Robust Journey Planner

Large-Scale data science for real world data

COM-490

Our Startup

"Travel Your Way, Every Day."

What our planner looks like

Choose your Trip information

Choose your city

Select city...

Choose your departure stop

Select departure stop...

Choose your arrival stop

Select arrival stop...

Departure Day

2024/05/27

Departure Time

16:02
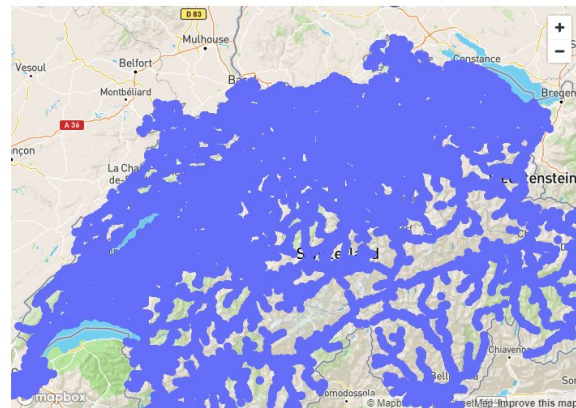
Arrival Time

19:02

Whats the weather ?

- ☀️
- ☁️
- ⛈️

Temperature in °C

15

−10   30

Plan my trip

Temporaire

Robust Journey Planner

What our planner looks like

Speaker

ROBUST JOURNEY PLANNER

Choose your Trip information

Choose your city

Lausanne gare 10km radius

Choose your departure stop

Prilly-Chasseur

Choose your arrival stop

Lausanne, Ouchy-Olympique

Departure Day

2024/05/27

Departure Time

16:04

Arrival Time

19:04

Whats the weather ?

Temperature in °C

14

−10   30

Plan my trip

Temporaire

Robust Journey Planner

Your trip from Prilly-Chasseur to Lausanne, Ouchy-Olympique is planned for `2024-05-27` at `16:04:00` and you will arrive at `19:04:00` with a ☁ weather.

The probability of success for each change is : [0.75, 0.65]

The probability of success for the trip is : `0.48750000000000004`

The probability of success for each change (with predicted delay taken into account) is : [1, 1]

The probability of success for the trip (with predicted delay taken into account) is : `1`

BARGHORN Jérémy

ROBUST JOURNEY PLANNER

Data models used to represent the infrastructure

Data Preprocessing and Graph Construction

|	trip_id	arrival_time	departure_time	stop_id	stop_sequence	route_id	stop_name	stop_lat	stop_lon	route_desc	arrival_stop_id	departure_stop_id	arrival_time_seconds	departure_time_seconds
444152	1.TA.91-1R-Y-j24-1.1.H	18:57:00	19:01:00	8501120:0:1	2	91-1R-Y-j24-1	Lausanne	46.516775	6.629513	EXT	8501120:0:1	NaN	68220	68460
450005	1.TA.91-2E-Y-j24-1.17.H	19:45:00	19:45:00	8501120:0:6	1	91-2E-Y-j24-1	Lausanne	46.516521	6.629019	TGV	8501120:0:6	NaN	71100	71100
552697	1.TA.91-2H-Y-j24-1.2.H	13:43:00	13:43:00	8501120:0:4	1	91-2H-Y-j24-1	Lausanne	46.516669	6.629055	IC	8501120:0:4	NaN	49380	49380
561787	1.TA.91-37-Y-j24-1.1.H	18:23:00	18:23:00	8501120:0:7	1	91-37-Y-j24-1	Lausanne	46.516459	6.629001	TGV	8501120:0:7	NaN	66180	66180
473577	1.TA.91-3D-Y-j24-1.10.H	08:08:00	08:14:00	8501118:0:4	3	91-3D-Y-j24-1	Renens VD	46.536355	6.581067	IR	8501118:0:4	8501120:0:5	29280	29640

