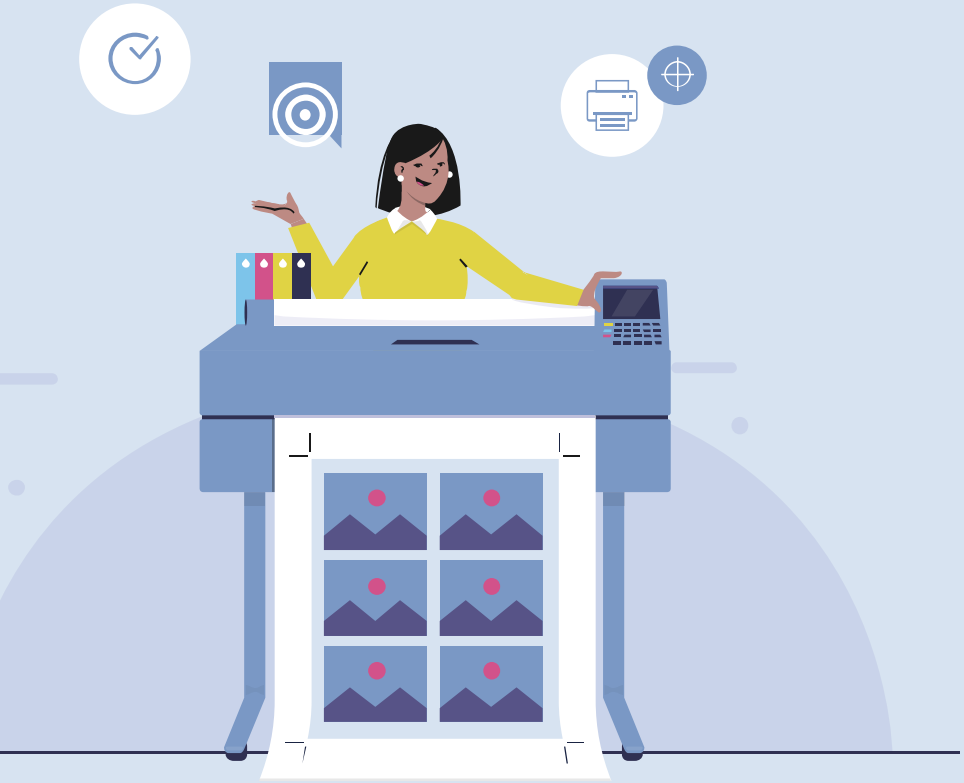
Spatio-temporal Forecasting of Bus Arrival Times to Downstream Stops

Group 07

- 210172N - Anushanga Sharada
- 210173T - Deshitha Gallage
- 210189X - Helith Gothatuwa



01

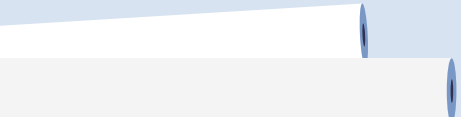
Introduction

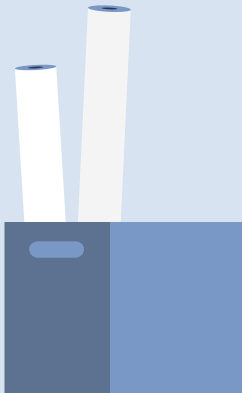
Project Overview

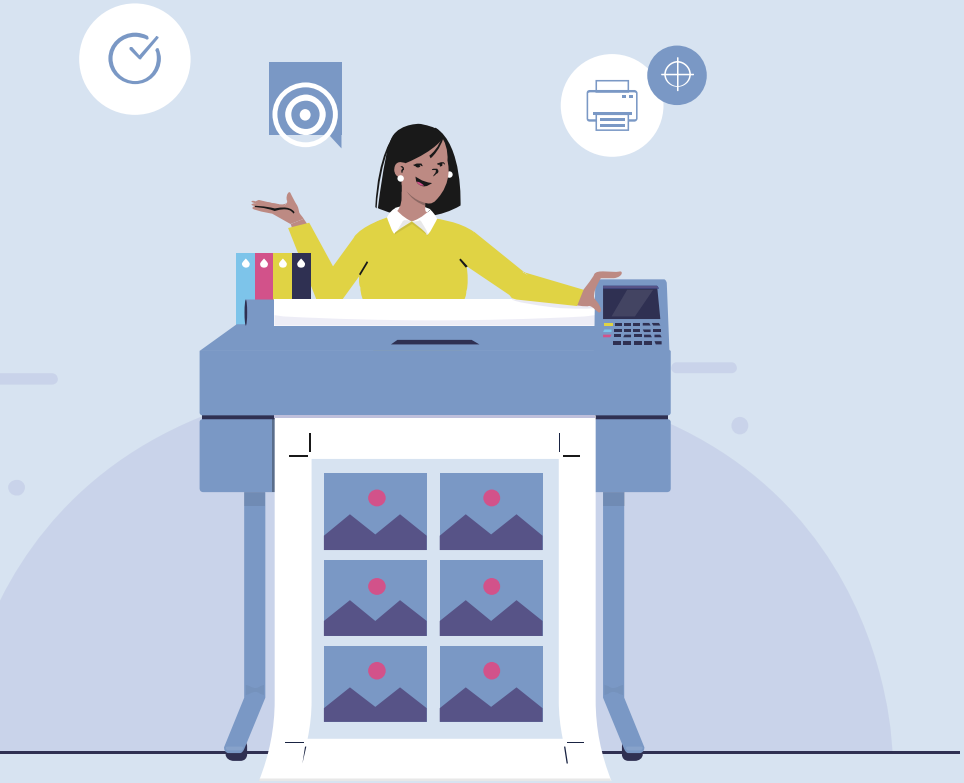
This project aims to predict bus arrival times using **spatio-temporal** data from NYC MTA buses, enhancing public transportation reliability and user satisfaction by reducing wait times and improving schedule adherence.

