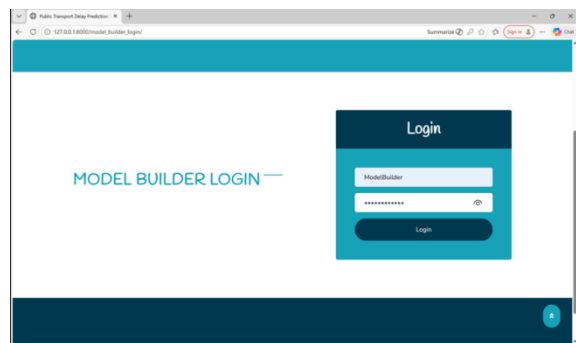


Figure 9 illustrates the Model Builder Login Interface

This figure represents the login page for the Traffic Analytics or Model Builder role. Authorized users can log in to preprocess data and build the machine learning model. This separation of roles improves system security and modularity.

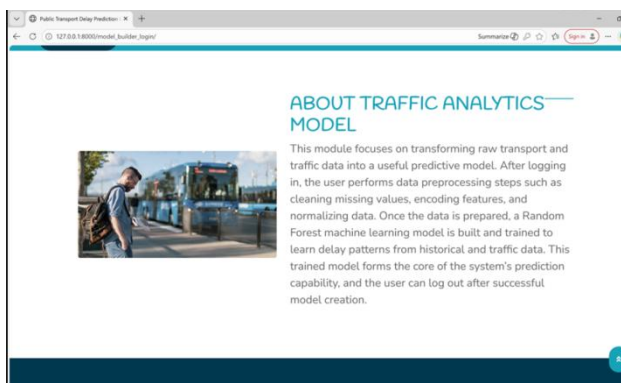


Figure 10 illustrates the Traffic Analytics Model Page

This figure explains the Traffic Analytics Model module, where data preprocessing steps such as cleaning missing values, encoding features, and normalization are performed. After preprocessing, the Random Forest machine learning model is trained using historical and traffic data. This trained model forms the core of the system's delay prediction capability.

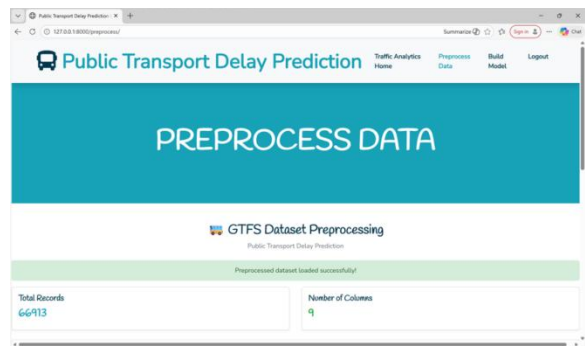


Figure 11 illustrates the Data Preprocessing Interface

This figure shows the preprocessing module of the Traffic Analytics Model. It confirms successful loading of the preprocessed GTFS dataset and displays summary statistics such as total number of records and number of columns. This step ensures that raw transport data is cleaned, structured, and ready for model training.

Walt Disney World: Key Performance Indicators

107.541.61001@preprocessed

Summary

Sign in

100%

Dataset Columns

stop_id

stop_name

stop_id

stop_name

trip_id

time

speed

number_of_stops

degree_of_congestion

hour

minute

Preprocessed Data Preview (First 10 Rows)

	stop_id	stop_name	stop_id	stop_name	trip_id	time	speed	number_of_stops	degree_of_congestion	hour	minute
0	36156	36709	3732	0.037222	14.479585	9.0	3	9.0	13.0		
1	36156	36709	355	0.032222	12.232736	9.0	2	9.0	3.0		
2	36156	36709	23	0.058333	6.757130	9.0	0	9.0	15.0		
3	36156	36709	5293	0.038311	11.727281	9.0	2	9.0	5.0		
4	36156	36709	22	0.058333	6.757130	9.0	0	9.0	0.0		
5	36156	36709	2238	0.040556	9.712050	9.0	1	9.0	11.0		
6	36156	36709	4242	0.037222	10.989532	9.0	2	9.0	0.0		
7	36156	36709	4243	0.037222	10.989532	9.0	2	9.0	10.0		
8	36156	36709	2128	0.043889	8.980066	9.0	1	9.0	14.0		
9	36709	2321	5051	0.023889	24.774036	22.0	1	9.0	16.0		