Data Preprocessing Steps:
– Recent Timetable
– **Weekday** Services
– Correlate **stop** with **trips**
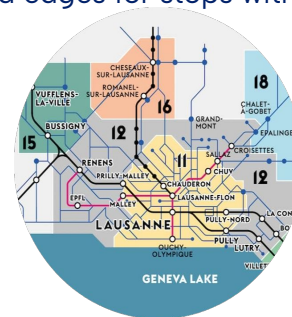– Business hours
– Stop pairs within **walking distance**
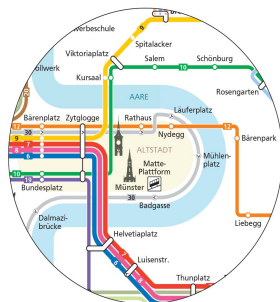
Reproducible Results
– All geo_shapes **objectid**
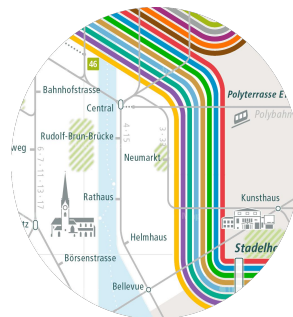
Graph Construction
– Each stop represents a **node**
– **Edges** between consecutive stops based on observed trip sequences
– We set **edge weights** as the travel time between two nodes
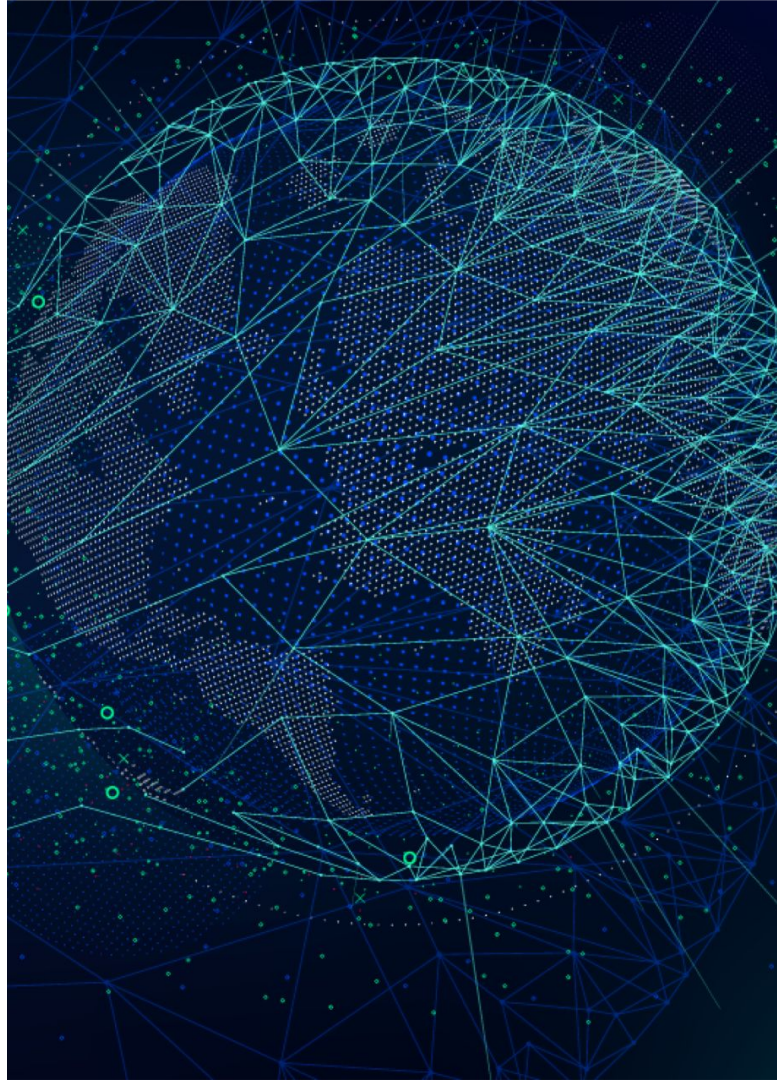– Added edges for stops within walk distance

Data for the Machine Learning model

Train data:
1% of istdaten trips in 2023

stop_id	arrival_time	lat/lon	avg_delay	stddev_delay	temerature	precipatations	delay_label
string	long	long	long	long	long	long	long

Inputs

Outputs

Processing

Data coming from the modified Djikstra algorithm :
– stop_id
– arrival_time
– lat/lon

Data provided by the user :
– temperature
– precipitations

Data specific to geoshape :
– avg_delay
– stddev_delay

Discrete predicted delay between 0 and 45 minutes

Given the transition time at a stop and the predicted delay get the **probability of a successful change**.

