Siim Esko
International Project Development Lead
siim.esko@positium.com

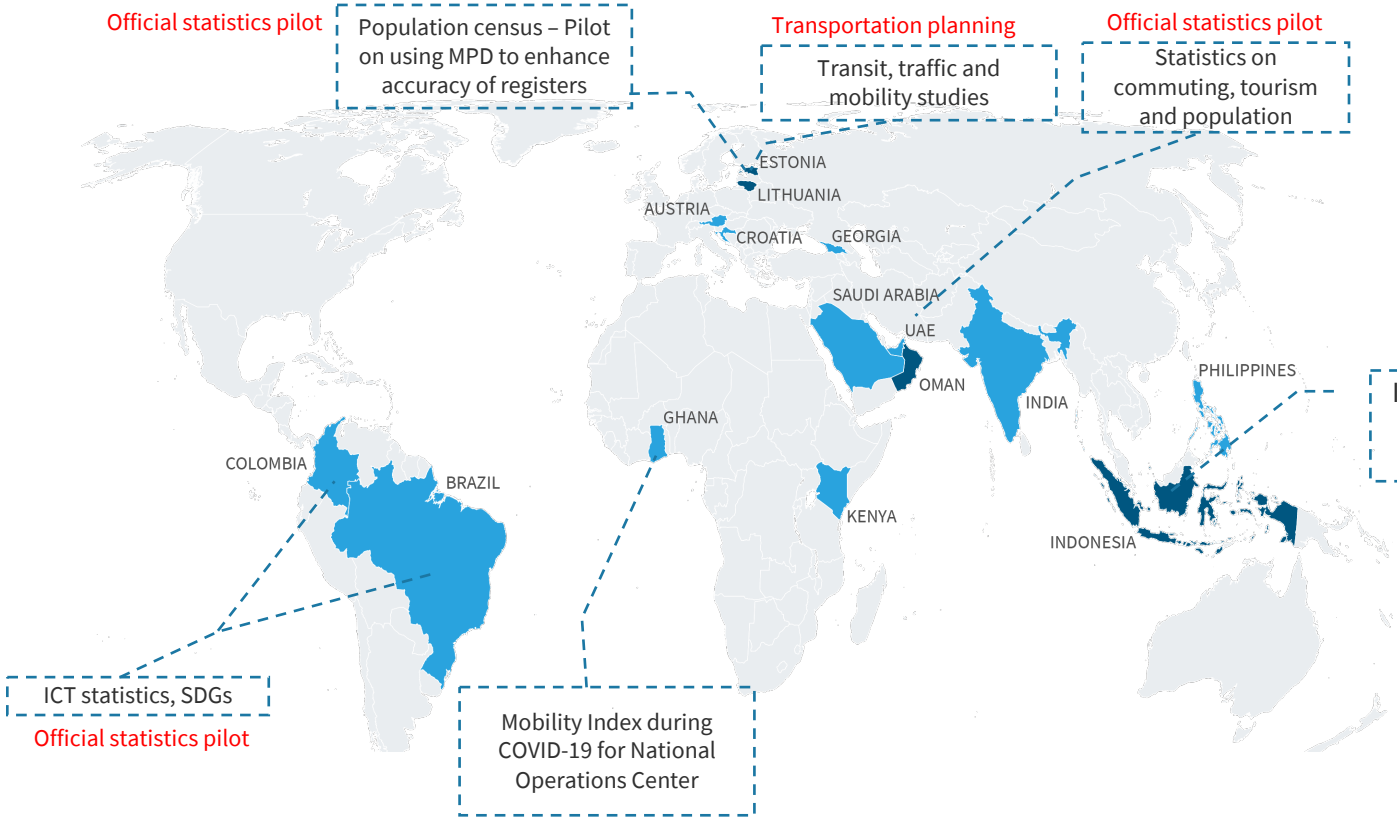
Modelling: Integration of Mobile Positioning Data with Other Datasets

Big Data integration with data for public transport and traffic counter data

UN Big Data Regional Hub for Africa

Webinar of the UN Regional Hub for Africa
11 April 2024

Mobile Positioning Data Projects Globally – Positium



Key stats:
20+ mobile operators
15+ countries
300+ projects
20+ use cases
1 Billion+ population

Inbound, outbound and domestic tourism methodology and technology

Official statistics production

Fundamental questions in urban planning and transport

	Where are people?	Where do people move?
Long-term	Where people live and work	Home-work commuting
Short-term	Where people spend the day and night	Daily movement
Micro-term	Where people spend every 15 minutes	15-minute movement
Data sources	Registries & census (static) Surveys (static)	Traffic loops PT smart cards Mobile apps

Mobile positioning data (MPD)

MPD - the best dataset to connect dynamic population and origin-destination data on a strategic scale

Transport use case

Urban transport analysis: Tartu City Public Transport

Client: Tartu City Government

Objective: Big data-driven public transport remodelling

Data sources: 20 layers of data, including mobile positioning data, registries, public transportation check-ins, land use data, survey.

