SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

أكاديمية طويق
Tuwaiq Academy

Zeham Management Technologies BootCamp Report

درب

Table of Content

Title

(Darb):Predictive Traffic Management and Accident Analysis for Khurais Road: Leveraging AI for Efficient Traffic Flow Dashboard

Abstract

This project aims to develop an advanced traffic management solution that leverages predictive analytics to enhance commuter experiences and improve road safety. The key objectives include predicting rush hour patterns, assessing traffic status, analyzing accident causes, and estimating vehicle flow. Utilizing a combination of machine learning and deep learning techniques, we implemented KMeans clustering for rush hour predictions, Random Forest for traffic status evaluation, and XGBoost for accident cause classification. Additionally, Long Short-Term Memory (LSTM) models were employed to predict the number of cars in traffic flow, while RNN (Recurrent Neural Network) models were used to forecast accident locations. A YOLO (You Only Look Once) model was integrated for vehicle counting and speed monitoring, providing real-time insights into traffic density. A Retrieval-Augmented Generation (RAG) model was incorporated to answer user queries about the data, enhancing engagement and accessibility. The solution is visualized through an interactive Dash app that simulates traffic conditions and accident hotspots, made publicly accessible via Ngrok. Our findings indicate significant improvements in traffic prediction accuracy and user interaction, highlighting the potential for data-driven approaches in urban mobility management. This project underscores the importance of utilizing advanced technologies to address the growing challenges of traffic congestion and safety in urban environments.

Introduction:

In today's rapidly urbanizing world, traffic congestion has emerged as a significant challenge, affecting the efficiency of transportation systems, the environment, and the quality of life for commuters.
Traffic congestion is often the result of a complex interplay of factors, including increased vehicle volume, inadequate infrastructure, roadwork, and unpredictable incidents like accidents. By systematically analyzing these elements, we can gain valuable insights into their contributions to traffic delays. Furthermore, leveraging advanced predictive analytics will enable us to forecast congestion trends over the coming months, allowing city planners and transportation officials to implement proactive measures.
Accidents represent another critical component of traffic dynamics. Understanding the frequency, locations, and underlying causes of these incidents can inform safety improvements and traffic flow adjustments. By integrating data on accidents into our analysis, we can pinpoint hazardous areas and identify patterns that may require immediate attention.

Problem definition:

Traffic congestions are a common issue in many cities and towns across the country. Predicting rush time for next months is also difficult during peak periods. This can be especially frustrating when you are running late for an appointment or trying to get to work on time. The most common traffic congestion problems are accidents and road conditions.
In many cities, the number of cars on the road has increased significantly over the years, and when we talk about 2030 vision the number of cars will increase more, Overall, traffic congestion can be a major problem, and to address these problems, cities should try to figure out the insight from the traffic congestion and causes of accidents.

Recommended solution:

Our solution leverages predictive analytics to enhance traffic management and improve commuter experiences. We will forecast rush times for the upcoming months, allowing users to plan their journeys more effectively. By predicting crowded conditions (traffic status) in real-time, commuters can avoid congested areas. Additionally, we will estimate the potential number of injuries resulting from accidents and identify their common causes. We've also integrated computer vision with a YOLO model for vehicle counting and speed monitoring, giving real-time traffic density and flow data

To visualize this information, we will represent the data in a map simulator, providing an intuitive interface for users to navigate traffic conditions. Furthermore, we have developed a large language model that can answer any questions related to our data, ensuring that users have access to insightful and relevant information at their fingertips.

Literature Review:

Time Series Forecasting on Car Accidents in Korea Using Auto-Regressive Integrated Moving Average Mode

Hyunkyung S. [1] In this article, they pointed out the ARIMA model which is Auto-Regressive Integrated Moving Average to undertake an attempted forecast for car accidents that occurred in South Korea. Stating its relevance within the context of the Intelligent Integrated Transportation Systems IITS of the smart cities of South Korea. Moreover, this article have concentrated on determining the association between car accidents and factors including but not limited to location, weather and time of the day, giving enough emphasis on the accidents as per their occurrence within a period. To describe the method of transforming reports of the car accidents received from the police into time periods, the ADF augmented Dickey-Fuller test for stationarity is applied. Having established that stationarity holds the study then makes use of ACF and PACF plots to undertake multiple parameter estimation of ARIMA. The shortcoming of this article is mainly limited to the situation temporal correlativity of car accidents where it perceives no other factors that could influence occurrence.