ROBUST JOURNEY PLANNER
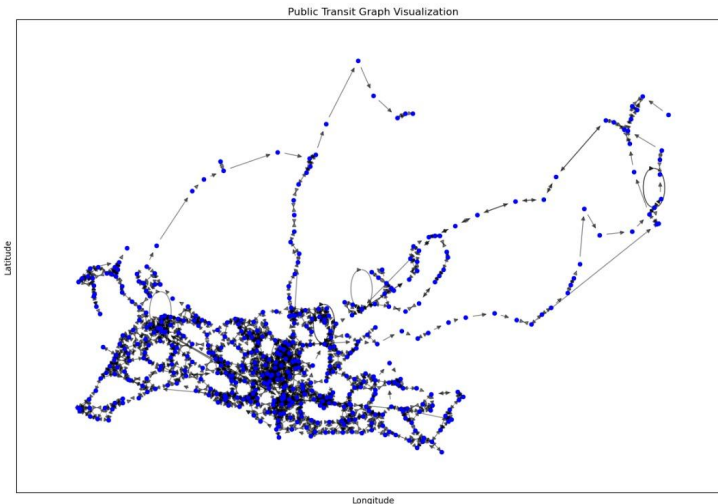
Speaker

ROBUST JOURNEY PLANNER

What is our routing algorithm ?

Graph for the Routing Algorithm

- **Create a node for each stop_id**

- **Edges are each connections**

- **Edge's weights are difference of arrival and departure times.**

- **Walking edges for stops within 500 m**

Public Transit Graph Visualization

Graph of lausanne area

Modified Dijkstra's Algorithm:

- **Output Multiple Path (Top-K Paths) given an arrival time and 2 stops**

- **Priority Queue Based on Departure Time**

- **Time Constraints**   `latest_arrival_time`
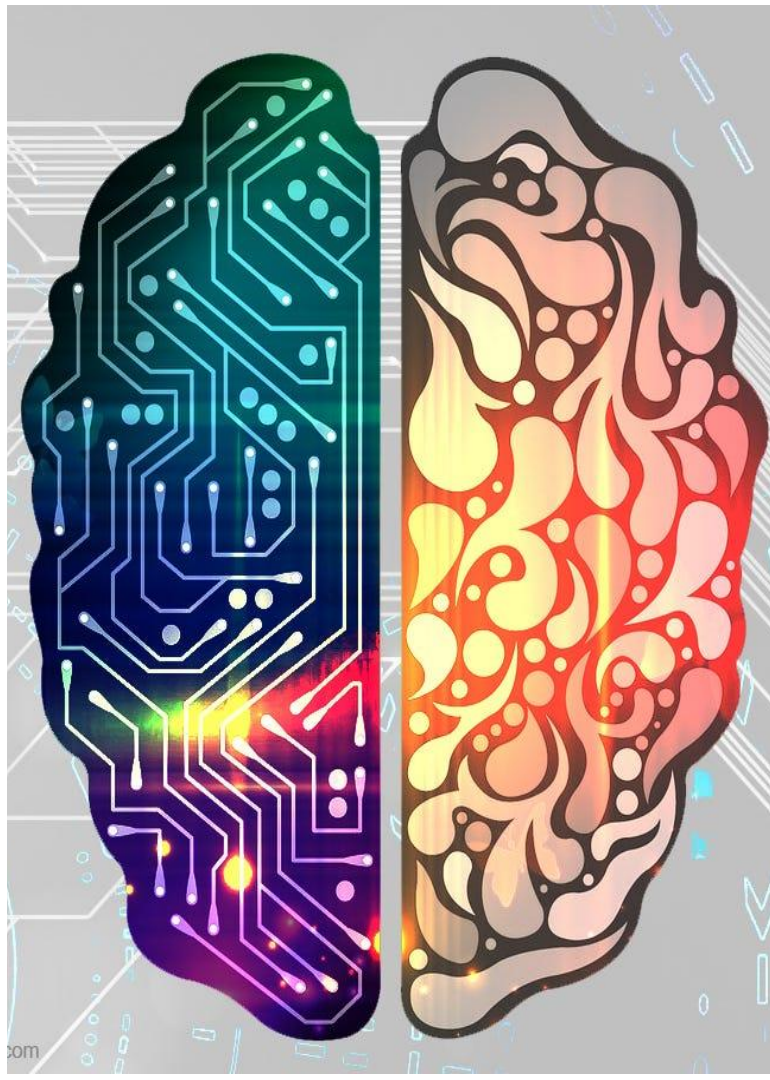
- **Path Recording**   (stop_id, stop_name, type_of_transport, departure_time, arrival_time)