Figure 12 illustrates the Preprocessed Dataset Preview

This figure presents a preview of the first few rows of the preprocessed dataset along with selected feature columns such as stop IDs, trip ID, speed, congestion degree, hour, and minute. It allows verification of data correctness after preprocessing and feature engineering.

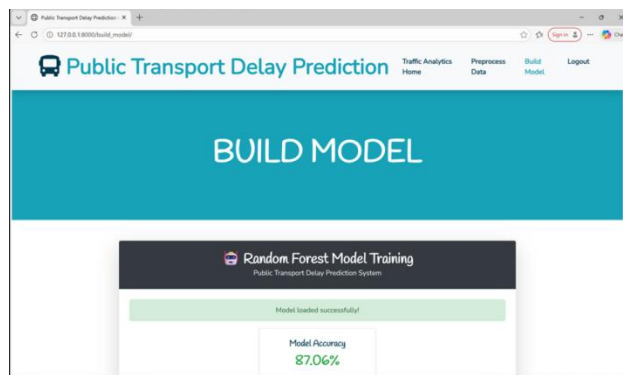


Figure 13 illustrates the Model Building Interface

This figure shows the Random Forest model training page. After preprocessing, the model is built using the prepared dataset, and a successful model loading message is displayed. The interface also highlights the achieved model accuracy, confirming effective learning of delay patterns.

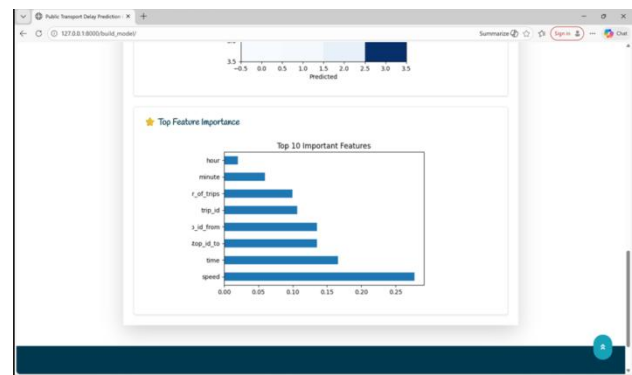


Figure 15 illustrates Feature Importance Analysis

This figure displays the top important features identified by the Random Forest model. Features such as speed, travel time, stop IDs, number of trips, and temporal attributes contribute significantly to prediction decisions. This analysis improves interpretability and validates the relevance of selected features.

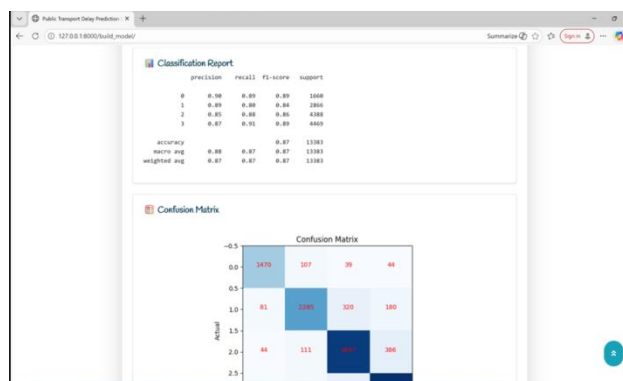


Figure 14 illustrates the Classification Report and Confusion Matrix

This figure presents the evaluation results of the trained Random Forest model. It includes precision, recall, F1-score, and support values for each class, along with a confusion matrix visualization. These metrics demonstrate the reliability and classification performance of the proposed delay prediction model.