Scope of the Project

- Predicting bus arrival times depends on spatial and temporal patterns among stops and transport networks.
- Using state-of-the-art spatio-temporal forecast algorithms improves prediction accuracy.
- The project integrates these algorithms to provide real-time, precise bus arrival information.
- A comparative study will identify the most effective algorithm.







02

Dataset Selection

Dataset

NYC MTA

We used the NYC MTA buses data stream service, which provides bus location, route, and stop data at 10-minute intervals, along with scheduled arrival times for accurate prediction and analysis.

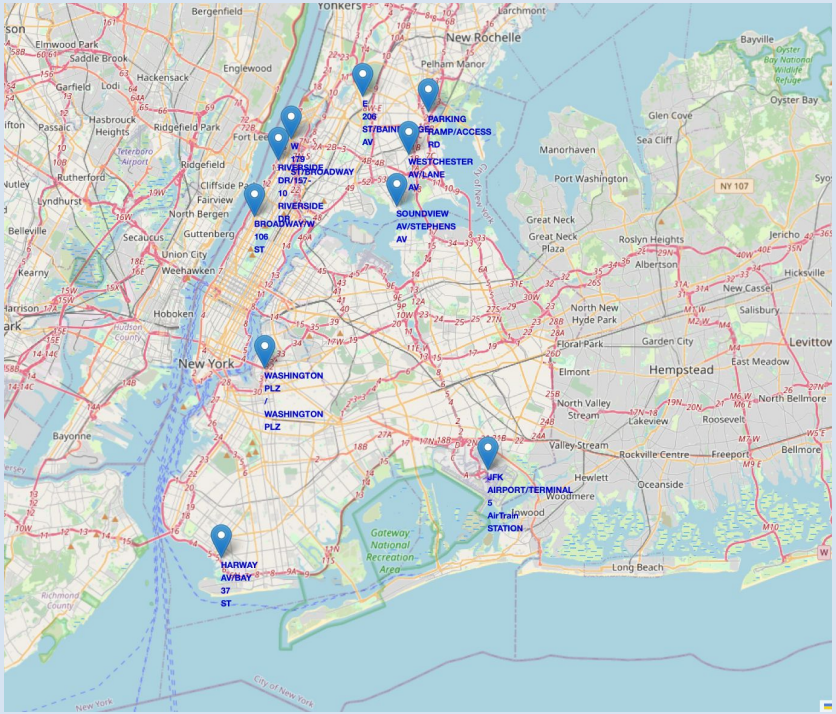
Why not local data

Because of the limitations of not having enough bus data with **multiple routes and crossroads** in local datasets.

Dataset Insights

Bus stops with the highest delays were identified by measuring the difference between scheduled and expected arrival times.

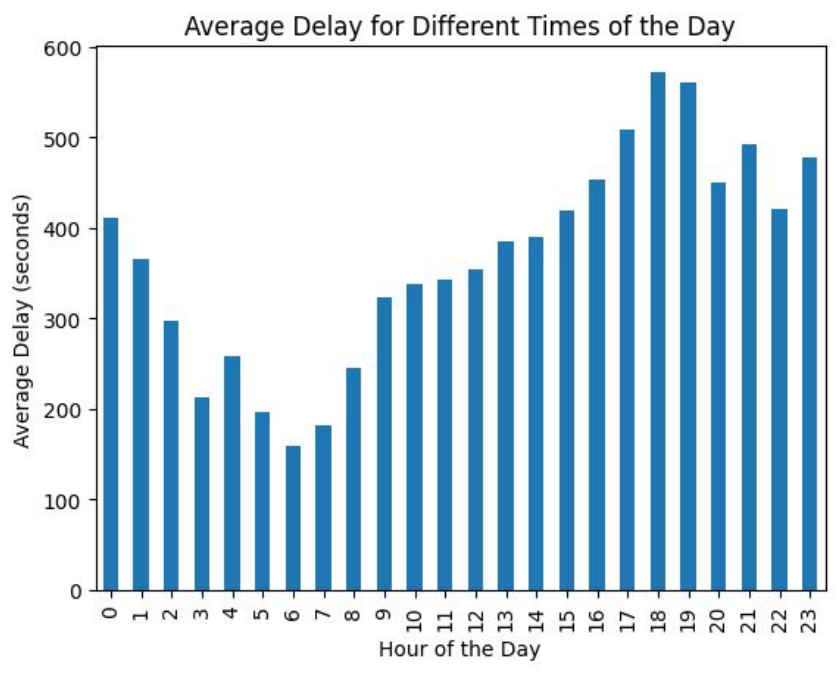
These stops reveal critical points for potential transit improvements.



Dataset Insights (cont.)

Average Delays by Time of Day:

Evening delays are notably higher, likely due to increased traffic as people travel from work to home.



Dataset Preprocessing

- **Handling Missing Values:**

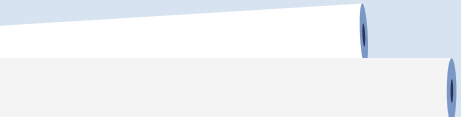
  Columns with missing values were removed from the dataset due to the abundance of available data. This approach allowed us to maintain data integrity without significantly impacting the analysis.

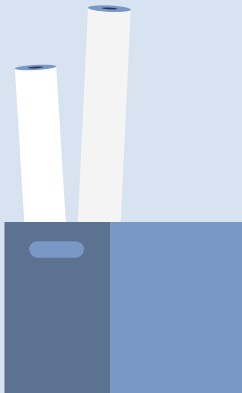
- **Date and Time Handling:**

  We addressed inconsistencies in the `ScheduledArrivalTime`, such as handling values like `24:` hours by converting them to `00:` and adjusting the date accordingly.

- **Outlier Detection:**

  Identified and removed extreme values in arrival times and other features to reduce noise and improve model accuracy.





Dataset Preprocessing (cont.)

