TRAFFIC PATTERN ANALYSIS USING GPS DATA IN PYTHON

D.Srinith[1,] G.Sai Kumar Reddy[2,] P. Raghavendra Krishna[3,] Rajkumar B Patil[4.]

Computer Science and Engineering, Vardhaman college of Engineering, Shamshabad road, Kacharam, Telangana, India, 501218.

Abstract:

In present-day urbanized, high-speed environments, it's is imperative to know traffic flow for effective management of transportation. In this project, traffic flow is studied using GPS data, providing useful insights regarding congestion areas, most congested routes and busiest traffic times. By using GPS coordinates, timestamp, and speed data, we are able to trace the movement of vehicles across different zones and spot traffic congestion spots. Involves the gathering and processing of huge amounts of location data from vehicle, smartphones and other GPS devices. The addition of external conditions such as weather and road accidents make's us more precise in our analytics. Employing visualization methods, we can also create interactive maps, graphs and charts that help in the interpretation of results. Lastly, this study aims enhance traffic management methodologies through intelligent urban planning, enhancing routing planning, and processing, and promoting better urban development for smooth and efficient traffic.

Introduction:

Traffic congestion is among the largest issues in contemporary cities, resulting in lost time, more fuel consumption and greater pollution. As urban populations continue to grow, the demand for efficient transportation system is higher than ever before. Conventional traffic monitoring techniques, including manual surveys and fixed-point sensors, tend to be ineffective in offering large scale information. As more devices become available with GPS support, we can now access extensive location data and utilizes it for analyzing and improving traffic flow. Our project uses vehicle GPS data, mobile and navigation system GPS data to reveal interesting facts regarding traffic management. By capturing GPS coordinates, timestamps, and speed, we can trace traffic flow in different zones, detect congestion hotspots, and find peak hours of traveling. Adding external variables such as road conditions, weather, and accidents improves our accuracy. The core objectives of this project to create data-driven options that can prove useful to urban planners, traffic authorities and even common commuters in making better choices. From analyzing the data we can forecast future congestion patterns, provide recommendations and help shape smarter cites. Logistics and business can also be improved by their delivery efficiency through reduction of delays. In addition to merely locating congestions, this can enhance urban mobility. With visualization tools, we can also create interactive maps, graphs and charts that enable easy interpretation of results. This can decrease travel time, decrease fuel usage and encourage sustainability. Through the integration of technology, data analytics and predictive modeling, this project in the future enables cites capable of managing traffic and resulting in a smoother, safer and more efficient transportation system for everyone.

Literature Survey:

Paper Title	Summary of Findings	Methodology Used	Key contributions	Limitations	Future Scope
A Low-Cost GPS-Data enhanced approach for traffic network Evaluations	GPS data improves real-time traffic monitoring accuracy and reduces costs	GPS data collection, preprocessing, machine learning for traffic flow estimation, validation with real-world data	Enhances traffic monitoring offers a cost-effective solution, improves congestion estimation	Limited to specific test regions; GPS inaccuracies affect results	Extending to larger datasets, integrating with IoT-based traffic systems
Corridor level Mobility analysis using Data	GPS data effectively Analyzes mobility at corridor levels and detects congestion trends	GPS trajectory collection, time-space diagrams, speed profile analysis	Identifies congestion patterns, helps in urban mobility planning	GPS errors and missing data affect results	Enhancing predictive Capabilities using AI and real-time data fusion
Analyzing spatiotempor-al Congestion patterns on Urban roads Based on Taxi GPS data	GPS data from taxis reveals urban congestion hotspots and traffic flow variations	Fuzzy C-means clustering, spatial regression modeling on taxi GPS data	Provides insights into congestion formation, helps in transport policy planning	Limited to taxi movement data; does not consider private or public transport	Incorporati-ng diverse vehicle GPS data, improving congestion prediction accuracy
Travel Time prediction for Traveler Information System in Heterogeneous Traffic	GPS based machine learning models accurate predict time in traffic conditions	GPS trajectory collection, machine learning models	Improves travel time prediction, enhances traffic affects model accuracy	Inconsistent data in heterogeneous traffic affects model accuracy	Using deep learning for more accurate predictions, incorporating weather and accident data

Methodology:

1.Data Collection:

To conduct traffic pattern analyze we require data which is collected from various sources including GPS devices installed in vehicles, smartphone GPS data from navigation apps, Traffic sensors at some locations and from external datasets like weather reports and road incidents.