- **Walking Segments Handling**

- **2 min minimum change requirement**

How are delays predicted ?

Prediction Pipeline

- Random Forest Classifier model for multiclass classification.
- 46 classes (0 to 45) representing the delay in minutes.
- The data is split into training and testing sets with an 80–20 ratio.
- Goal: Predict the delay (in minutes) for a given transport, departure and arrival time and the weather condition
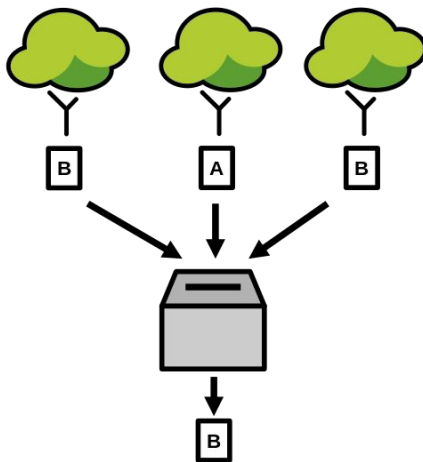
Prediction Pipeline

Pipeline

- StandardScaler for feature scaling.
- RandomForestClassifier for classification.
- Parameter Grid for Hyperparameter Tuning:
    - n_estimators (number of trees in the forest): [20, 50, 100, 200].
    - max_depth (maximum depth of the tree): [5, 10, 20].
- Cross-Validation Configuration:
    - GridSearchCV is used to perform hyperparameter tuning with 3-fold cross-validation, optimizing for accuracy.
- Output : A number between 0 and 45

Speaker

ROBUST JOURNEY PLANNER
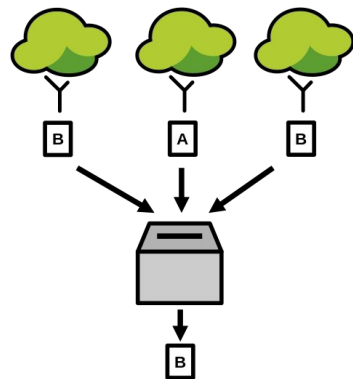
How well does our model perform ?

And potential future improvements

Model Evaluation

How we evaluated the model:

- Accuracy: The ratio of correctly predicted instances to the total instances.
- Precision: The ratio of correctly predicted positive observations to the total predicted positives.
- Recall: The ratio of correctly predicted positive observations to the all observations in actual class.
- F1 Score: The weighted average of Precision and Recall.
- AUC (Area Under the ROC Curve): Measures the ability of the model to distinguish between classes.
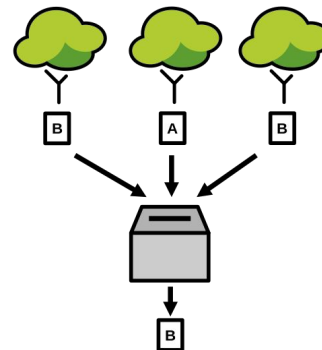
Preliminary Results

| Accuracy	Precision	Recall	F1 Score	AUC
0.52	0.47	0.52	0.47	0.97

ROBUST JOURNEY PLANNER

Result Analysis

- Random guessing would yield an accuracy of approximately 1/#classes = 1/46 ≈ 0.022.
- Accuracy = 0.52 is significantly better than random guessing
    - Indicates that the model is capturing some patterns in the data despite the complexity introduced by having 46 classes.
- Precision = 0.47 suggests that the model has a relatively high rate of false positives, but it is still doing better than chance.
- Recall = 0.52 indicates that the model correctly identified 52% of the actual positive instances for each class.
- Maintaining a recall above 50% is challenging and shows that the model is fairly good at identifying actual positive cases.