Result

From data collection to start of operation of new bus lines in 2 years, resulting in +15% increased trips and 86% satisfaction rate

Key to Getting Data on Journeys – MPD for Mobility Demand

Challenge: Tackle data scarcity with big data as the core mobility dataset

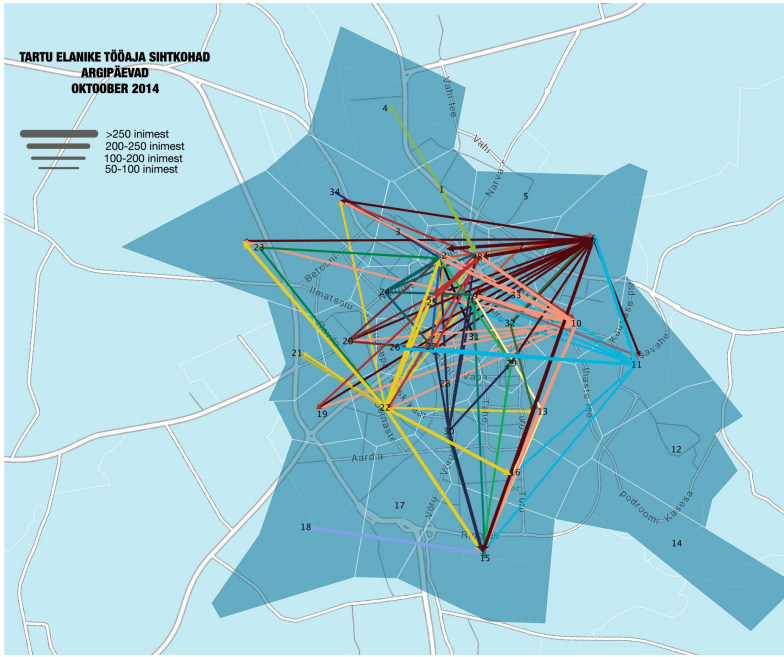
Data sources:

1. **Mobile positioning data** was the main data source for mobility demand, and for home and workplace figures

Tartu City

Population: 100,000

Movement combinations: 6,000,000



Origin-Destination Matrices between city districts from MPD

Key to Solving Data Scarcity – Data Integration

Challenge: Tackle data scarcity by using 20 layers of data in an integrated way

Data sources: Mobile positioning data is used to compare to:

2. **Registries** to validate population data (home, work, school),
3. **Public transportation check-ins** to measure unmet demand,
4. **Land use data layers** to characterize transport zones,
5. **Surveys** to add qualitative aspects,
6. **Journey planner application** to gather user feedback about old vs new network

Data layers integrated through location attributes

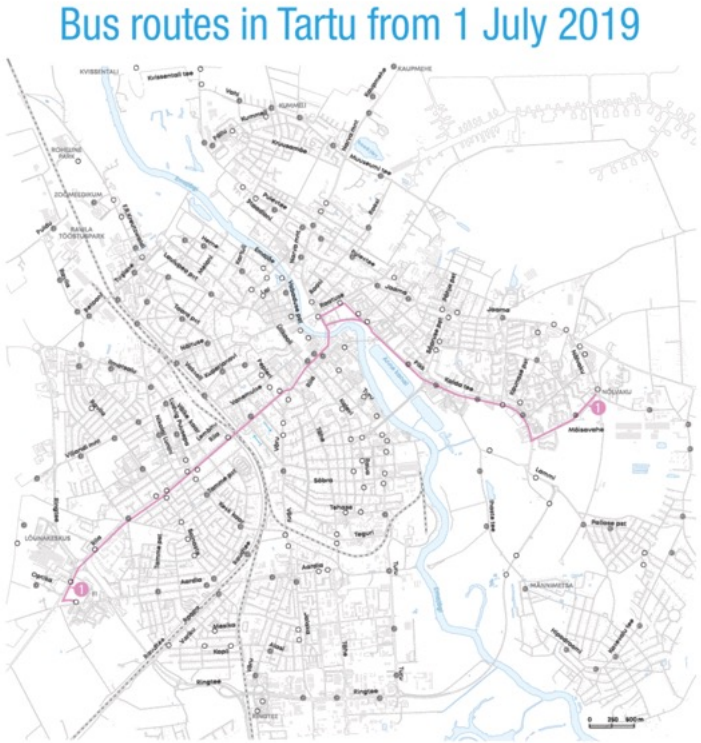
Key to a Time-Efficient Process – Data-First Approach