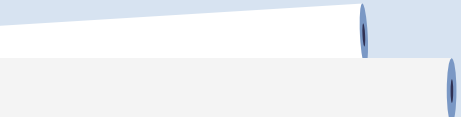
- **Encoding and Normalization:**

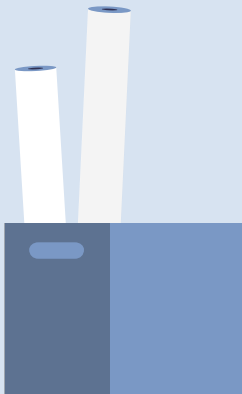
  We encoded categorical variables and normalized continuous features to prepare the data for model training.

- **Sampling:**

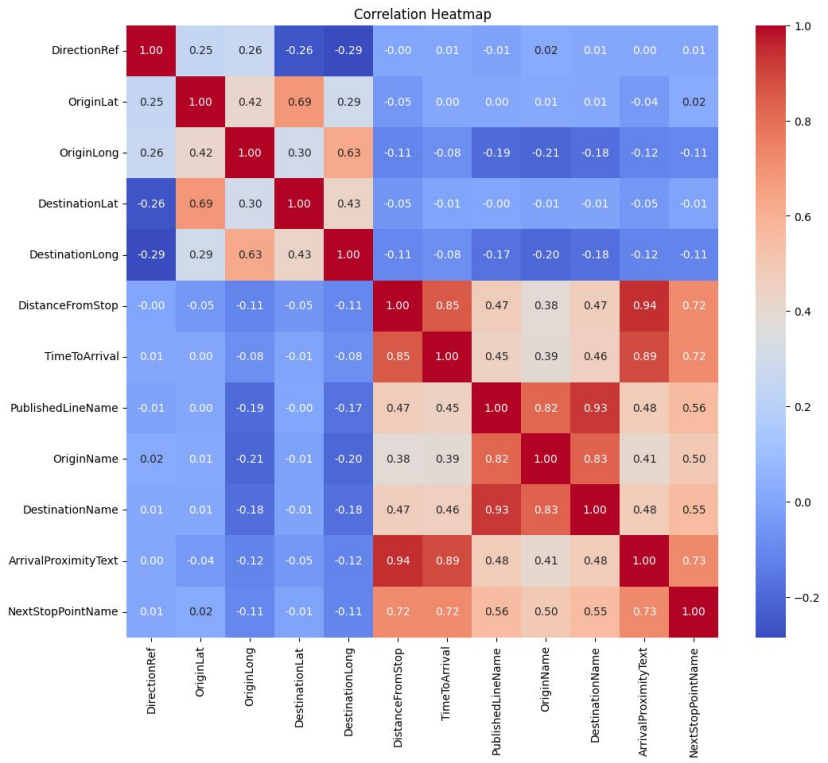
  A random sample of 500,000 records was selected from a dataset of 5 million MTA bus arrival time values. This sampling approach ensures that the subset maintains the statistical properties of the full dataset.

  Used stratified sampling method.





Feature Selection

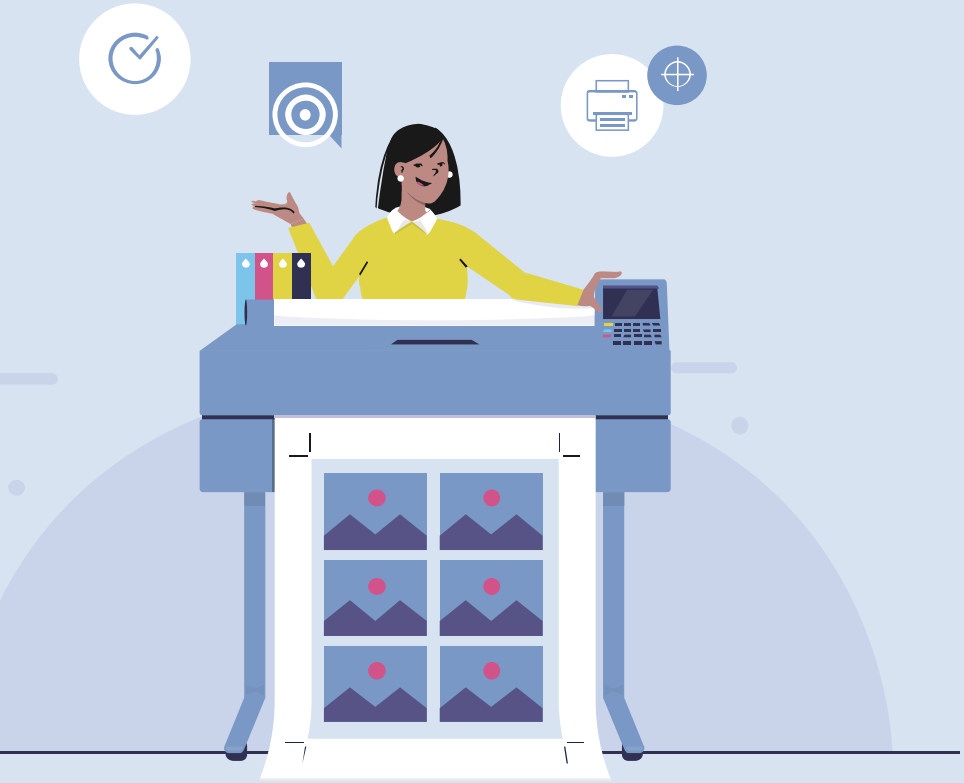


Low Correlation:

- Latitude and Longitude (Origin and Destination) have low correlation.
- Exclude these features as they add minimal predictive value.

High Correlation:

- Origin Name, Destination Name, Published Line Name, and Next Stop Name show strong correlations.
- Prioritize these features for improved model performance.



03

Model Selection

Existing Related Work

Following types of models are used in existing related work:

- Artificial Neural Networks
- Combination of CNN with RNN
- Graph Neural Networks

Also a more recent study using the Kandy bus data proposed a solution using ensemble model of XGboost and ConvLSTM.

From those we used previous approaches of GNN and CNN to compare with the proposed model Spacetimeformer.