Result Analysis

- F1 Score = 0.47
  - reflects a balance between precision and recall but indicates that both metrics are relatively moderate.
- AUC = 0.97 indicates that classifier has a 97% chance of correctly distinguishing between a randomly chosen positive instance and a randomly chosen negative instance.
  - model's predictions are reliable across different thresholds. This means the model is not only good at a specific threshold but performs well generally.

Speaker

ROBUST JOURNEY PLANNER

Performance evaluation

Robust Journey Planner
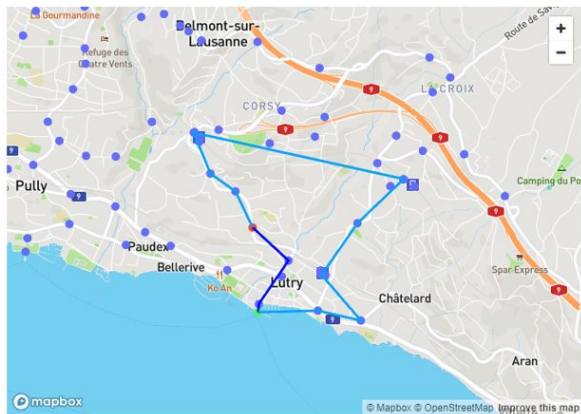
Your trip from Lutry (lac) to Lutry, Orzens is planned for `2024-05-28 21:32:51.061794` and you will arrive latest at `00:32:00` with a 🌧 weather.

The Optimal Path path proposed is:

Travel from Lutry (lac) to Lutry, port with the following mean of transport : walk. You have 1.0min to change the porbability of a successful change is 1

Travel from Lutry, port to Lutry, Les Champs with the following mean of transport : B. You have 6.0min to change the porbability of a successful change is 1

Travel from Lutry, Les Champs to Lutry, Orzens with the following mean of transport : walk. You have 5.0min to change the porbability of a successful change is 1

Total time of the path is 12.0 min and the overall probability of success is 1

SBB Path

Departure: Lutry (lac) (21:01) -> Arrival: Lutry, Orzens (21:12) | Duration: 0:11:00
From Lutry (lac) at 2024-05-28 21:01:00 to Lutry, Les Champs at 2024-05-28 21:11:00
    Walking or transfer
From Lutry, Les Champs at 2024-05-28 21:11:00 to Lutry, Orzens at 2024-05-28 21:12:00
    Transport: 68026 from platform None to platform None
    Category: B
    Operator: TL

Your trip from Ecublens VD, allée de Dorigny to Ecublens VD, Croset is planned for `2024-05-28 21:35:21.971324` and you will arrive latest at `00:35:00` with a ☁ weather.

The Optimal Path path proposed is:

Travel from Ecublens VD, allée de Dorigny to Lausanne, Sablons with the following mean of transport : B. You have 3min to change the porbability of a successful change is 0.8

Travel from Lausanne, Sablons to Lausanne, Bourdonnette nord with the following mean of transport : B. You have 3min to change the porbability of a successful change is 0.8

Travel from Lausanne, Bourdonnette nord to Chavannes-R., Dorigny with the following mean of transport : walk. You have 5.0min to change the porbability of a successful change is 1

Travel from Chavannes-R., Dorigny to Chavannes-R., Talluchet with the following mean of transport : B. You have 3.0min to change the porbability of a successful change is 1

Travel from Chavannes-R., Talluchet to Renens VD, Censuy with the following mean of transport : walk. You have 5.0min to change the porbability of a successful change is 1

Travel from Renens VD, Censuy to Renens VD, gare sud with the following mean of transport : B. You have 2.0min to change the porbability of a successful change is 1

Travel from Renens VD, gare sud to Ecublens VD, Epenex/Pont-Bleu with the following mean of transport : walk. You have 5.0min to change the porbability of a successful change is 1

