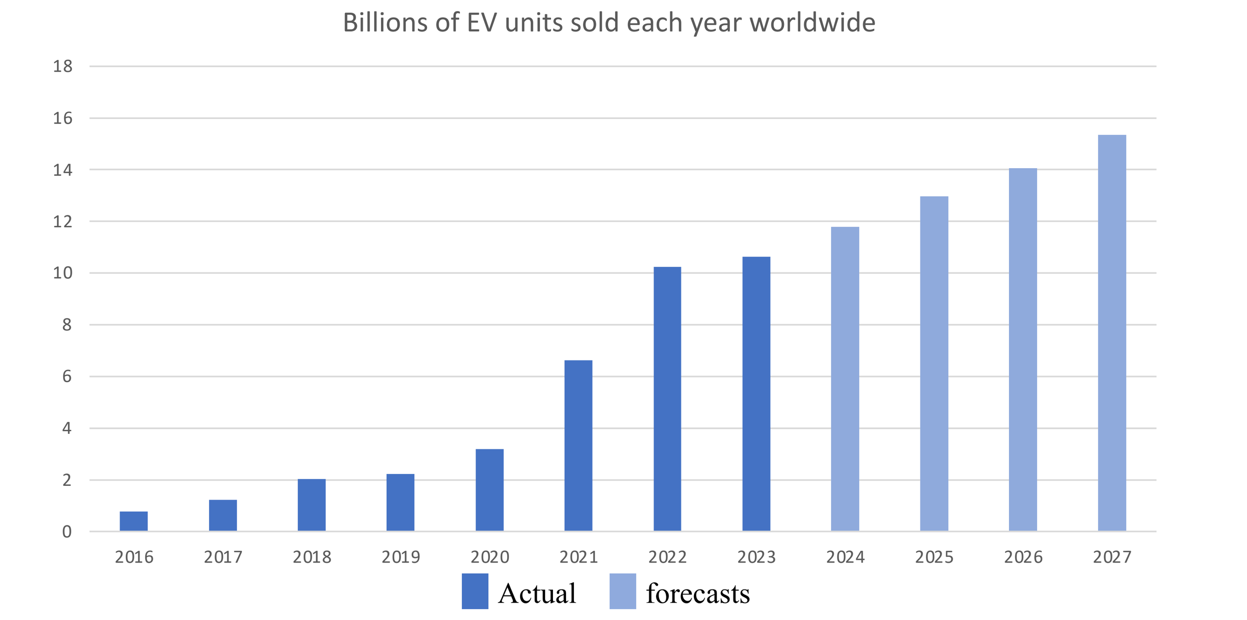
Strength in Numbers

Assessing customer EV readiness using Federated Learning

■ Students: Mattia Ravasio, Paul Theron ■ MIT Advisor: Prof. Retsef Levi ■ BMW Group: Dr. Maka Karalashvili

1. Background and use case

The world is seeing an increasing market trend in EV sales



Source: https://www.statista.com/outlook/mmo/electric-vehicles/worldwide

- The BMW Group corporate strategy team is interested in assessing **how electric vehicle ready are certain traditional petrol engine drivers**. To do this they built a simulation.
- To improve their current simulation, they needed a better way to model the charging behavior. Thus we delivered a **machine learning model** that would **predict charging events at the end of driving sessions.**
- The BMW innovation and emerging technologies team was interested in **assessing the potential of Federated Learning**, a privacy compliant way to train models .
- Seeing the potential in the simulation of the corporate strategy team, we decided to test the federated learning technology by **comparing a centralized and distributed modelling approach on the simulation use case.**

Simulation steps

1. Select a given petrol engine driver
2. Select an EV model for the customer
3. Assess when the customer needs to charge

Final deliverables

Federated model
For the innovation and emerging technologies team

Centralized model
For the simulation

2. Data

The data we used can be divided into 3 parts:

Vehicles
- General information about the vehicle
- Geographical context of Home and important locations
- Customer segmentation and clustering

Drives
- General information about the drive
- Time/ Geographical context of the drive

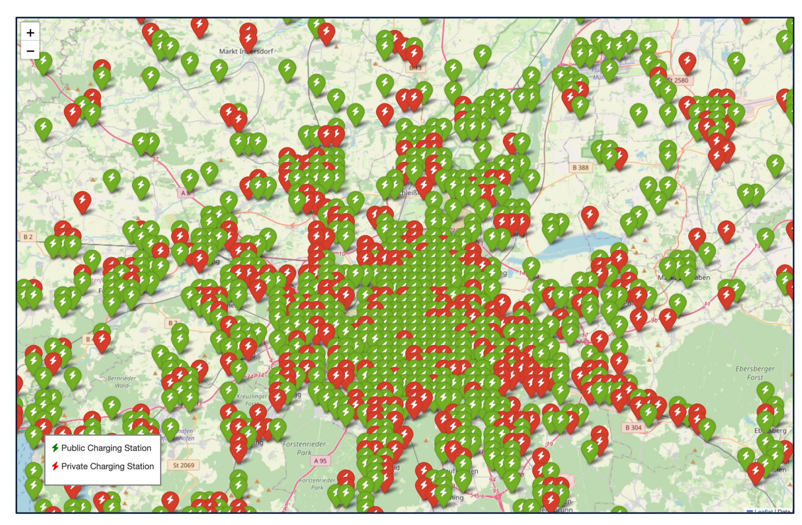
Charges
- General information about the charge
- Time/ Geographical context of the charge
- Previous/following drive information