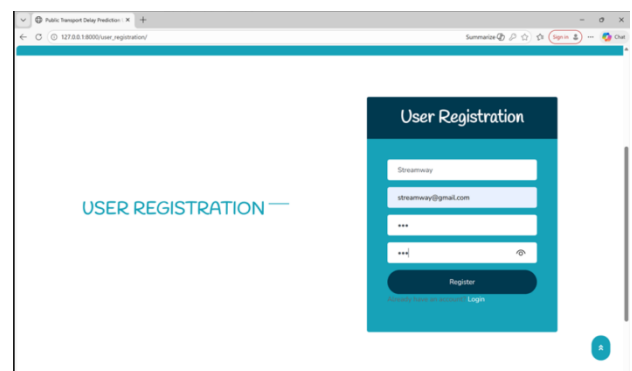


Figure 16 illustrates the User Registration Page

This figure shows the user registration interface, where new users can create an account by providing basic details such as username, email, and password. This module enables secure access for end users who wish to predict public transport delays.

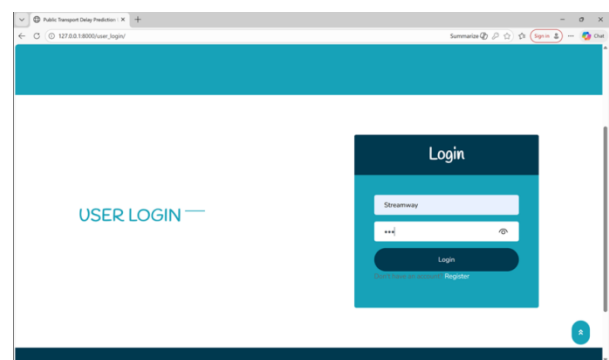


Figure 17 illustrates the User Login Interface

This figure represents the login page for registered users. After authentication, users gain access to the prediction module, ensuring secure and controlled system usage.

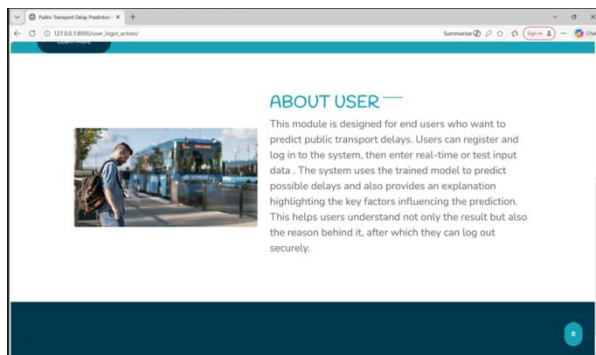


Figure 18 illustrates the About User Page

This figure explains the functionality available to end users. It describes how users can input travel data, obtain delay predictions, and receive explanations for the predicted results, thereby improving transparency and user understanding.

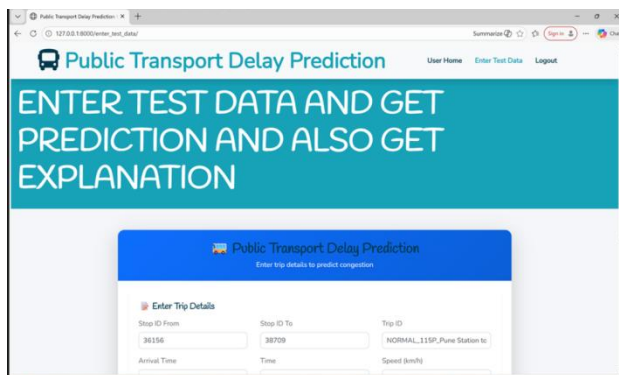


Figure 19 illustrates the Test Data Input Interface

This figure shows the interface where users enter test or real-time trip details such as stop IDs, trip ID, speed, and time parameters. These inputs are forwarded to the trained model for delay prediction.