Travel from Ecublens VD, Epenex/Pont-Bleu to Ecublens VD, Suchet-Forêt with the following mean of transport : B. You have 1.0min to change the porbability of a successful change is 1

Travel from Ecublens VD, Suchet-Forêt to Ecublens VD, Parc with the following mean of transport : B. You have 3min to change the porbability of a successful change is 0.8

Travel from Ecublens VD, Parc to Ecublens VD, Croset with the following mean of transport : walk. You have 6.0min to change the porbability of a successful change is 1

Total time of the path is 36.0 min and the overall probability of success is 0.5120000000000001

The Alternative path proposed is:

Travel from Ecublens VD, allée de Dorigny to Lausanne, Sablons with the following mean of transport : B. You have 3min to change the porbability of a successful change is 0.8

Travel from Lausanne, Sablons to Lausanne, Bourdonnette nord with the following mean of transport : B. You have 3min to change the porbability of a successful change is 0.8

Travel from Lausanne, Bourdonnette nord to Chavannes-R., Dorigny with the following mean of transport : walk. You have 5.0min to change the porbability of a successful change is 1

Travel from Chavannes-R., Dorigny to Chavannes-R., Talluchet with the following mean of transport : B. You have 3.0min to change the porbability of a successful change is 1

Travel from Chavannes-R., Talluchet to Renens VD, Censuy with the following mean of transport : walk. You have 5.0min to change the porbability of a successful change is 1

Travel from Renens VD, Censuy to Renens VD, gare sud with the following mean of transport : B. You have 2.0min to change the porbability of a successful change is 1

Travel from Renens VD, gare sud to Ecublens VD, Epenex with the following mean of transport : walk. You have 0.0min to change the porbability of a successful change is 1

Travel from Ecublens VD, Epenex to Chavannes-R., Crochy with the following mean of transport : M. You have 1.0min to change the porbability of a successful change is 1

Travel from Chavannes-R., Crochy to Ecublens VD, Croset with the following mean of transport : walk. You have 1.0min to change the porbability of a successful change is 1

Total time of the path is 23.0 min and the overall probability of success is 0.6400000000000001

SBB Paths proposed

Departure: Ecublens VD, allée de Dorigny (21:01) -> Arrival: Ecublens VD, Croset (21:19) | Duration: 0:19:00

From Ecublens VD, allée de Dorigny at 2024-05-28 21:01:00 to St-Sulpice VD, Castolin at 2024-05-28 21:08:00
    Transport: 1144 from platform None to platform None
    Category: B
    Operator: MBC Auto
From St-Sulpice VD, Castolin at 2024-05-28 21:08:00 to St-Sulpice VD, Venoge nord at 2024-05-28 21:10:00
    Walking or transfer
From St-Sulpice VD, Venoge nord at 2024-05-28 21:10:00 to Ecublens VD, Croset at 2024-05-28 21:19:00
    Transport: 33721 from platform None to platform None
    Category: B
    Operator: TL

Departure: Ecublens VD, allée de Dorigny (20:51) -> Arrival: Ecublens VD, Croset (21:14) | Duration: 0:23:00

From Ecublens VD, allée de Dorigny at 2024-05-28 20:51:00 to Lausanne, Bourdonnette at 2024-05-28 20:54:00
    Transport: 1145 from platform None to platform None
    Category: B
    Operator: MBC Auto
From Lausanne, Bourdonnette at 2024-05-28 20:54:00 to Lausanne, Bourdonnette nord at 2024-05-28 20:59:00
    Walking or transfer
From Lausanne, Bourdonnette nord at 2024-05-28 21:00:00 to Renens VD, gare sud at 2024-05-28 21:05:00
    Transport: 25253 from platform None to platform None
    Category: B
    Operator: TL
From Renens VD, gare sud at 2024-05-28 21:09:00 to Ecublens VD, Croset at 2024-05-28 21:14:00
    Transport: 33139 from platform None to platform None
    Category: B
    Operator: TL

Limitations and Benefits

General trends:

      Proposes many changes

      Longer trip times

Benefits:

      ML model to calculate probability of making trip

Thank you for listening !