The collected data included essential attributes such as Geospatial Coordinates (Latitude, Longitude), Timestamp, vehicle speed, Direction, Traffic Density and Weather Conditions.

2.Data Preprocessing:

Before analysis, the raw GPS data undergoes a series of preprocessing steps to ensure quality and reliability.

Data Cleaning:

Duplicate records are eliminated from the data to ensure accuracy. Incorrect or missing values were handled using outlier detection.

Data Transformation:

Timestamps were converted into a standard datetime format. Data was filtered to focus on relevant geographic areas.

Data Aggregation:

Data was aggregated into hourly and daily intervals. Traffic data was categorized based on vehicle type.

3.Traffic Flow Analysis:

Traffic flow patterns were analyzed using spatiotemporal visualization techniques such as Congestion intensity at different locations displayed by Heatmaps. Identified peak traffic hours using Time-series plots. And, analyzed average speeds across different times of the day.

4.Pattern Detection and Anomaly Identification:

To identify hidden traffic patterns and anomalies, the following techniques were applied such as

Clustering is used to group the data based on similarities. K-means clustering was used to detect grouped traffic patterns based on speed, location and density.

```
dbscan = DBSCAN(eps=0.5, min_samples=5)#anamoly detction
df["Anomaly"] = dbscan.fit_predict(features_scaled)
anomalies = df[df["Anomaly"] == -1]
print(f"Detected {len(anomalies)} anomalies like roadblocks, accidents, etc.")
```

F(a)

DBSCAN was used to detecting high-traffic regions and anomalies such as accidents, roadblocks and unexpected congestion. Anomaly detection done by statistical methods detecting deviations from normal traffic conditions. Machine learning was used to outlier detection and unusual patterns.

5.Predictive Modeling:

To predict future trends, various machine learning models were used such as linear regression to predicate traffic density based on historical data. LSTM were trained on historical traffic data. Weather conditions and road incidents were incorporated to enhance forecasting accuracy.

6.Data Visualization and Dashboard Development:

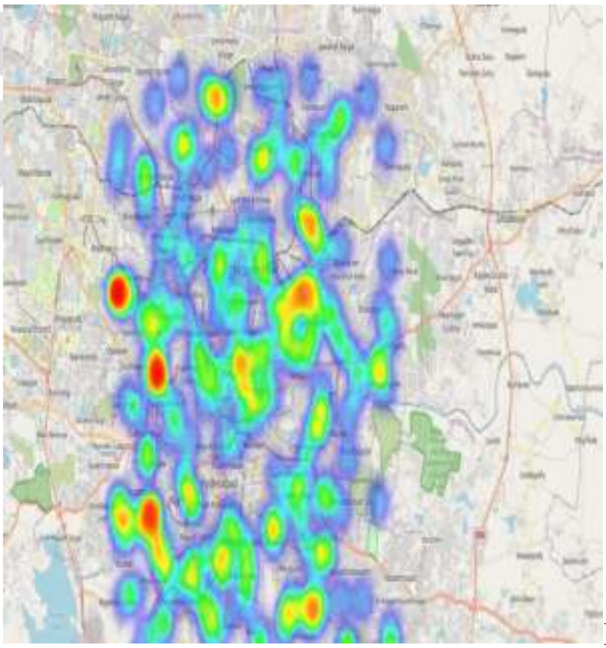
To present the findings, an interactive web-based dashboard was developed using Flask and Folium. Heatmaps were generated for congestion hotspots. Line graphs & bar charts were used for time-series analysis. Real-time traffic alerts were displayed for high-congestion areas. The dashboard provided real-time traffic monitoring and route optimization suggestions for effective traffic management.