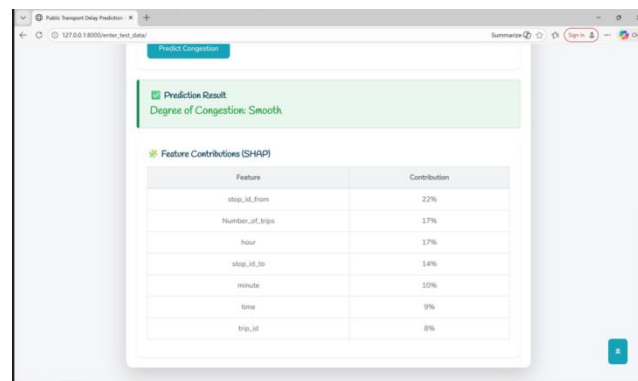


Figure 20 illustrates the Prediction Result with Explanation

This figure presents the final prediction output, displaying the predicted degree of congestion or delay status. Additionally, SHAP-based feature contribution values are shown, explaining how each feature influenced the prediction. This enhances trust and interpretability of the machine learning model.

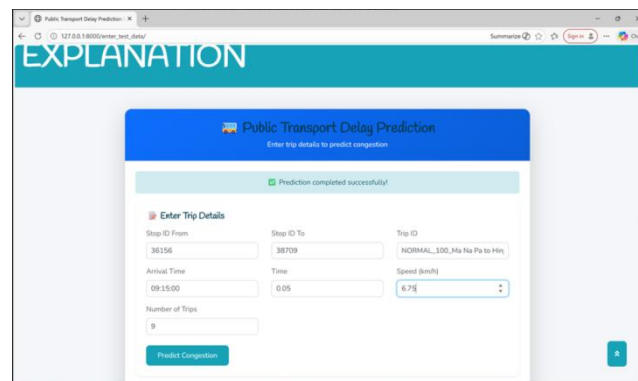


Figure 21 illustrates the User Input Interface for Explanation

This figure shows the explanation module where users enter trip-related details such as stop IDs, arrival time, travel time, speed, and number of trips. After submitting the inputs, the system confirms successful prediction initiation. This interface serves as the entry point for generating delay predictions along with explanations.

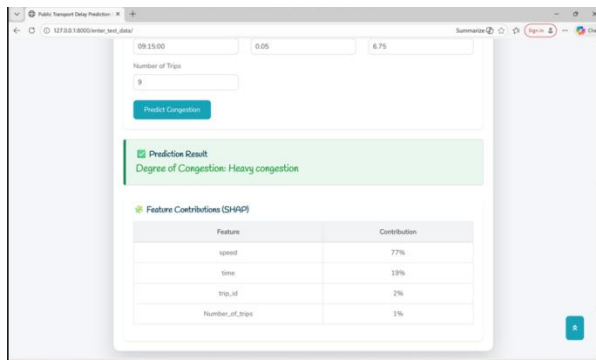


Figure 22 illustrates Heavy Congestion Prediction with SHAP Explanation

This figure presents a prediction result indicating **heavy congestion** for the given trip inputs. The SHAP-based feature contribution table shows that vehicle speed contributes the most (77%) to the prediction, followed by travel time (19%). This result demonstrates how low speed and increased travel time strongly influence congestion prediction.

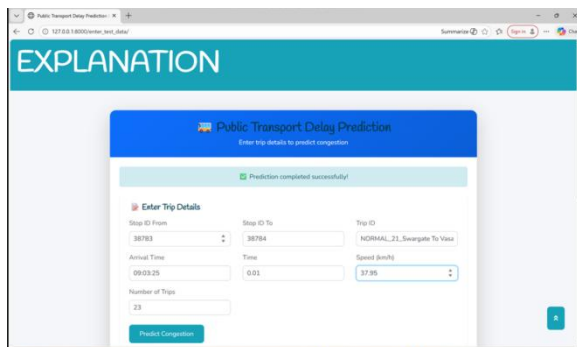


Figure 23 illustrates Another Test Case Input Scenario

This figure displays a different set of trip inputs entered by the user, including higher speed and different route parameters. It demonstrates the system's ability to handle multiple test cases dynamically and generate predictions for varying traffic conditions.