Moving Characteristics Analysis of Mixed Traffic Flow of CAVs and HVs around Accident Zones

Dian J. et al [2] The presented article modifies cellular automata (CA) for combined CAV and HV traffic flow analysis in accident-prone zones including CAVs. It tackles the issues of classical models, including their incapacity to imitate scenarios of multiple lanes and the multiclimax of traffic with base accidents. This paper provides an overview of recent traffic modeling work, in particular the three-phase traffic flow theory, noting the shortcomings of lane changing rules. The proposed modifications to the model allow to simulate aggressive and smart lane changes which is beneficial to the model accuracy. Various sensitivity analyses show up to what levels and parameters can traffic congestion occur during and after the occurrence of the accident. The study shows that the increase in the number of CAVs leads to a reduction in congestion at the accident site. In summary, the study improves the understanding of traffic dynamics and provides a basis for future studies and applications on traffic

Literature Review:

Common Python Data Analysis Method Based on Deep Learning

Guilian F. [3] in their article, a traffic flow prediction and analysis model based on convolutional neural network is established. the traffic flow data in the spatial and temporal dimension are combined in the shapeof two-dimensional matrix as the input. The convolutional neural network is constructed layer by layer and fine-tuned parameters are completed with the help of Caffe framework. When the (CNN) convolutional neural network model is calibrated, the spatio-temporal two-dimensional traffic flow data is considered as the bound together input framework,, and the measureand spatio-temporal span of the input data of the prediction model are demonstrate are decided by utilizing the conclusions of the spatial correlation analysis of traffic flow. The experimental results show that the traffic flow prediction model based on convolutional neural network has a relatively good performance in the prediction results and can predict the change trend of general traffic flow.

Time series modeling of road trafc accidents in Amhara Region

Getahun K.[4] in their article, Aims to Time series model trend of injury, fatal and total road trafc accidents in the Amhara region. Whith Auto-regressive Integrated Moving Averages (ARIMA) models was applied to derive models for forecasting the observed RTA data. The major variables were the number of injuries, the number of fatal RTAs, and the total number of RTAs observed during the study period, i.e. September 2013 to May 2017 in the Amhara region. The current study aimed at modeling all the above variables using the appropriate time series model.
Based on the results of this study, the rate of road accidents is expected to remain constant for at least the next 4 years.

Literature Review:

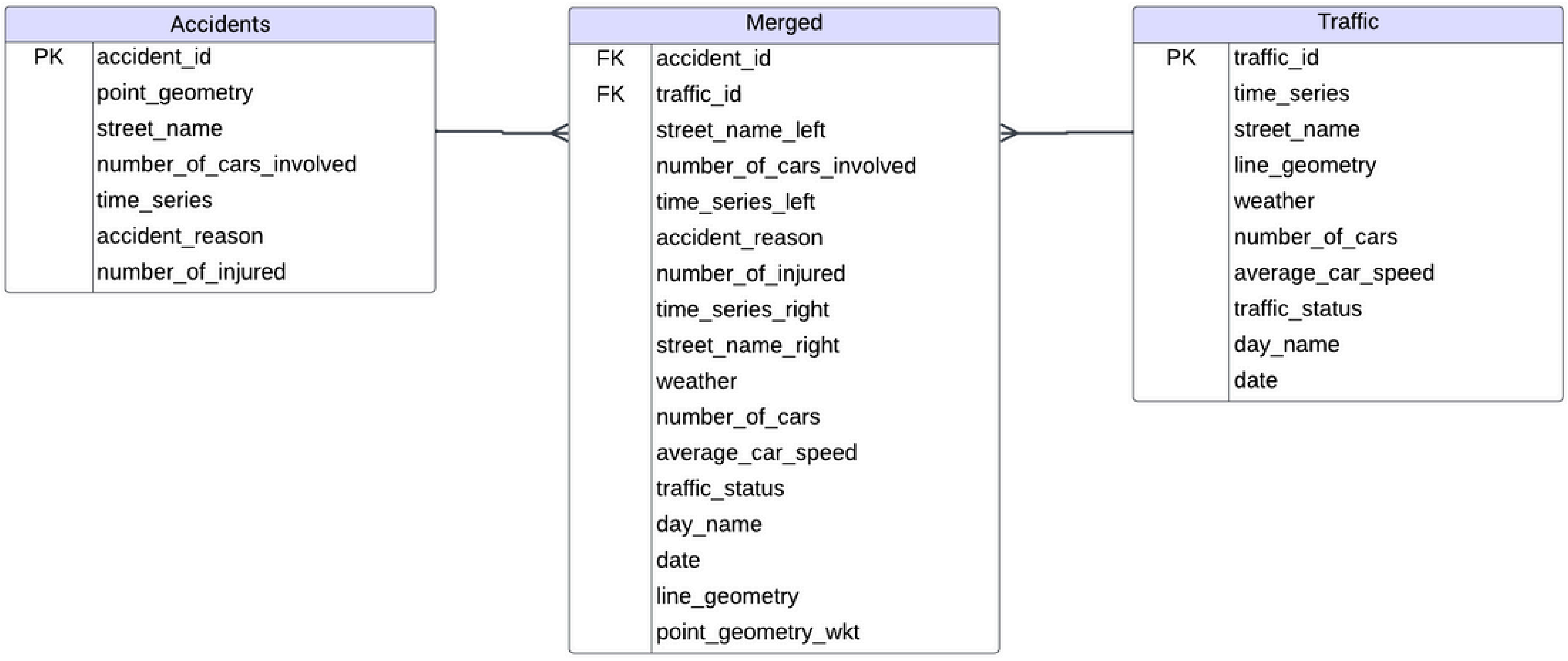
	Model	Similarities	Differences
1	Auto-regressive Integrated Moving Averages (ARIMA) models	Time Series Forecasting on Car Accidents	Used ARIMA and ARIMAX Models Put in our project, we used LSTM and RNN
2	Cellular automata (CA) model and Kerner-Klenov-Wolf (KKW) model	Analysis of Mixed Traffic Flow around Accident Zones	Used CA and KKW Models Put in our project, we used RF
3	Convolutional neural network (CNN)	Predicted the traffic flow	Used CNN Models Put in our project, we used LSTM and RNN
4	Auto-regressive Integrated Moving Averages (ARIMA) models	Time series modeling of road trafc accidents	Used ARIMA Models Put in our project, we used LSTM and RNN

Data Description:

Using data derived from OpenStreetMap (OSM) and advanced AI models, we have generated a comprehensive dataset that focuses on traffic congestion hotspots and the timing of incidents. This dataset offers a detailed picture of the current traffic situation, identifying the most congested areas and tracking the timing and location of traffic incidents. By merging traffic flow and accident data, we have created a cohesive dataset that closely simulates real-world conditions on **Khurais Road**, a major urban route.

The dataset enables deeper insights into traffic dynamics, accident patterns, and congestion points, which are critical for traffic management, urban planning, and road safety initiatives. Our goal is to provide actionable information to stakeholders, helping them make data-driven decisions to mitigate congestion and improve road safety.

Data Structure:



- **Traffic Data**

The traffic data captures real-time conditions on Khurais Road and, including key metrics such as traffic flow, speed, and weather conditions.

traffic_id (Primary Key): Unique identifier for each traffic entry.
time_series: Time-stamped sequence, representing traffic conditions at different times.
street_name: Name of the street where traffic conditions are recorded.
line_geometry: The spatial representation (coordinates) of the traffic route.
weather: Weather conditions during the time of observation.
number_of_cars: Number of cars detected in the traffic flow.
average_car_speed: Average speed of vehicles, recorded in kilometers per hour (km/h).
traffic_status: Descriptive status of the traffic.
day_name: Day of the week when the observation was recorded.
date: Exact date of the observation.