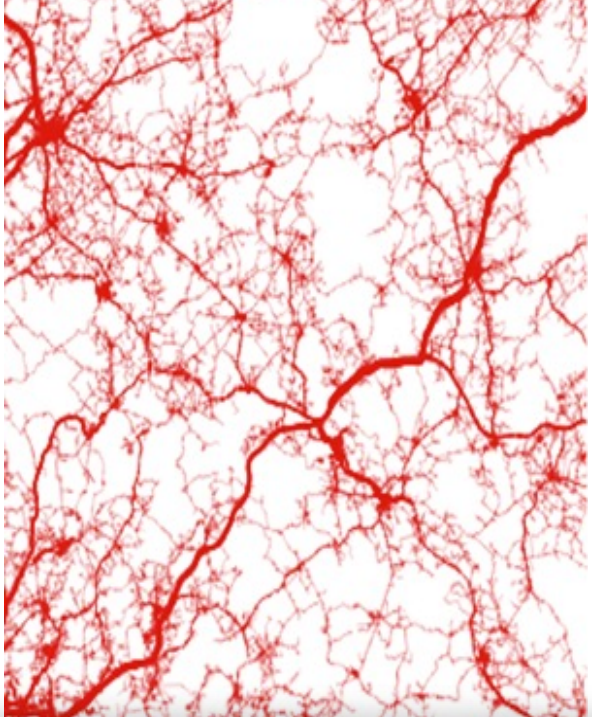
Challenge: 2 years from data collection to new routes in operation

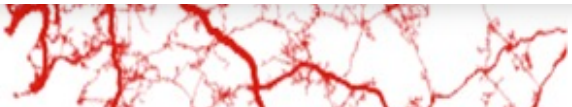
With a data-driven process, followed up with qualitative feedback, it was possible to achieve results:

- **2 years from data collection to starting** the new bus line network
- **30% more people** use public transport regularly
- **86% are satisfied** with the new network, across all age groups

And after 3 more years, the lines were expanded to nearby municipalities with a repeat of the process.







AADT use case

Estonian Transport Administration road maintenance plan with AADT based on MPD and counter data

Challenge: Independent tool to measure road traffic on all roads, not only those with permanent counters

The Estonian Transport Administration, in collaboration with us and Telia, a leading Mobile Network Operator (MNO) in Estonia, has developed an innovative solution to estimate the Annual Average Daily Traffic (AADT) for the entire region's roads.

Machine learning model was applied as an independent tool for AADT calculation, combining the strengths of both datasets.

Transportation planners rely on AADT numbers to make informed decisions about infrastructure investments and improvements for more than 4000 road segments.

“From the beginning of our partnership, the team at Positium demonstrated a strong understanding of our objectives and challenges, and their expertise in combining existing counter data with mobile positioning data resulted in a cost-effective, efficient, and accurate solution.”