Data sample selection:

- Sample extracted from vehicles situated in the **Munich geographical region**.
- Inclusion criteria: Drives conducted after **December 2021**.
- Picked a **subset of 110 drivers**, specifically those demonstrating the highest drive frequency within this area.
  - ➢ **Aggregate of 1 million drives included in the study.**
  - ➢ **Recorded data exhibits a 10% imbalance rate.**

3. Feature engineering

Geo-Spatial Feature engineering



Example of such features:
- Is stop in public location,
- How far is the closest public charging location I've previously charged in

Time-series Feature engineering

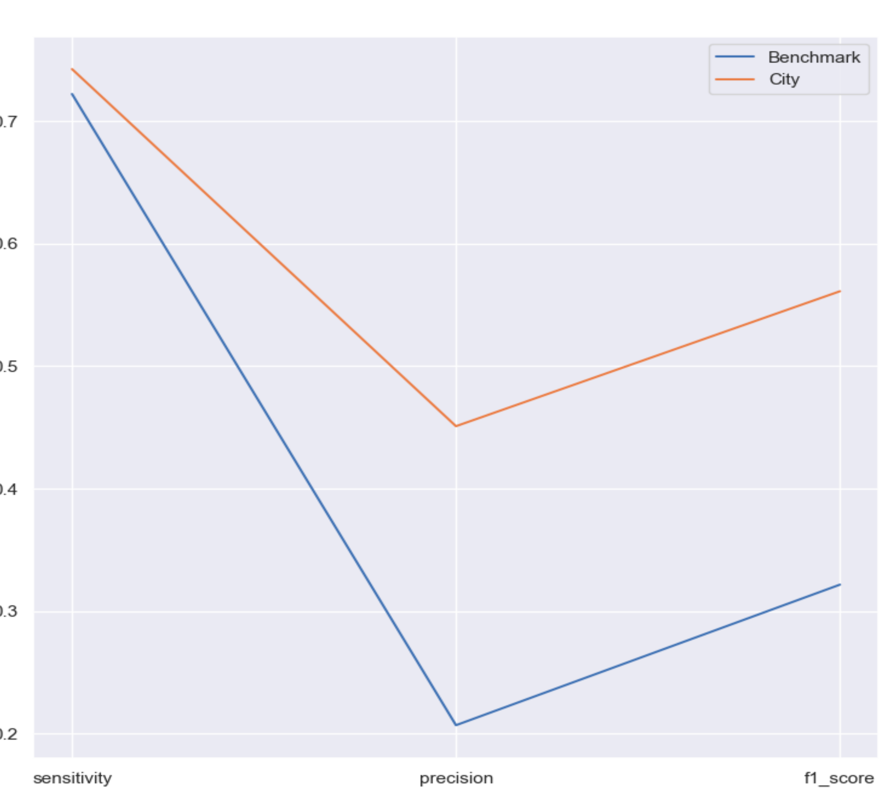
- Is my charge level low compared to my usual charging habits ?
- Is my stop in one of dominant charging location at the time of the day

4. Centralized model

Model pipeline

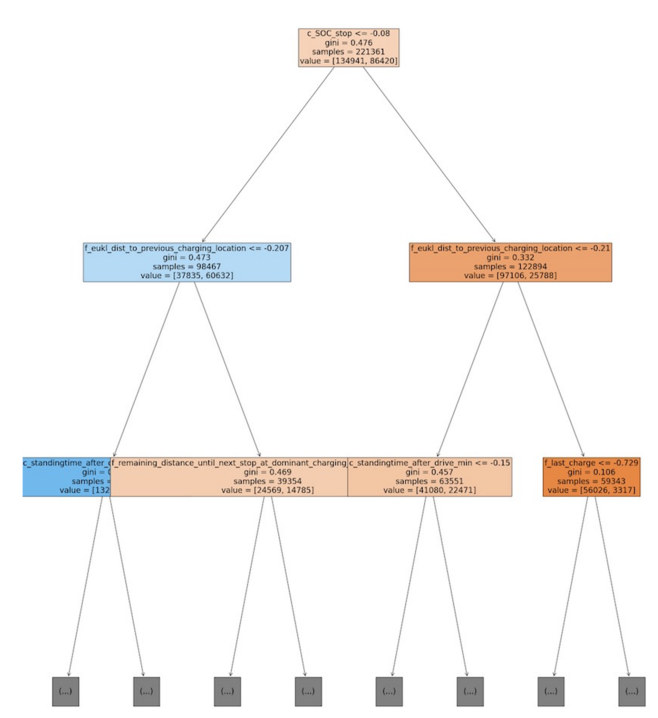
Fetching local data from AWS Athena ➤ Data quality check and feature engineering ➤ Correlated features filter ➤ Majority class under-sampling ➤ Training of the model on local data