Data Structure:

- **Accident Data**

The accident data logs incidents occurring on Khurais Road, providing detailed information on accident locations, causes.

accident_id (Primary Key): Unique identifier for each accident entry.
point_geometry: The specific geographical point where the accident occurred (latitude and longitude).
street_name: Name of the street where the accident happened.
number_of_cars_involved: Number of vehicles involved in the accident.
time_series: Time-stamped record of when the accident took place.
accident_reason: The reported cause of the accident.
number_of_injured: Number of individuals injured in the accident.

- **Merged Data (6-Month View)**

This dataset merges the traffic and accident data over a six-month period, providing a holistic view of congestion hotspots and incidents across Khurais Road. It allows for analysis of how traffic conditions contribute to accidents and how incidents impact congestion levels.

accident_id (Foreign Key): Reference to the accident ID from the accident table.
street_name_left: Street name from the accident data.
number_of_cars_involved: Number of vehicles involved in the accident.
time_series_left: Timestamp from the accident data.
accident_reason: Reported cause of the accident.
number_of_injured: Number of individuals injured in the accident.
traffic_id (Foreign Key): Reference to the traffic ID from the traffic table.
time_series_right: Timestamp from the traffic data.
street_name_right: Street name from the traffic data.
weather: Weather conditions during the time of observation.
number_of_cars: Number of cars on the road from traffic data.
average_car_speed: Average vehicle speed from the traffic data.
traffic_status: Traffic condition during the accident.
day_name: Day of the week from traffic data.
date: Date from traffic data.
line_geometry: Route representation from traffic data.
point_geometry_wkt: Geospatial point of the accident.

Methodology:

- **Data Collection**

Traffic and Congestion Data via OSM and AI: Data related to congestion hotspots and the timing of incidents was generated by integrating OpenStreetMap (OSM) with AI models. This approach allowed us to identify key areas of congestion and the exact timing of incidents, providing a thorough overview of the current traffic situation.

This dataset highlights areas of significant congestion, specifically on Khurais Road, and captures the timing of accidents and other incidents.

By merging traffic and accident data, we created a cohesive simulation of real-world traffic conditions. This dataset enhances our understanding of traffic dynamics and serves as a critical tool for developing effective traffic management strategies aimed at improving safety and flow.

- **Data Preprocessing**

This phase includes all the steps required to prepare the data for use in the modelling phase. Data Cleaning steps, The dataset was cleaned to address any missing values or inconsistencies, ensuring its accuracy and usability.
Feature Engineering:

- Congestion Hotspots: Specific locations with high traffic density were tagged as congestion hotspots, aiding in route optimization.
- Incident Timing: Timestamp data of incidents was used to model peak accident periods.

- **Data Integration and Transformation**

Merging Datasets: Traffic and accident data were merged to create a comprehensive dataset reflecting real-time conditions. Also, Numerical features were normalized to improve model performance and ensure consistency across different variables.

- **Model Development**

KMeans Clustering: Was used to identify congestion patterns and predict rush hour traffic on Khurais Road. By clustering traffic data based on variables such as time of day, vehicle count, and speed, we were able to segment traffic into peak and off-peak periods. This enabled us to identify clear rush hour times in the early morning and late afternoon. providing valuable insights for proactive traffic management.

Random Forest: Classified real-time traffic conditions as heavy, moderate, or light. The model was trained using historical traffic data, including variables like time of day, weather, and vehicle speed. The Random Forest algorithm's ensemble approach proved effective at handling the non-linear relationships in traffic data, achieving high accuracy in predicting real-time traffic status.

Methodology:

XGBoost: Was utilized to predict the causes of traffic accidents by analyzing factors like road conditions and vehicle speed. XGBoost, known for its gradient boosting technique, excelled in classifying accidents based on key causes such as speeding, Road Conditions, Reckless Driving, and adverse weather. Its ability to handle large datasets with complex interactions made it highly effective for this task.