– Reimo Tarkiainen, Head of Data and Analysis Dept, Transport Administration

1.5M
OD pairs from mobile network

100+ permanent counters
and 1100 temporary counter locations

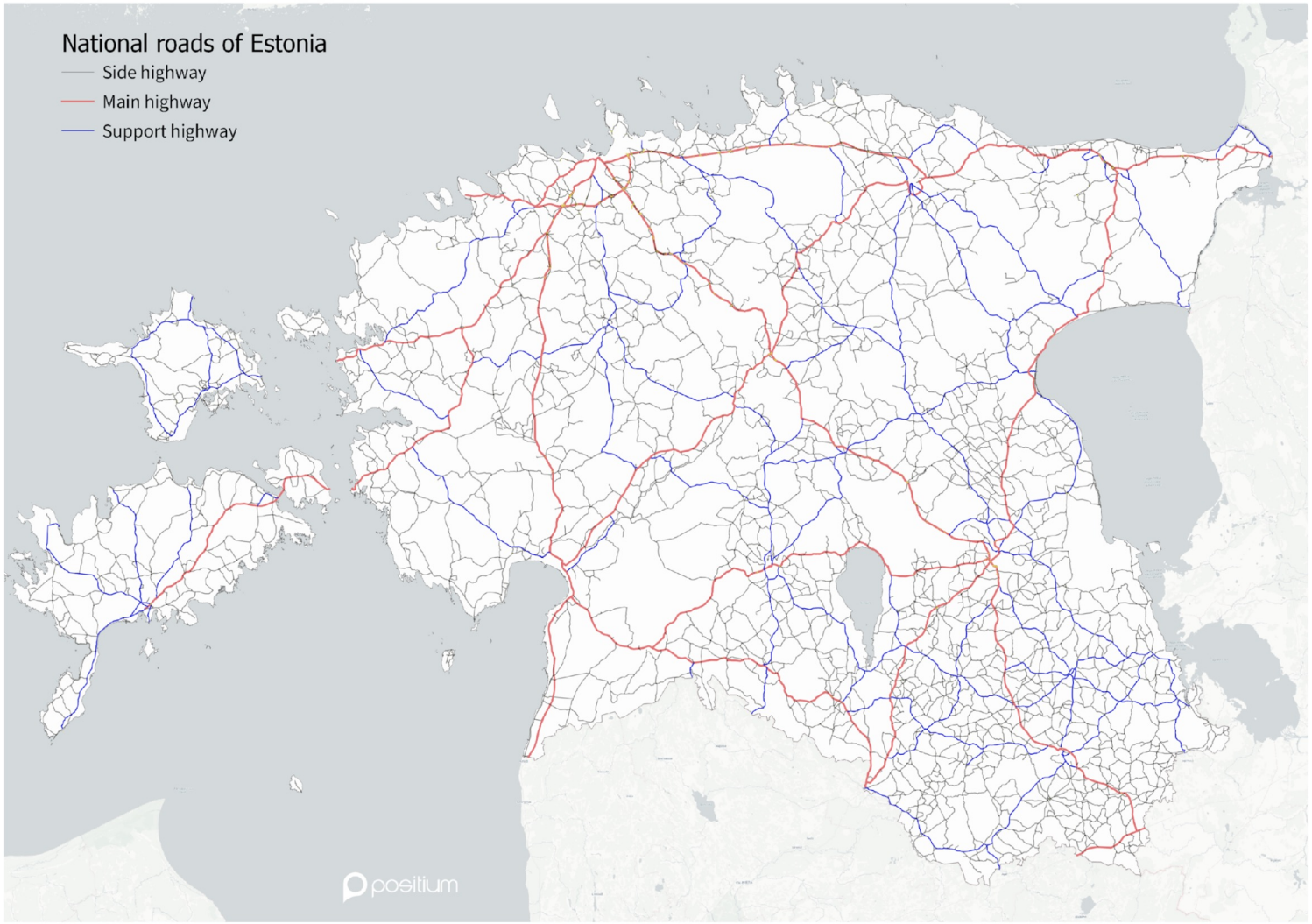
4000
road segments

DATA FROM MNO AND SENSORS

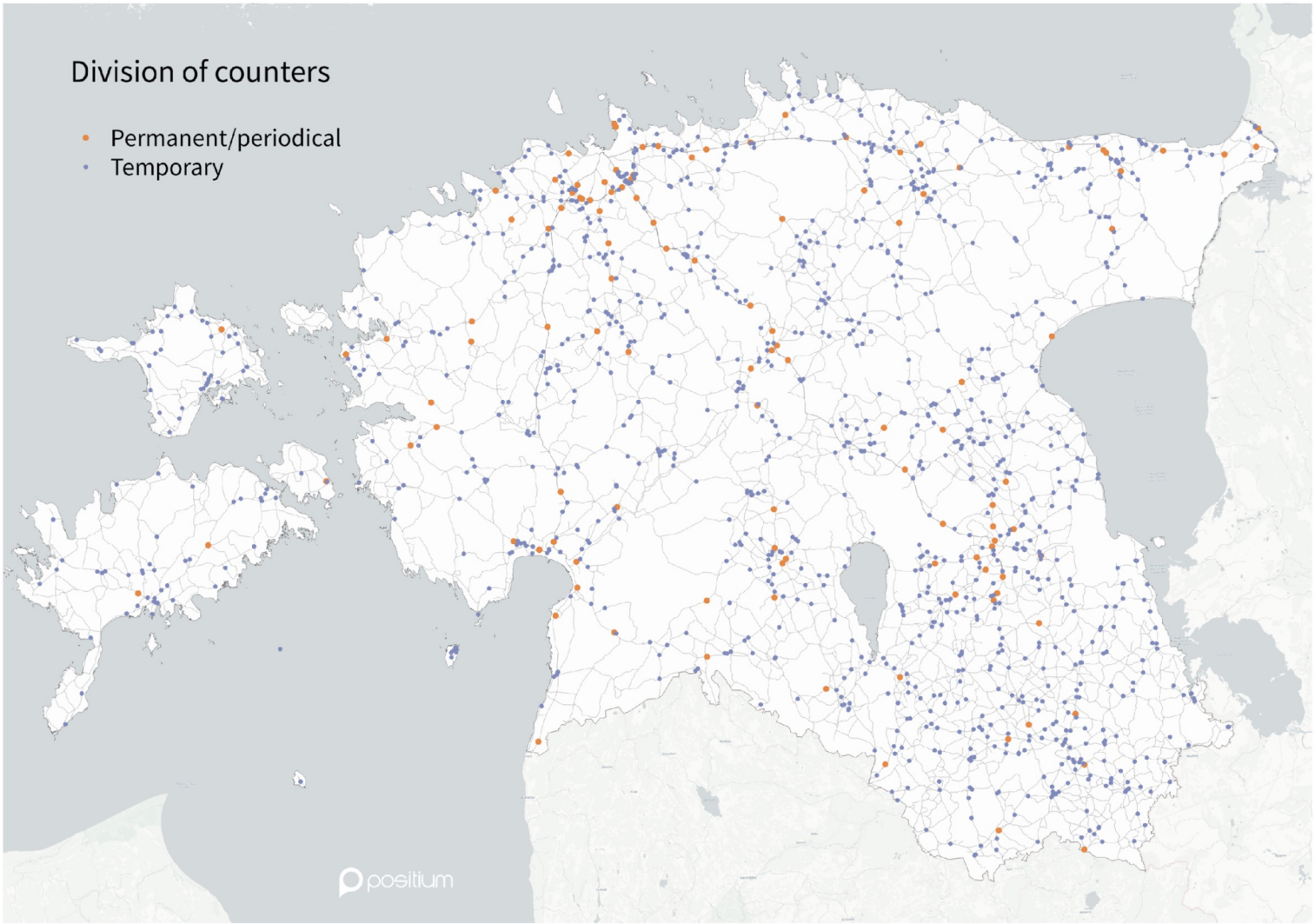
AADT FOR THE COUNTRY

METHODOLOGY DOCUMENT

QUALITY ASSESSMENT


National roads of Estonia
Side highway
Main highway
Support highway
positium

positium


Division of counters
Permanent/periodical
Temporary
positium

ADT delivery approach – 4 to 5 months

1. Data Collection

~ 6 weeks

2. ML Model Development

~ 5 weeks

3. ADT Estimation and Validation

~ 4 weeks

4. Final Results

~ 1-2 weeks

ACTIVITIES

1. Collect and store the data
2. Data qualification into types
3. Input data cleansing:
   - OD matrices
   - Counter locations
   - Counter results
   - Roads topology
4. Data prioritization methodology
5. Combination and grouping of different data types