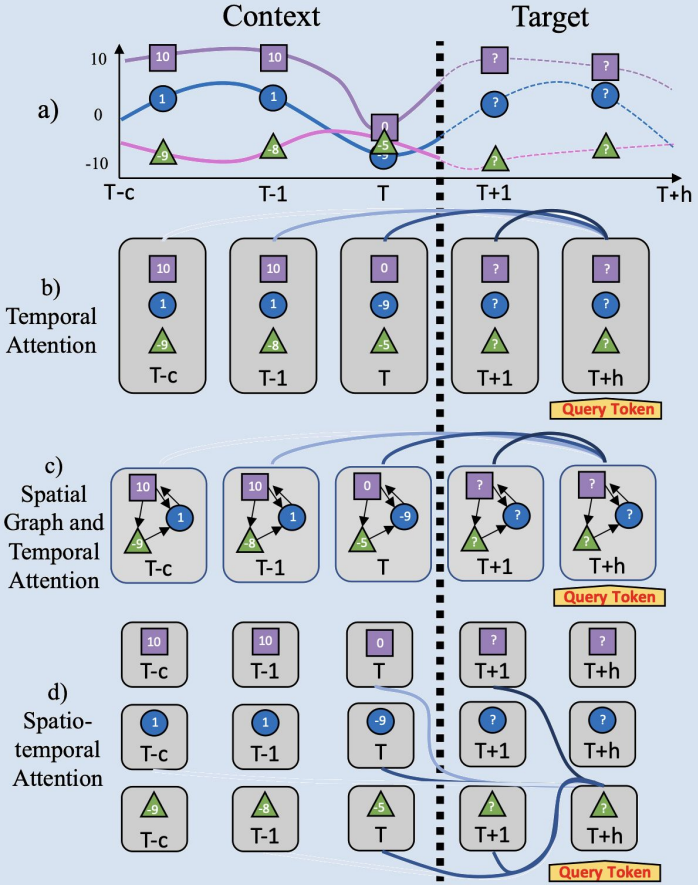
Approach

Spacetimeformer architecture

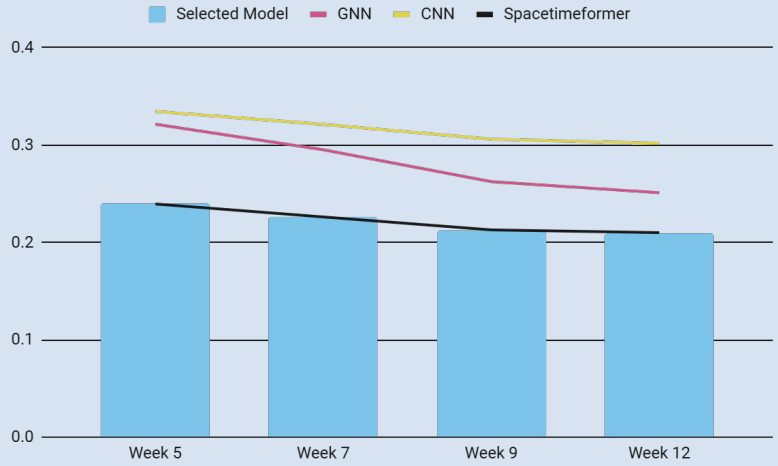
- A state-of-the-art model for dynamic spatiotemporal forecasting

Why Spacetimeformer ?

- Models both spatial and temporal relationships
- Adapts to Dynamic graphs
- Scalability



Model Evolution

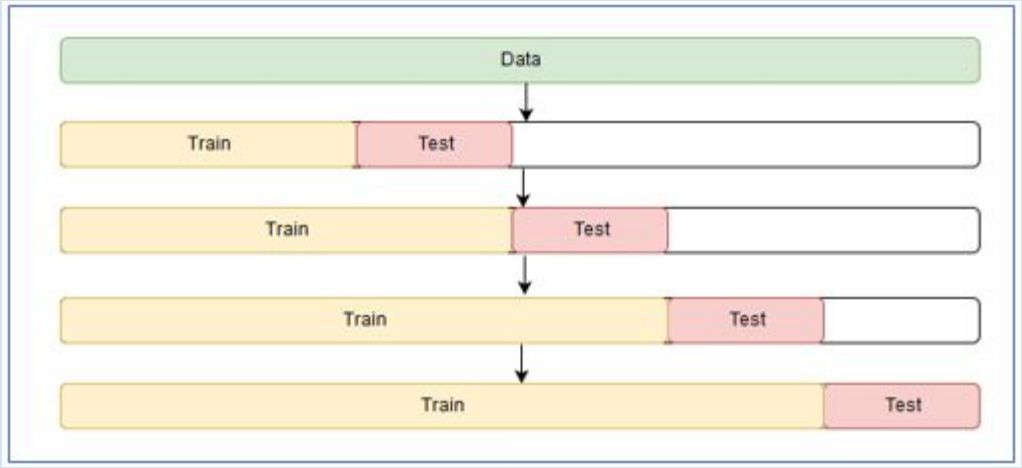


Evolution of RMSE values for each model

Week	GNN	CNN	Spacetimeformer
Week 4			
Week 5	0.3214	0.3345	0.2395
Week 7	0.2951	0.3212	0.2261
Week 9	0.2624	0.3061	0.2128
Week 12	0.251	0.3018	0.21