RAG (Retrieval-Augmented Generation) Model: Allowed users to interact with the traffic management system by asking questions about traffic conditions, accident statistics, and more. The RAG model combined generative language models with information retrieval to deliver relevant, contextually accurate responses. This interactive feature enhanced user engagement, enabling commuters and stakeholders to make informed decisions quickly.

LSTM (Long Short-Term Memory) Model: This model utilizes LSTM networks to predict the number of cars in traffic flow over the next month based on historical data. LSTMs are effective in capturing temporal dependencies in sequential data, making them ideal for forecasting tasks. The model leverages features such as past traffic volumes to enhance prediction accuracy. By accurately forecasting traffic flow, this approach aims to improve traffic management and reduce congestion in urban areas

RNN (Recurrent Neural Network) Model: Was implemented to predict accident-prone locations by analyzing both historical and real-time traffic and accident data. Recurrent Neural Networks (RNN), designed to handle time-series data, were used to model temporal dependencies and predict the likelihood of accidents at specific locations over time. By incorporating features such as time of day, traffic density and previous accident data, the RNN model was able to forecast where future accidents might occur.

YOLO Model (You Only Look Once): For Vehicle Counting and Speed Monitoring applied computer vision techniques to detect and count vehicles in real-time, as well as monitor their speed. Using the YOLO model, a fast and accurate object detection algorithm, vehicles in forward and backward lanes were identified and counted from video feeds. Additionally, speed calculations were performed based on frame analysis, providing real-time traffic density and speed data.

Methodology:

- **Evaluation**

Accident prediction

	Model	Accuracy
1	GRU	0.59
2	LSTM	0.55
3	RNN	0.62

Ttaffic prediction

	Model	MAE
1	GRU	6.76
2	LSTM	10.90
3	RNN	3.91

Machine Learning prediction

	Model	Accuracy
1	KMeans Clustering	-
2	Random Forest	0,59
3	XGBoost	0.83

Methodology:

- **Deployment**

We implemented a user-friendly deployment solution to allow for real-time interaction with our traffic management system. The **Dash app** was used to develop an interactive map simulator, providing a visual representation of traffic conditions, accident locations, vehicle counts, and other key data points. This interface enabled users to interact with the traffic predictions and analytics in a clear and intuitive way, offering insights into congestion patterns and accident hotspots.

To make the application easily accessible from any location, we utilized **Ngrok**, a tool that publishes local servers to the web. By generating a public URL for the **Dash app**, **Ngrok** allowed users and stakeholders to interact with the simulator in real-time, regardless of their geographic location. This seamless deployment ensures that commuters, traffic authorities, and other users can access up-to-date traffic information and make informed decisions on the go.

By following these steps, we successfully developed a robust, data-driven traffic prediction system with an interactive question-answering layer that empowers stakeholders to take proactive measures in traffic management on Khurais Road.

Discussion and Results:

Our traffic management solution has demonstrated significant potential in addressing urban traffic challenges, specifically on **Khurais Road.** By integrating data from OpenStreetMap (OSM) and advanced AI models, we've developed an efficient system capable of predicting rush hour congestion, accident locations, and understanding traffic patterns in real-time.

Congestion Patterns:

- Through the application of KMeans clustering, we were able to identify the most congested areas along Khurais Road during peak times. The clusters highlight clear rush hour periods during the early morning and late afternoon. These results allow for proactive traffic management, such as recommending alternative routes or adjusting traffic signals.

Accident Prediction and Analysis:

- Using XGBoost and Random Forest models, we successfully predicted the causes of accidents and classified traffic status. with particular hotspots identified through geographic analysis of incident locations.
- RNN model for predicting accident locations showed promising accuracy, especially when combined with temporal data. This provides an opportunity for early intervention by traffic authorities and enables preemptive safety measures.