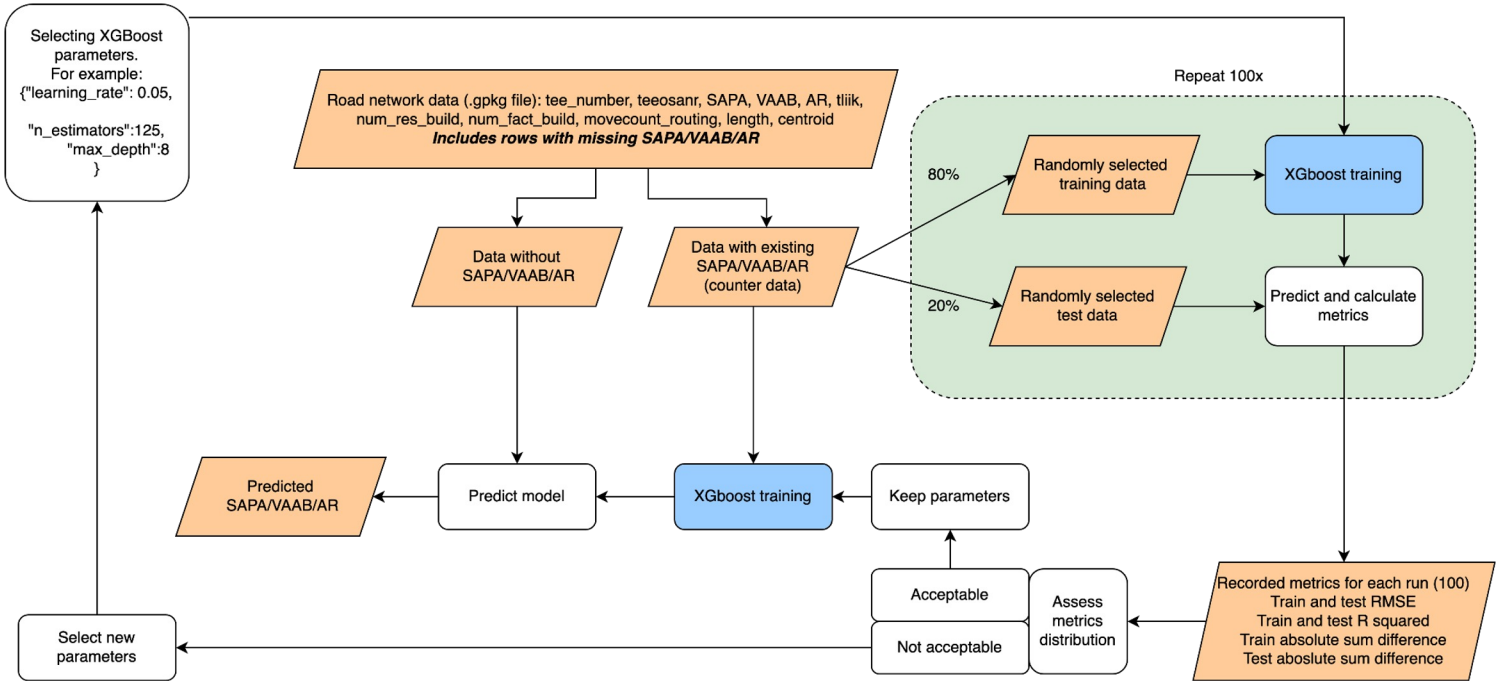
1. Develop the approach for utilization of sensors data, OD matrices and survey data in one model
2. Find ADT values for roads with all available data
3. Test various approaches to achieve the best quality and granularity of the results

> XGBoost

1. Iterative modelling and analysis generation
2. Validation of the results

1. Result delivery
2. Tool delivery
3. Assist with results analysis and using results for business purposes
4. Comprehensive training