Scalability evaluation



Difference in the performance of the model on the same region it was trained on compared to a different region

Model interpretability

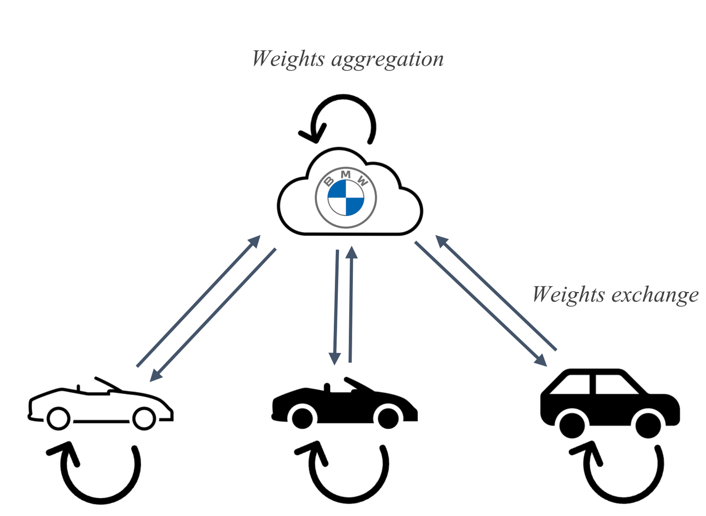


Decision tree to interpret the model results

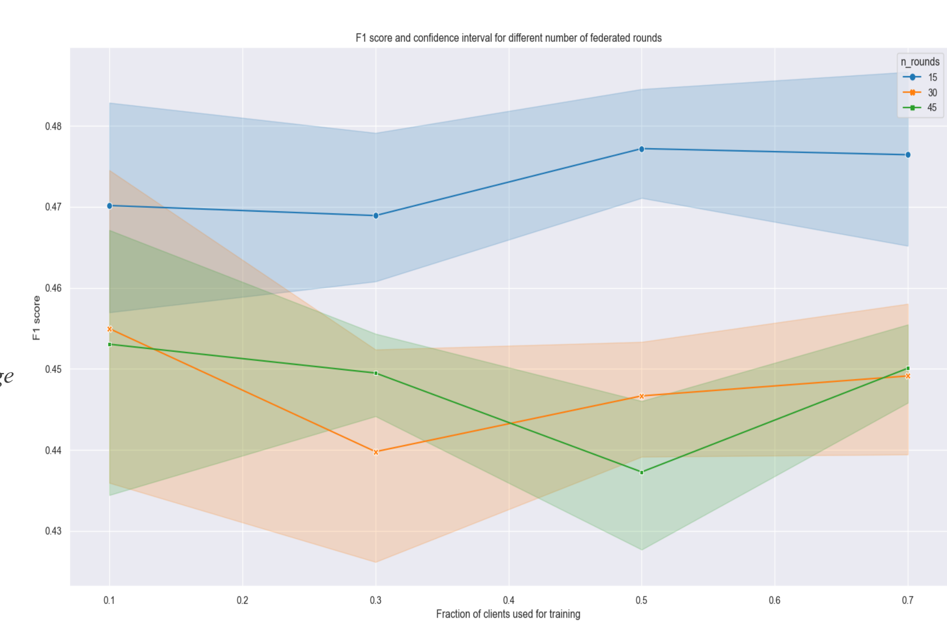
5. Federated model

The necessity for high-performing models that **safeguard user privacy** is vital for BMW's future, in part due to the forthcoming European GDPR. Addressing this, the project employed Federated Learning to predict user charging behavior, examining if it could **match the efficacy** of a centralized model without accessing private data, and further **exploring any limitations.**

- Explored and understood the influence of number of federated rounds, aggregation strategies, number of clients, non IID data
- Compared FL results to a centralized baseline, trained with and without user's private features



Federated Learning framework



Influence of number of federated rounds

6. Results

63% enhancement on the baseline facilitating more accurate customer targeting.

Bridged **77%** of the performance divide between a private centralized model and a traditional centralized model.

Provided valuable insights regarding trade-offs, fostering experimental opportunities in Federated Learning for the innovation team.

Performance of the best model