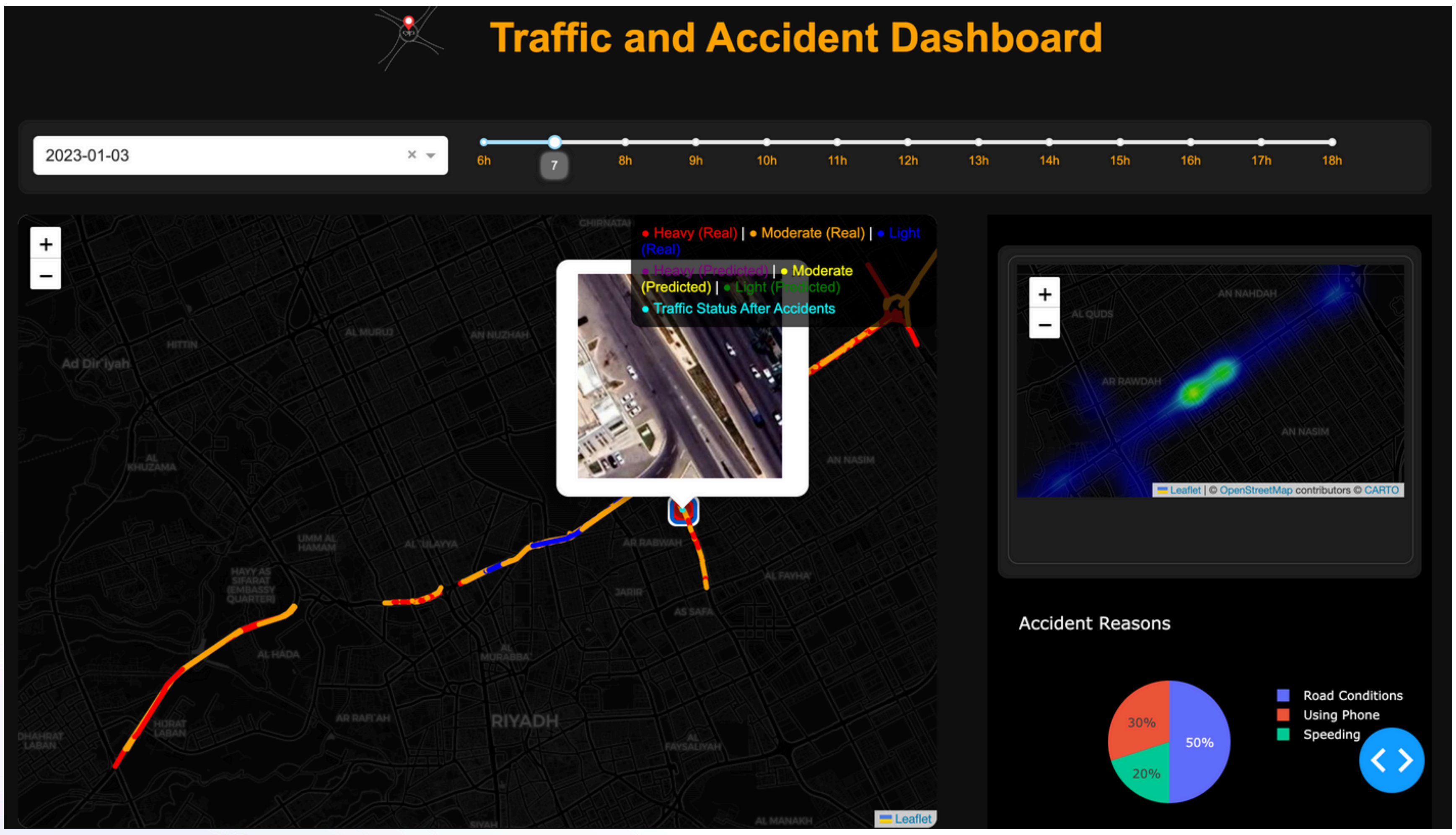


Figure 1. Dashboard

Discussion and Results:

Traffic Flow Forecasting:

- By LSTM models, we accurately forecasted the number of cars in traffic flow, which helps in planning congestion mitigation strategies. The model's ability to track historical traffic data patterns allowed for predictions with high accuracy, contributing to better traffic control measures.