Modelling

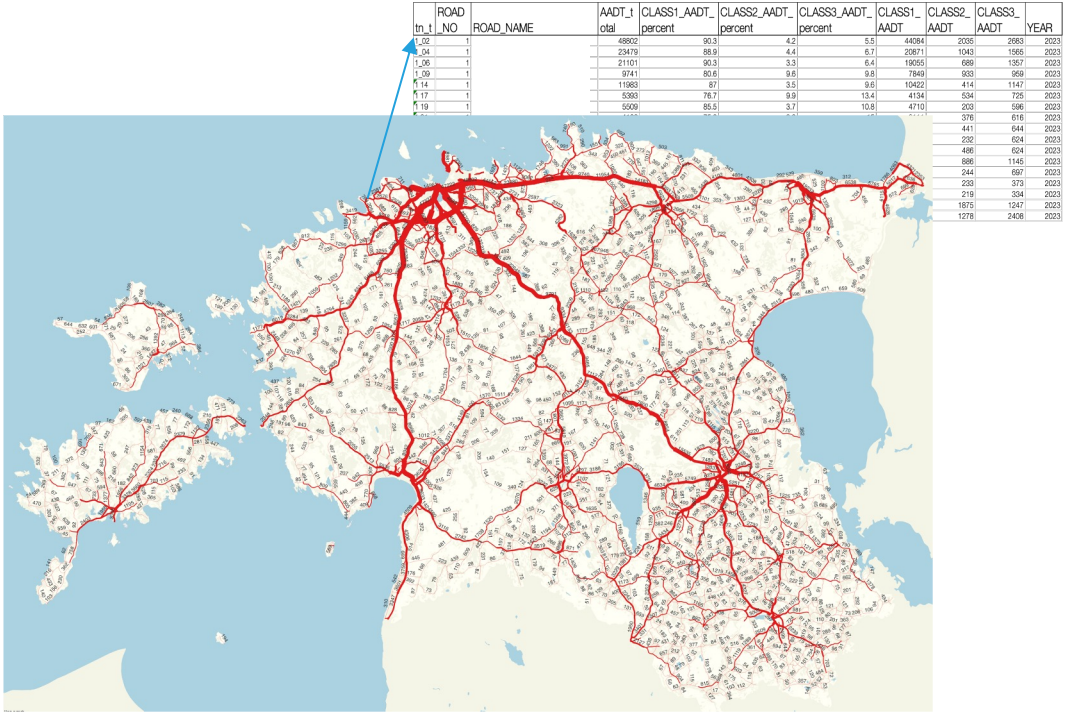

Selecting XGBoost parameters. For example: {"learning_rate": 0.05, "n_estimators":125, "max_depth":8 }
Road network data (.gpkg file): tee_number, teeosanr, SAPA, VAAB, AR, tliik, num_res_build, num_fact_build, movecount_routing, length, centroid
Includes rows with missing SAPA/VAAB/AR
Repeat 100x
80%
Randomly selected training data
XGboost training
Data without SAPA/VAAB/AR
Data with existing SAPA/VAAB/AR (counter data)
20%
Randomly selected test data
Predict and calculate metrics
Predicted SAPA/VAAB/AR
Predict model
XGboost training
Keep parameters
Acceptable
Not acceptable
Assess metrics distribution
Recorded metrics for each run (100)
Train and test RMSE
Train and test R squared
Train absolute sum difference
Test aboslute sum difference
Select new parameters

Result - Total AADT

Deliverables:

1. Results on AADT divided into vehicle classes for all roads
2. Self-service tool:

- Project methodology on 45 pages
- 4 Jupyter notebooks for Quality Assurance purposes
- 8 Python scripts



tn_t	ROAD_NO	ROAD_NAME	AADT_total	CLASS1_AADT_percent	CLASS2_AADT_percent	CLASS3_AADT_percent	CLASS1_AADT	CLASS2_AADT	CLASS3_AADT	YEAR
1_02	1		46802	90.3	4.2	5.5	44084	2035	2683	2023
1_04	1		23479	88.9	4.4	6.7	20871	1043	1565	2023
1_06	1		21101	90.3	3.3	6.4	19055	689	1357	2023
1_09	1		9741	80.6	9.6	9.8	7849	933	959	2023
1_14	1		11983	87	3.5	9.6	10422	414	1147	2023
1_17	1		5393	76.7	9.9	13.4	4134	534	725	2023
1_19	1		5009	85.5	3.7	10.8	4710	203	596	2023
								376	616	2023
								441	644	2023
								232	624	2023
								486	624	2023
								886	1145	2023
								244	697	2023
								233	373	2023
								219	334	2023
								1875	1247	2023
								1278	2408	2023

Evaluation Criteria

- ✓ Allowed difference on main roads +/- 8 %.
- ✓ Allowed difference on supporting roads +/- 15 %.
- ✓ Allowed difference on side roads +/- 30%

“The estimation of Annual Average Daily Traffic (AADT) using machine learning and mobile positioning data has been extremely successful and has significantly improved our ability to make informed decisions regarding infrastructure investments and improvements.”
– Reimo Tarkiainen, Head of Data and Analytics, Transport Administration

positium

Average Hourly Traffic example from a European city

Key Takeaways

Recap: Tartu City's Transformation

Key Points:

- Successful integration of MPD led to an optimized bus network.
- Increased ridership, improved efficiency, and user satisfaction.
- Demonstration of fast and effective urban transport remodelling.

2 years
to
86% satisfaction
rate

Recap: Innovations in AADT Estimation

Key Points:

- ML and MPD integration for comprehensive traffic analysis.
- Achieved high accuracy in AADT across diverse road types.
- Cost-effective and efficient approach for large-scale application.

Extend 100
counters to 4000
road segments
with MPD

Thank you!

We strive for the future where every country in the world benefits from mobile positioning data for the good of society

For further information, please contact:

Siim Esko: siim.esko@positium.com

UN Big Data Regional Hub for Africa