DATA SCIENCE CANVAS

Team: Data Dorks | Project name: TADA: Traffic Analysis DArmstadt | Iteration: 2 | Date: 01.02.24

Problem Statement

Congested roads and slow-moving traffic are problems faced by commuters in Darmstadt. The existing challenges necessitate innovative solutions to optimize traffic flow, reduce environmental impacts, and enhance overall transportation efficiency in the city.

Solution

Through our application, historical data can be used to predict future traffic flow. As such, this application can provide insights for traffic department staff and help them improve urban transportation in various ways.
For instance, critical junctions can be identified early on before congestion occurs, traffic can be analyzed over long-term for urban planning and to improve traffic concepts, redirections for construction sites or events can be planned, and traffic light signals can be balanced which leads to less slow-moving traffic.

Users & Use

Our application is aimed at control center agents monitoring and controlling traffic in Darmstadt. Moreover, it offers support in recognizing and planning critical situations. This relieves and supports them in their work.
Indirectly, road users and residents of Darmstadt will benefit the most, as they will be able to move around in the city more quickly and safely and plan routes reliably.

Data Acquisition

"Datenplattform Darmstadt" provides traffic data in CSV format for one month, an overview map of all intersections, a detailed map for each intersection showing the location of sensors, and a JSON file containing sensor configuration data.

An extended dataset of two years was made available after a mail request.

Analytics Formulation

Input: Traffic sensor data from "Datenplattform Darmstadt"
Output: Number of cars per hour (past / present / predicted)
Methodology: Machine learning models: Prophet, Random Forest, LightGBM, CatBoost.
Regression for time-series data was implemented.

Modeling

We performed time-series data prediction for the number of cars per hour. For this purpose, we evaluated the following models: Prophet, LightGBM, CatBoost, and Random Forest. Based on the metrics, Random Forest was deemed as best to use for our application.
After prediction, data was binned to different congestion categories (low / high / medium).

Data Preparation

First, we had to gain deep understanding of the dataset (rudimentary / missing explanation), then data cleansing was performed e.g. intersections that are out of service due to roadworks, merging data from various intersections, connecting which sensors are installed where and at which intersection, and interpolation for missing values.

Evaluation

Evaluation was carried out using a combination of error metrics: MAE, MSE. RMSE, R^2, SMAPE, and additionally, a confusion matrix (for the congestion categories). This approach allows for a nuanced evaluation of the model's accuracy, error magnitude, and predictive performance.

Success Criteria

Objective: the model performs well with respect to the evaluation metrics. It identifies congestion in historical data and provides reliable predictions.
Subjective: users can easily recognize critical moments in the current and future traffic situation. Furthermore, there is a user interface representation that shows the traffic in different colors, based on the amount of congestion, and provides an accurate traffic prediction. The predictions can keep up with and outrun predictions from other tools such as Google Maps and Waze.

Constraints

Updating the model with real time data is a constraint. Unpredictability exists due to regional factors like strikes. Also, there are variations in the sensors. Sensor hardware failure leads to imputation of data which leads to falsification of the prediction. In addition, there is ambiguity in the information about the dataset, sensors, and their positions.

Business Value

A regulated flow of traffic increases the safety and well-being of all road users and residents.
Unlike in other services, the end user is a control center agent from the road authority of Darmstadt. Monitoring and prediction are key in planning the future traffic flow. Agents can determine other ways to organize traffic and difficult traffic conditions beforehand. This can be applied to road planning and instantly adjusting traffic lights.
Utilizing the city sensor data allows precise recording of the traffic situation which results in more accurate future predictions.

MVP

The MVP covers a small major area of Darmstadt: five junctions around the city center and their corresponding traffic flow. It demonstrates the amount of present and future congestion. The traffic situation on routes between these intersections is also displayed.

Key Actors

Internal: traffic sensor data
Customer stakeholders: traffic department of Darmstadt (and other cities), vehicular commuters, general public
External: employees from the mobility office or road planning department of Darmstadt.

Technology stack

Backend: Python and Python libraries (including numpy, pandas, sklearn, prophet, catboost, lightgbm, matplotlib)
Frontend: Figma, Anvil, Python
Project Management: Trello, Google Meet, Telegram
Development: Github, Jupyter Notebook, Google Colab, Visual Studio Code