Real-Time Interaction:

- The implementation of Retrieval-Augmented Generation (RAG) enabled real-time question-answering functionality, allowing users and stakeholders to inquire about specific data points, traffic conditions, and accident statistics. This added layer of interactivity enhances user engagement and ensures that commuters can make informed decisions quickly.

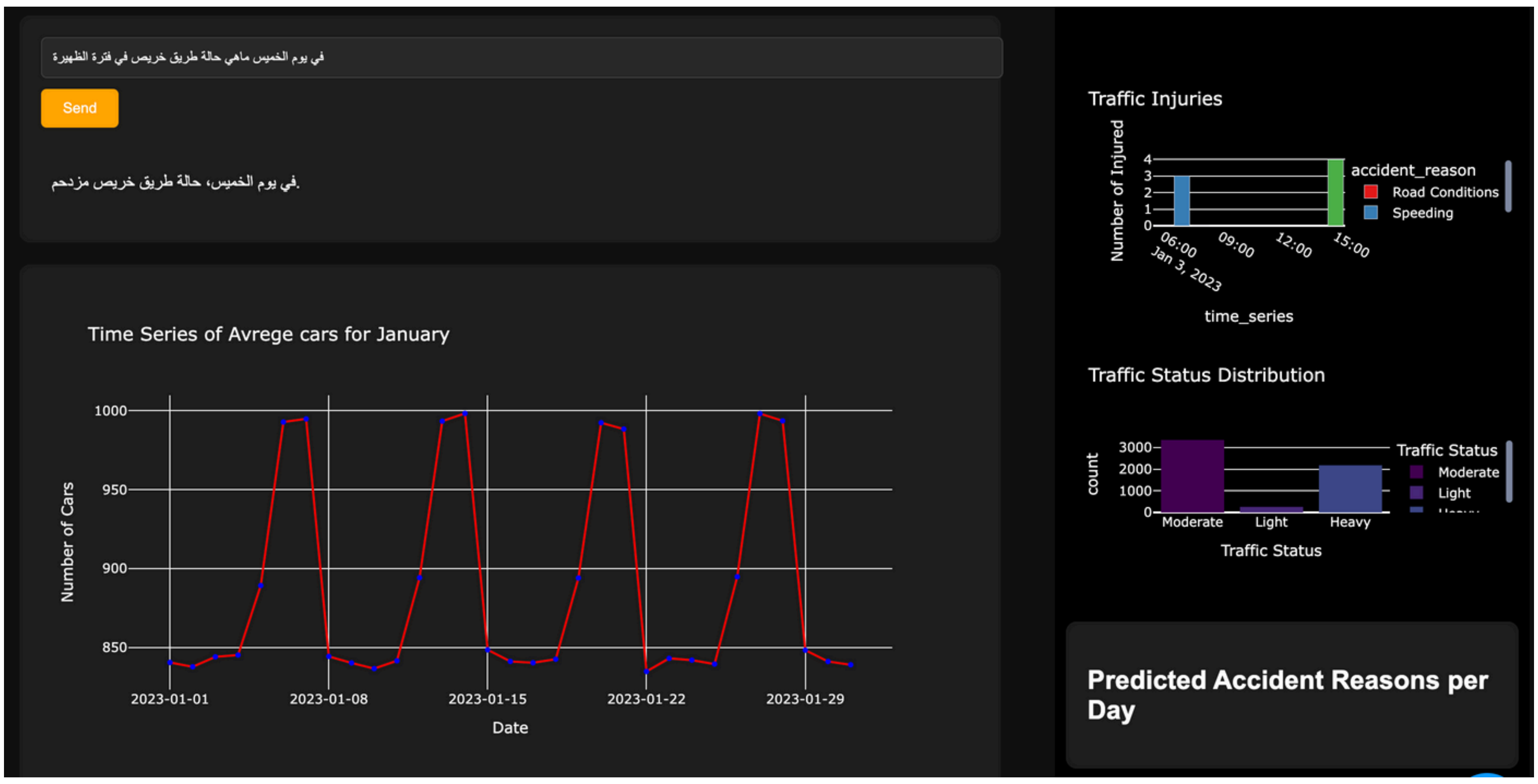


Figure 2. Dashboard

Discussion and Results:

Computer Vision for Vehicle Counting and Speed Monitoring:

- We integrated computer vision technology using a YOLO (You Only Look Once) model to count vehicles in forward and backward lanes and monitor their speeds. This provided real-time insights into traffic density and flow conditions, further improving the accuracy of traffic monitoring and allowing authorities to take action in heavily congested or unsafe areas.