Hyperparameter Tuning

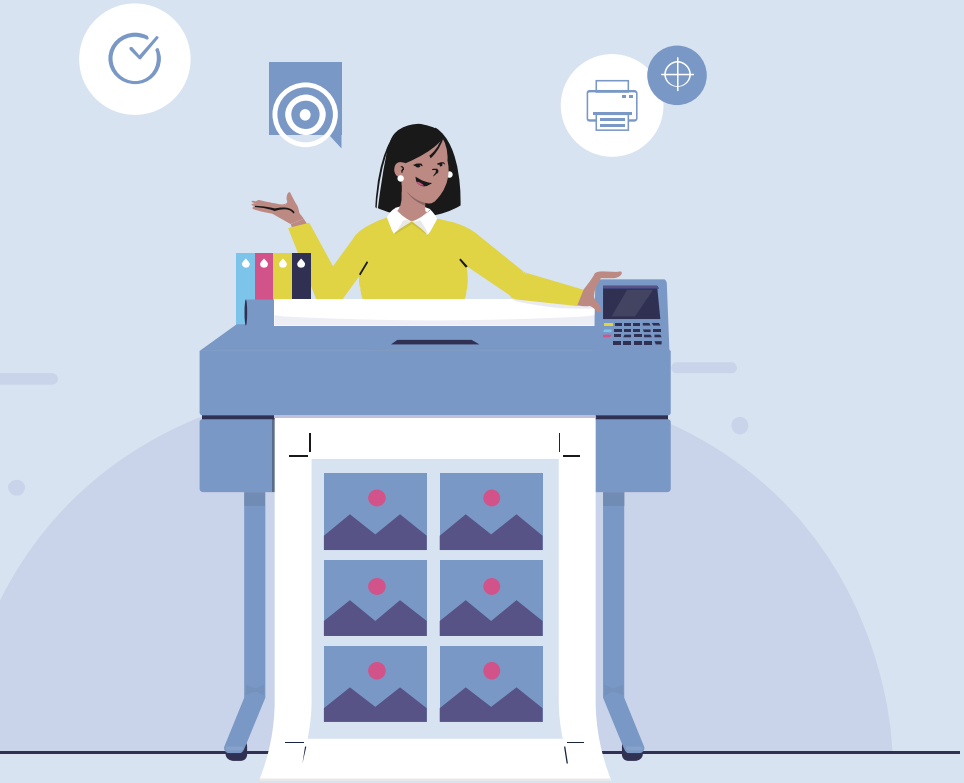
- Spacetimeformer has a set of important hyperparameters that should be tuned for better performance.
- We employed time series cross-validation to identify the most effective hyperparameters for our model.



Final Model Comparison

The table below compares the performance of three models — GNN, CNN and Spacetimeformer, used for bus arrival time prediction, highlighting their accuracy, Mean Absolute Error (MAE), and Mean Squared Error (MSE).

Model	MAE	MSE	RMSE
GNN	0.1889	0.0630	0.2510
CNN	0.1978	0.0911	0.3019
Spacetimeformer	0.1187	0.0364	0.1907

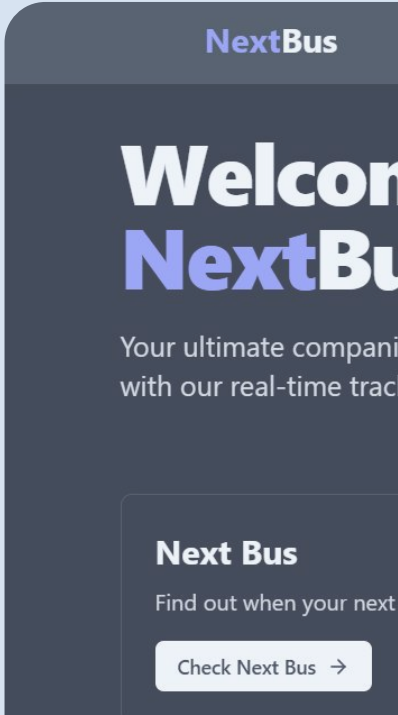


04

Dashboard

Overview

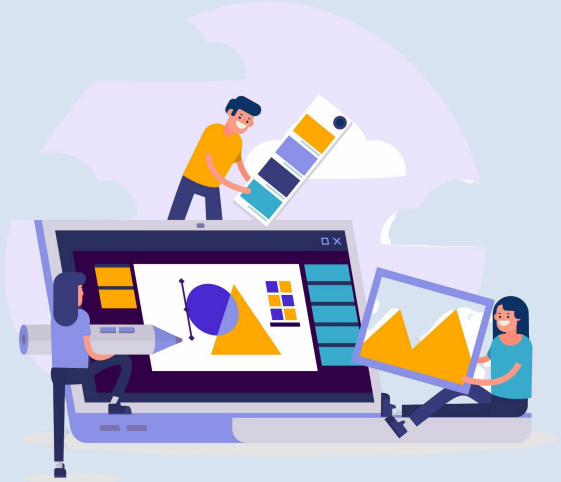
- **React & Vite**
  - Fast and Responsive UI: Built for seamless user interactions.
  - Modern Development Stack: Optimized for high performance and efficiency.
- **Testing with Vitest**
  - Ensures all components function as intended.
- **User-Friendly Design**
  - Attractive Interface: Simple and engaging layout for users.
- **Mobile Responsiveness:**
  - Optimized for smartphones, ensuring easy access for passengers on the go.