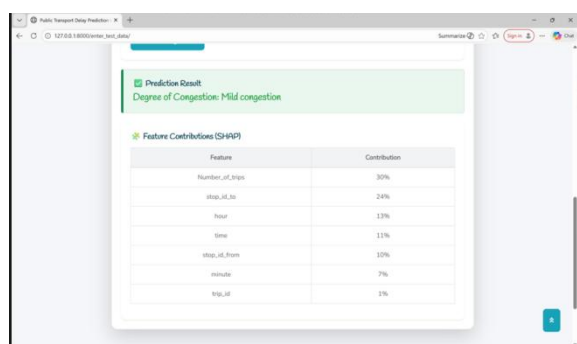


Figure 24 illustrates Mild Congestion Prediction with Feature Contributions

This figure shows a **mild congestion** prediction outcome. The SHAP explanation indicates that the number of trips and destination stop ID have significant influence, while temporal features such as hour and time have moderate impact. This confirms that congestion levels vary based on route load and time-related factors.

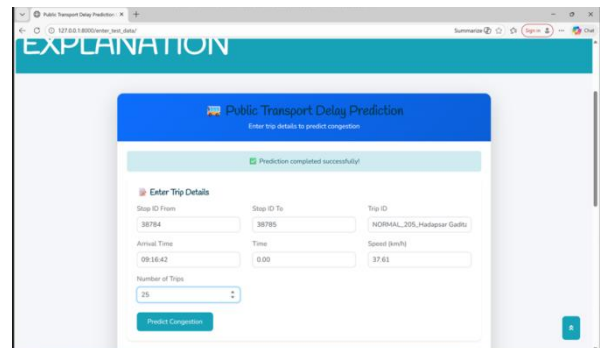


Figure 25 illustrates a New Test Input with Updated Parameters

This figure presents another test input scenario with higher speed and increased number of trips. It highlights how the system accepts real-time-like inputs and processes them consistently through the trained Random Forest model.

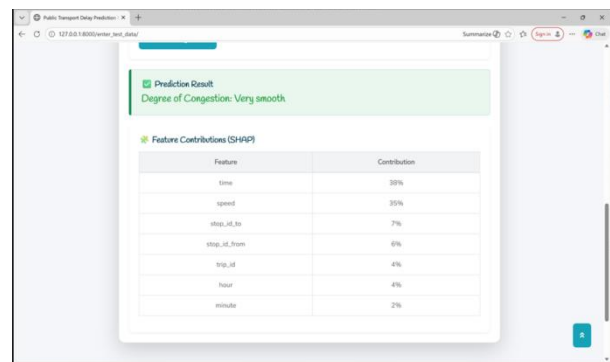


Figure 26 illustrates Very Smooth Traffic Prediction with Explanation

This figure shows a prediction result indicating **very smooth traffic conditions**. The SHAP feature contributions reveal that travel time (38%) and speed (35%) are the dominant factors leading to this outcome. Minimal contribution from route and temporal features indicates free-flow traffic conditions.

5. CONCLUSION

In order to overcome the shortcomings of conventional schedule-based methods, this study introduced a machine learning-based public transportation delay prediction system that efficiently makes use of past transportation records and traffic-related data. The system precisely learns delay patterns under various traffic situations by combining feature engineering, data preprocessing, and a Random Forest model inside a web-based platform. The experimental findings show consistent system execution across all functional units and dependable prediction performance.

Additionally, by identifying important variables impacting delay forecasts, such as speed, trip duration, and route features, explainable AI employing SHAP promotes transparency. Both passengers and transportation authorities may make well-informed decisions because to this explanation capacity, which also increases user confidence. All things considered, the suggested approach enhances trip planning, lowers uncertainty, and advances the creation of intelligent and dependable public transportation systems.

6. FUTURE SCOPE

The proposed public transport delay prediction system can be further enhanced by integrating additional real-time data sources such as weather conditions, road construction updates, accidents, and live traffic feeds. Incorporating these factors will improve the model's ability to handle dynamic and unforeseen situations, thereby increasing prediction accuracy. The system can also be extended by exploring advanced deep learning models such as LSTM, CNN, or hybrid approaches to better capture

temporal and spatial dependencies in transportation data.