Figure 3. Dashboard

Challenge:

Data Availability:

- A significant challenge in this project was the lack of data that combines traffic conditions and accident reports within the same time series and geographic coordinates. these datasets were often disconnected. Traffic data and accident data were typically recorded independently, with no unified framework to track both simultaneously.

Data Integration:

- Integrating multiple data sources such as accident reports, and real-time traffic feeds was a complex task. Ensuring data consistency and addressing issues like varying coordinate systems (for geographic data), and inconsistent timestamps posed significant hurdles during the integration phase.

Limited Resources:

- Training complex models, especially deep learning architectures like LSTM and RNN, was constrained by limited GPU resources. These models require substantial computational power to handle large datasets and perform efficient training iterations.

Conclusion and Future Work:

Our traffic management solution offers critical insights into current and expected traffic conditions, including congestion patterns, accident locations, timings, and causes. By leveraging predictive analytics and computer vision for vehicle counting and speed monitoring, we empower commuters and traffic authorities with actionable information that enhances decision-making and improves road safety. As we refine our models and expand our capabilities, our goal is to transform traffic management into a data-driven process, facilitating smoother commutes and enhancing overall mobility in urban areas.

Key achievements of the project include:

- Accurate rush hour predictions and identification of congestion hotspots.
- Insights into accident causes and locations, enabling targeted interventions.
- Real-time traffic flow forecasting for smoother commutes.
- Computer vision for vehicle counting and speed monitoring using a YOLO model, providing real-time data on traffic density and flow conditions.
- An interactive map simulator that visualizes traffic conditions in real-time, combined with a language model for answering user queries.

Future Work

As we continue to evolve and enhance our traffic management solution, several areas of improvement and expansion are identified:

Predicting Arrival Time:

- In the future, we aim to extend our predictive models to estimate not only traffic flow but also arrival times for commuters. By factoring in real-time data such as road conditions and incidents, users will receive accurate time estimates for their journeys.

User Behavior Analysis:

- To further refine our solution, analyzing user behavior such as route choices, departure times, and responses to congestion alerts will provide a deeper understanding of commuting patterns. This will enable more personalized traffic predictions and suggestions.

Real-Time Data Updates:

- While our system currently integrates real-time data, we aim to improve the frequency and granularity of these updates. Continuous data streaming from various sources, including Internet of Things (IoT) devices such as traffic sensors and vehicle telematics, will enhance the accuracy and timeliness of our predictions.

Scalability to Other Regions:

- Expanding our solution to cover additional regions beyond Khurais Road will be a key goal. By scaling up the dataset and adapting the models, we aim to provide predictive traffic insights for Riyadh and all cities of the Kingdom of Saudi Arabia

Reference:

[1] Shin, H.(2019). Time Series Forecasting on Car Accidents in Korea Using Auto-Regressive Integrated Moving Average Model. Journal of Convergence for Information Technology.

[2] Jinga, D. Yaoa, E. Chen, R.(2023). Moving Characteristics Analysis of Mixed Traffic Flow of CAVs and HVs around Accident Zones. Physica A Statistical Mechanics and its Applications.

[3] Feng, G.(2021). Common Python Data Analysis Method Based on Deep Learning. Journal of Physics: Conference Series.

[4] Getahun, K.(2021). Time series modeling of road trafc accidents in Amhara Region. Journal of Big Data.

Team:

Alyaa Bajaber

Khuzama Alsenan

Anas Alwohaib

Saleh Alanazi

GROUP 4