Key Features

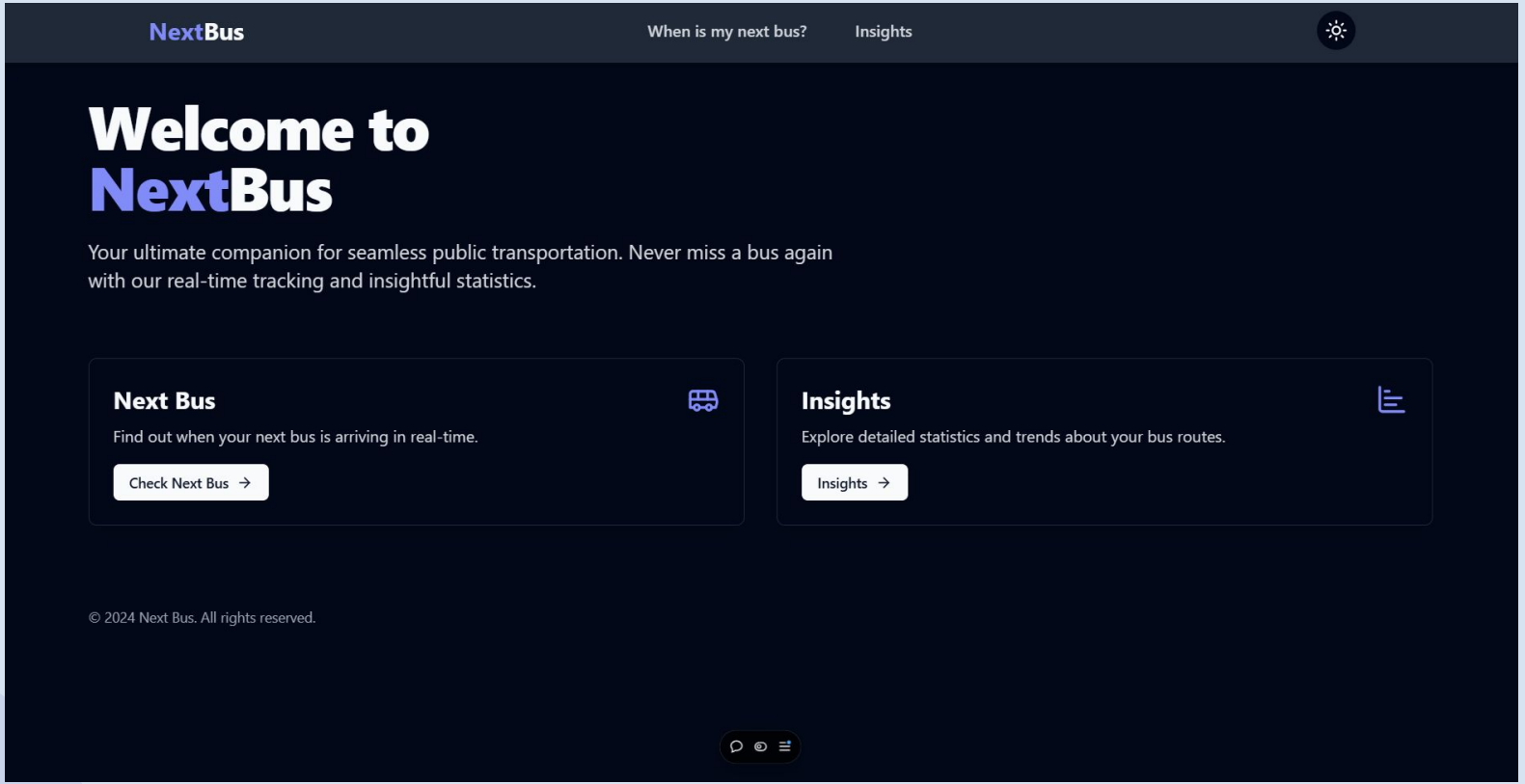
- **Bus Insights**
  - Access schedules and patterns daily, weekly, and hourly.
  - Gain detailed insights to enhance travel efficiency.
- **Helpful Suggestions**
  - Smart Travel Tips: Recommendations for optimal travel times.
  - Reduced Wait Times: Strategies to help users avoid delays.
- **Color Coding**
  - **Green**: Indicates a short waiting time for timely planning.
  - **Red**: Highlights a long waiting time for quick status checks.