In addition, the system can be deployed as a mobile application or integrated into existing smart city platforms to provide real-time delay notifications to passengers. Future improvements may include multi-city scalability, integration with GPS-enabled IoT devices for live tracking, and continuous model updating using streaming data. Enhancing the explainable AI component with visual dashboards and user-friendly insights can further improve usability. Overall, these advancements will make the system more robust, scalable, and suitable for real-world intelligent transportation systems.

REFERENCES

- [1] K. Gkiotsalitis and O. Cats, "Public transport planning adaption under the COVID-19 pandemic crisis: literature review of research needs and directions," *Transp. Rev.*, vol. 41, no. 3, pp. 374–392, May 2021, doi: 10.1080/01441647.2020.1857886.
- [2] X. Y. Yan, C. Zhao, Y. Fan, Z. Di, and W. X. Wang, "Universal predictability of mobility patterns in cities," *J. R. Soc. Interface*, vol. 11, no. 100, Nov. 2014, doi: 10.1098/RSIF.2014.0834.
- [3] R. Hamdi et al., "The State-of-the-Art of Urban Climate Change Modeling and Observations," *Earth Syst. Environ.* 2020 44, vol. 4, no. 4, pp. 631–646, Nov. 2020, doi: 10.1007/S41748-020-00193-3.
- [4] H. Nugroho, "A Review: Data Quality Problem in Predictive Analytics," *IJAIT (International J. Appl. Inf. Technol.)*, vol. 7, no. 02, p. 79, Aug. 2023, doi: 10.25124/ijait.v7i02.5980.
- [5] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 45, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

- [6] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 27, 2015, doi: 10.1038/nature14539.
- [7] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.
- [8] H. Wicaksono, M. U. Nisa, and A. Vijaya, "Towards Intelligent and Trustable Digital Twin Asset Management Platform for Transportation Infrastructure Management Using Knowledge Graph and Explainable Artificial Intelligence (XAI)," 2023 IEEE Int. Conf. Ind. Eng. Eng. Manag. IEEM 2023, pp. 528–532, 2023, doi: 10.1109/IEEM58616.2023.10406401.
- [9] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003, doi: 10.5555/944919.944968.
- [10] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001, doi: 10.1214/aos/1013203451.
- [11] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
- [12] N. S. Altman, "An introduction to kernel and nearest neighbor nonparametric regression," *Am. Stat.*, vol. 46, no. 3, pp. 175–185, 1992, doi: 10.1080/00031305.1992.10475879.
- [13] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, no. 1, pp. 79–82, Dec. 2005, doi: 10.3354/cr030079.
- [14] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: 10.5194/gmd-7-1247-2014.
- [15] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/j.ijforecast.2006.03.001.
- [16] Kormáksson, Matthías, Luciano Barbosa, Marcos R. Vieira, and Bianca Zadrozny. "Bus travel time predictions using additive models." In 2014 IEEE international conference on data mining, pp. 875–880. IEEE, 2014. <https://doi.org/10.1109/ICDM.2014.107>
- [17] Kulla, Elis, Soushi Morita, Kengo Katayama, and Leonard Barolli. "Route lifetime prediction method in VANET by using AODV routing protocol (AODV-LP)." In *Complex, Intelligent, and Software Intensive Systems: Proceedings of the 12th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2018)*, Springer International Publishing, p. 3–11. 2019. https://doi.org/10.1007/978-3-319-93659-8_1
- [18] Rajput, Prashant, Durga Toshniwal, and Apeksha Aggarwal. "Improving infrastructure for transportation systems using clustering." In *Big Data Analytics: 5th International Conference, BDA 2017, Hyderabad, India, Proceedings 5*, Springer International Publishing, p. 129–143. 2017. https://doi.org/10.1007/978-3-319-72413-3_9
- [19] Chien, Steven I-Jy, and Chandra Mouly Kuchipudi. "Dynamic travel time prediction with real-time and historic data." *Journal of transportation engineering* 129, no. 6 (2003): 608–616. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:6\(608\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(608))