7. Reporting & Decision-Making:

Insights derived from the analysis were used for urban Planning: Identifying areas for road expansion or traffic signal optimization. Traffic Management: Recommending alternate routes based on congestion predictions. Real-time alerts for accidents and roadblocks.

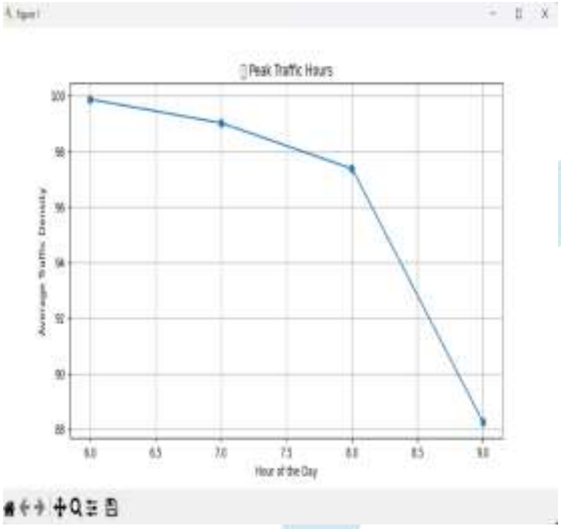
Result:

The project successfully able to visualize the traffic movements across different regions, frequently travelled and least travelled routes. The analysis points main bottleneck locations where traffic slows down. The generated heatmaps from GPS data shows the areas which has high traffic congestions during peak hour.



F(b)

This heatmaps [f(b)] show that how congestions occurred in the specific areas. Red spots show's high congestion, green shows low traffic congestion and yellow-orange shows moderate traffic. Successfully identifying peak traffic hours with congestion to help in better traffic management.



F(c)

This graph f(c) shows that high traffic occurs between 6pm to 8pm done by time series analysis. The Road accidents and weather conditions were also included to visualize congestion to better traffic management. Also detected unusual patterns using DBSCAN that indicates road closures and accidents. K-means clustering helped to identify constant high traffic areas to helping in better urban planning. The project abled to forecasting of future traffic conditions based on historical data. The accuracy of the project is very good. The MAE is 2.5 for the solution of this project.

This figure(d) shows the interaction output of the project which gives details of density of traffic, accidents and peak hours of the specific GPS coordinates.

	Timestamp	Latitude	Longitude	Speed	Direction	Weather	Road_Incident	Vehicle_Type	Traffic_Density	Hour_of_Day	Traffic_Cluster
0	2025-03-25 06:00:00	[illegible]	[illegible]	32	East	Fog	NaN	Truck	187	6	2
1	2025-03-25 06:01:00	[illegible]	[illegible]	68	West	Cloudy	NaN	Bus	120	6	1
2	2025-03-25 06:02:00	[illegible]	[illegible]	32	South	Cloudy	NaN	Bus	73	6	0
3	2025-03-25 06:03:00	[illegible]	[illegible]	27	North	Clear	NaN	Car	98	6	1
4	2025-03-25 06:04:00	[illegible]	[illegible]	56	North	Cloudy	NaN	Bus	140	6	1
5	2025-03-25 06:05:00	[illegible]	[illegible]	32	North	Clear	NaN	Truck	94	6	3
6	2025-03-25 06:06:00	[illegible]	[illegible]	26	West	Clear	NaN	Truck	97	6	2
7	2025-03-25 06:07:00	[illegible]	[illegible]	22	West	Rain	NaN	Bus	178	6	2
8	2025-03-25 06:08:00	[illegible]	[illegible]	88	South	Cloudy	NaN	Car	131	6	1
9	2025-03-25 06:09:00	[illegible]	[illegible]	29	East	Fog	Road Work	Car	113	6	2
10	2025-03-25 06:10:00	[illegible]	[illegible]	40	North	Cloudy	Accident	Truck	173	6	2
11	2025-03-25 06:11:00	[illegible]	[illegible]	20	North	Cloudy	NaN	Truck	129	6	2
12	2025-03-25 06:12:00	[illegible]	[illegible]	78	East	Cloudy	Heavy Traffic	Truck	28	6	0