Visual Design

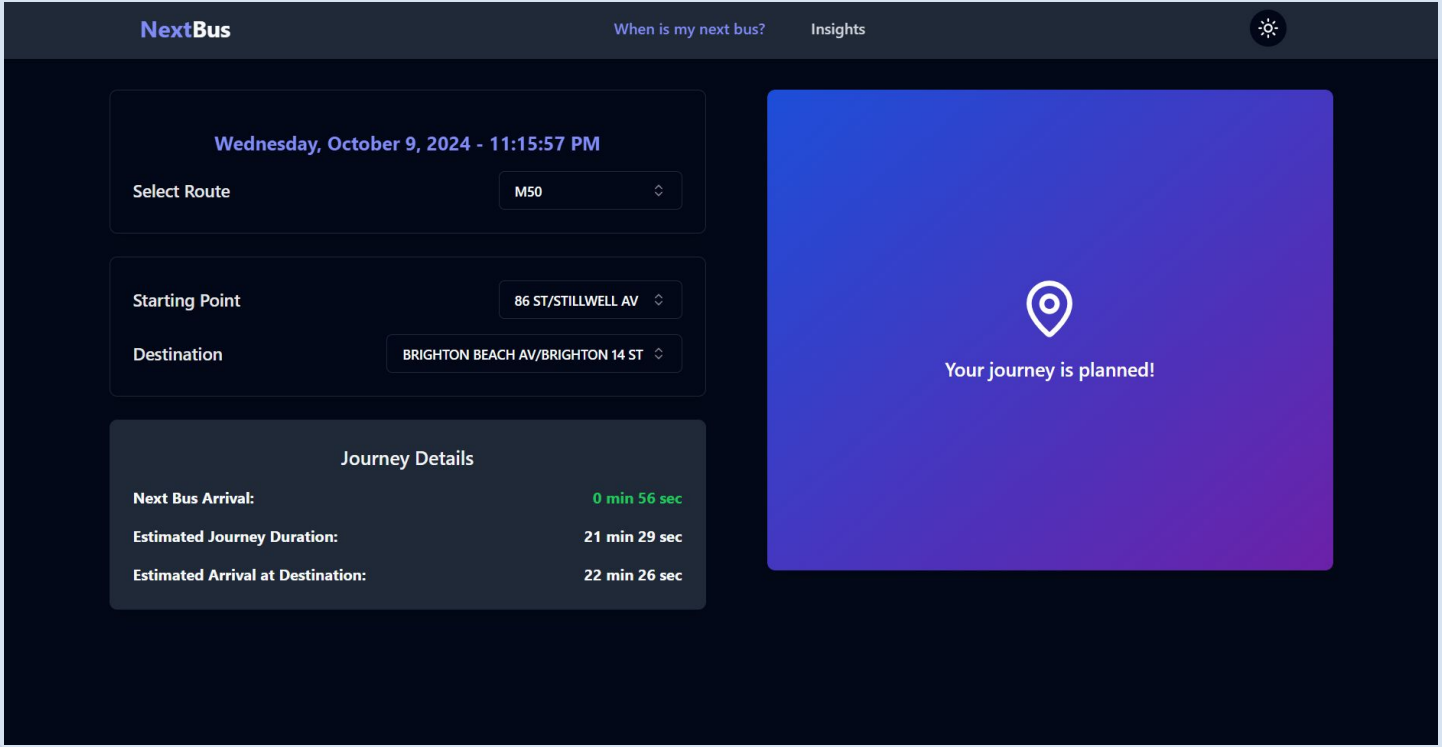


- **Attractive Color Palette**
  - Create a welcoming and friendly atmosphere.
- **Dark and Light Mode**
  - User Preference: Option to switch between modes for comfort in different lighting conditions.
- **Clear Data Display**
  - Straightforward Presentation: Easily digestible schedules and trends.
- **Easy Navigation**
  - Simple Layout: Ensures a seamless and enjoyable user experience.
  - Mobile-Friendly Design: Ensures passengers can easily navigate and access information on their devices.