- [20] Vanajakshi, Lelitha, Shankar C. Subramanian, and R. Sivanandan. "Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses." *IET intelligent transport systems* 3, no. 1 (2009): 1-9. <https://doi.org/10.1049/iet-its:20080013>
- [21] Williams, Billy M., and Lester A. Hoel. "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results." *Journal of transportation engineering* 129, no. 6 (2003): 664-672. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:6\(664\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664))
- [22] Shalaby, Amer, and Ali Farhan. "Prediction model of bus arrival and departure times using AVL and APC data." *Journal of Public Transportation* 7, no. 1 (2004): 41-61. <https://doi.org/10.5038/2375-0901.7.1.3>
- [23] Lin, Yongjie, Xianfeng Yang, Nan Zou, and Lei Jia. "Real-time bus arrival time prediction: Case study for Jinan, China." *Journal of Transportation Engineering* 139, no. 11 (2013): 1133-1140. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000589](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000589)
- [24] Khan, Mahnoor, Nadeem Javaid, Muhammad Nabeel Iqbal, Muhammad Bilal, Syed Farhan Ali Zaidi, and Rashid Ali Raza. "Load prediction based on multivariate time series forecasting for energy consumption and behavioral analytics." In *Conference on Complex, Intelligent, and Software Intensive Systems*, Cham: Springer International Publishing, p. 305-316. 2018. https://doi.org/10.1007/978-3-319-93659-8_27
- [25] Majumdar, Sharmila, Moez M. Subhani, Benjamin Roullier, Ashiq Anjum, and Rongbo Zhu. "Congestion prediction for smart sustainable cities using IoT and machine learning approaches." *Sustainable Cities and Society* 64 (2021): 102500. <https://doi.org/10.1016/j.scs.2020.102500>
- [26] He, Peilan, Guiyuan Jiang, Siew-Kei Lam, and Yidan Sun. "Learning heterogeneous traffic patterns for travel time prediction bus journeys." *Information Sciences* 512 (2020): 1394-1406. <https://doi.org/10.1016/j.ins.2019.10.073>
- [27] Shoman, Maged, Armstrong Aboah, and Yaw Adu-Gyamfi. "Deep learning framework for predicting bus delays on multiple routes using heterogeneous datasets." *Journal of Big Data Analytics in Transportation* 2 (2020): 275-290. <https://doi.org/10.1007/s42421-020-00031-y>
- [28] Wu, Canrong, Yang Liu, Yueying Yang, Peng Zhang, Wu Zhong, Yali Wang, Qiqi Wang et al. "Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods." *Acta Pharmaceutica Sinica B* 10, no. 5 (2020): 766-788. <https://doi.org/10.1016/j.apsb.2020.02.008>
- [29] Zhang, Dalin, Yunjuan Peng, Yumei Zhang, Daohua Wu, Hongwei Wang, and Hailong Zhang. "Train time delay prediction for high-speed train dispatching based on spatio-temporal graph convolutional network." *IEEE Transactions on Intelligent Transportation Systems* 23, no. 3 (2021): 2434-2444. <https://doi.org/10.1109/TITS.2021.3097064>
- [30] Yu, Bin, Zhen Guo, Sobhan Asian, Huaizhu Wang, and Gang Chen. "Flight delay prediction for commercial air transport: A deep learning approach." *Transportation Research Part E: Logistics and Transportation Review* 125 (2019): 203-221. <https://doi.org/10.1016/j.tre.2019.03.013>

Author Profiles



Mr. Y. Naga Malleswarao Completed his Masters of Technology from JNTUK, MSC(IS) from ANU, BCA from ANU and MTech from JNTUK. He has System Administrator, Networking Administrator and Oracle Administrator. He also a web developer and python developer, Currently working has an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His area of interest include Artificial Intelligence and Machine Learning.



Mrs. K. Pavani Working as Assistant & Head of Department of MCA, in SRK Institute of technology in Vijayawada. She done with MCA, M. Tech in Computer Science. She has 10 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes AI ML, etc



Ms. T. Swapna is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc.(computer science) from P.B Siddhartha college of arts and science vijayawada. Her area of interest are Machine Learning with Python.