F(d)

Apart from identifying congestion trends and predicting traffic conditions, the systems offer's actionable insights for urban infrastructure planning by exploiting geospatial data. By detecting clusters of high traffic density through latitude and longitude coordinates, the model highlights recurring slowdowns. These observations allow the system to suggest particular interventions, like the development of flyovers at interactions where traffic from different directions always overlaps and congestion continues even after adjustments in signal timings. Likewise, widening of roads is proposed for extended periods, which means existing road capacity inadequate.

Flyover Suggestions (High-Density Points)

	Latitude	Longitude	count
0	[illegible]	[illegible]	1
1	[illegible]	[illegible]	1
2	[illegible]	[illegible]	1
3	[illegible]	[illegible]	1
4	[illegible]	[illegible]	1
5	[illegible]	[illegible]	1
6	[illegible]	[illegible]	1
7	[illegible]	[illegible]	1

F(e)

The flyover recommendation module identified intersections and bottleneck-intensive points where traffic from multiple directions meet, causing congestion. From F(e) coordinates like latitude 17.319500, longitude 78.491814 and latitude 17.387532, longitude 78.492508 were identified as key flyover candidates. These locations are usually aligned with high intersection complexity or junctions with heavy signals.

Road Widening Suggestions (Consistently High Traffic Areas)

	latitude	longitude	count
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

F(f)

The road expansion recommendations were derived for linear sections of roads that showed uniformly high traffic load over a period, as opposed to spikes or one-off occurrences. From F(f), latitude 17.315521, longitude 78.523850 and latitude 17.327234, longitude 78.495012 are sections where traffic intensity is sustained throughout the day, and mean speeds are low. Hence suggested for upgradation.

Future Extensions & Conclusion:

The project is good enough to find insights of traffic patterns but to enhance more the project, we can propose the AI-powered route optimization using reinforcement learning to recommend the alternate paths. And also deploying AI at traffic signals to dynamically adjust the time for stop and go based on the congestion. Also adding social media and events data to refine the congestion predictions. Building an app to provide traffic recommendations. This is study successfully demonstrates the use of GPS data and Machine learning algorithms to analyse traffic patterns. By conducting data preprocessing, clustering, anomaly detection, and predictive modelling, we effectively identified traffic zones, peak hours. Accuracy depends on data quality and real time updates like weather, events, etc.

References:

1.Andrienko G., & Andrienko N. (2008), spatio-temporal aggregation for visual analysis of movements. IEEE Visual Analytics Science and Technology, pp 51-58, IEEE Computer Society Press.

2. Atoniou, C., & Koutsopoulos, H.N. (2006). Estimation of traffic dynamics models with machine learning methods. Transportation Research Record, volume 1965, pp 103-111, Washington, DC.

3. Azimi, M., & Zhang, Y. (2010). Categorizing freeway flow conditions using clustering methods. Proceedings of the 89Th Annual Meeting of the Transportation Research Board, Washington, DC.

4. Dodge, S., Weibel, R., & Lautenschutz, A.k. (2008). Towards a taxonomy of movement patterns. Information Visualization, 7(3-4), 240-252.

5.Bar-Gera, H. (2007). Evaluation of a cellular phone-based system for measuring traffic speeds and travel times.

6.Jenelius, E., & Koutsopoulus, H. N. (2013). Travel time estimation for urban road networks using low-frequency GPS probes. Transportation Research part B: methodological, 53, 64-81.

7.Wang, y., Papageorgiou, M., & Messmer, A. (2008). Real-time freeway traffic state estimation based on extended Kalman filter: A general approach. 42(3),283-301.

8. Google Traffic Analysis Report (2021). How GPS data and AI improve real-time traffic predictions. Google Research Blog.

9. Waze Data Science (2020). Crowdsourced GPS data for real-time traffic management. Waze Analytics Blog.