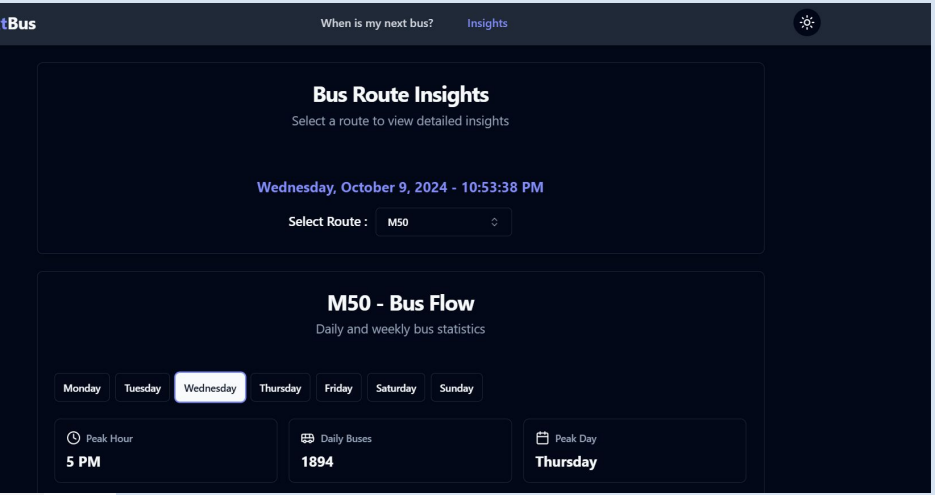
User Dashboard - Landing Page

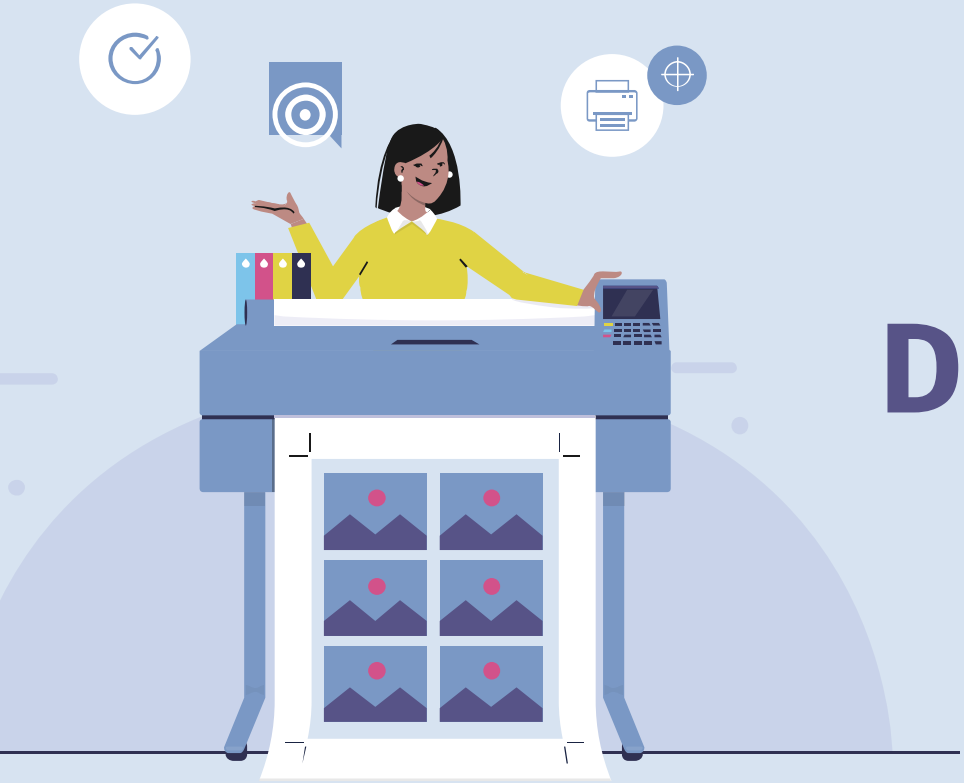


User Dashboard – NextBus



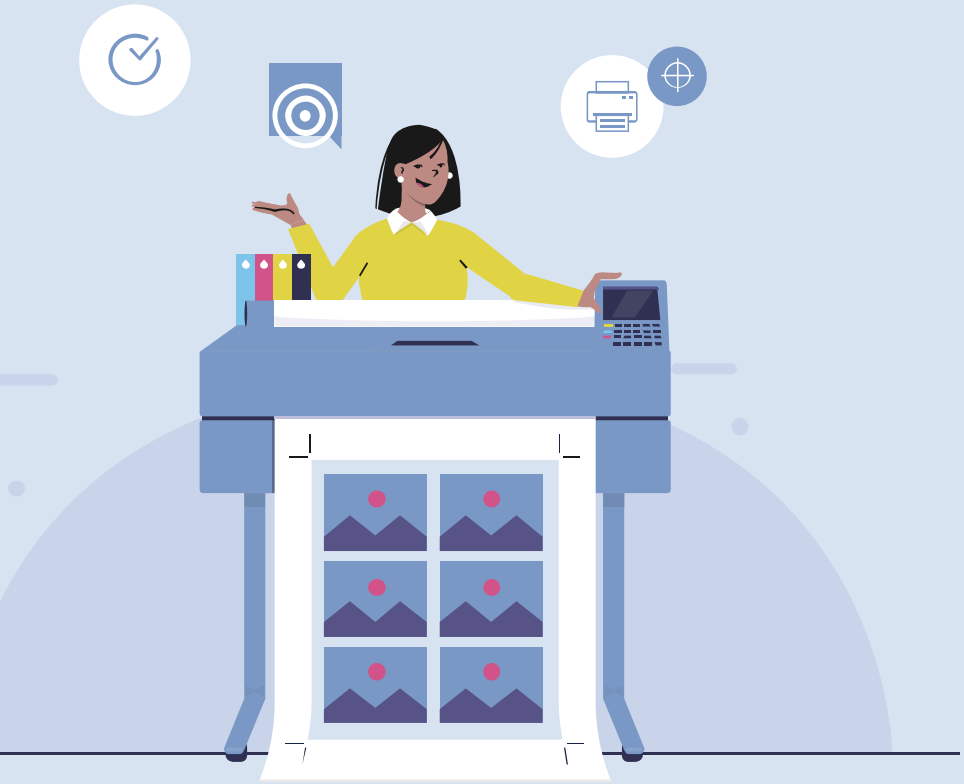
User Dashboard - Insights





05

Demonstration



06

Future Directions

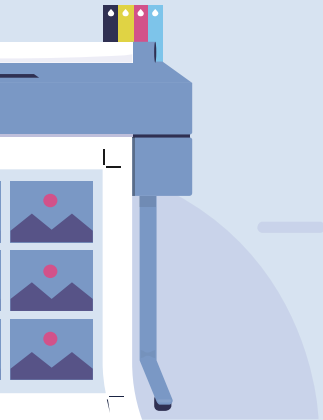
Future Directions

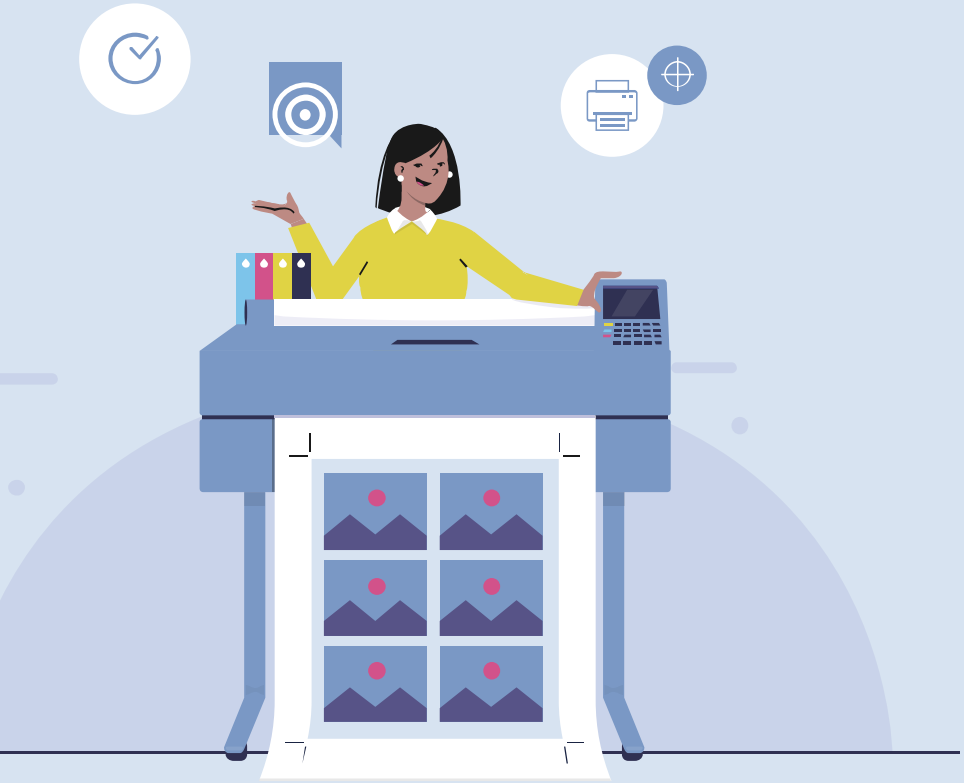
- **Creating a Local Dataset:**

  Our goal is to solve Sri Lanka's transportation challenges, but a suitable dataset with necessary features was unavailable. Using NYC bus data, we demonstrated feasibility, and a future local dataset could enhance prediction accuracy for Sri Lankan bus arrivals.

- **Incorporating Environmental Data:**

  The current dataset lacks weather and other environmental conditions. Integrating this data in future models would improve prediction accuracy by capturing external factors impacting bus arrivals.





07

Contribution

Individual Contribution

Anushanga Sharada
210172N

- Reviewed literature on spatio-temporal forecasting.
- Developed a responsive React dashboard and optimized data fetching.
- Integrated frontend with the FastAPI backend and organized backend routes.
- Supported data cleaning by resolving time format issues and automating conversions.

Deshitha Gallage
210173T

- Reviewed literature on spatio-temporal forecasting.
- Led data cleaning, feature engineering, and sampling.
- Built FastAPI backend with Firebase Firestore with testing.
- Assisted in model selection and training.
- Deployed backend and frontend on Vercel with API integration.

Helith Gothatuwa
210189X

- Reviewed literature on spatio-temporal forecasting.
- Assisted in data cleaning, extracting useful insights from dataset.
- Developed existing approaches like CNN and GNN to compare with the proposed model.
- Assisted in Spacetimeformer training and testing.

Thank You!