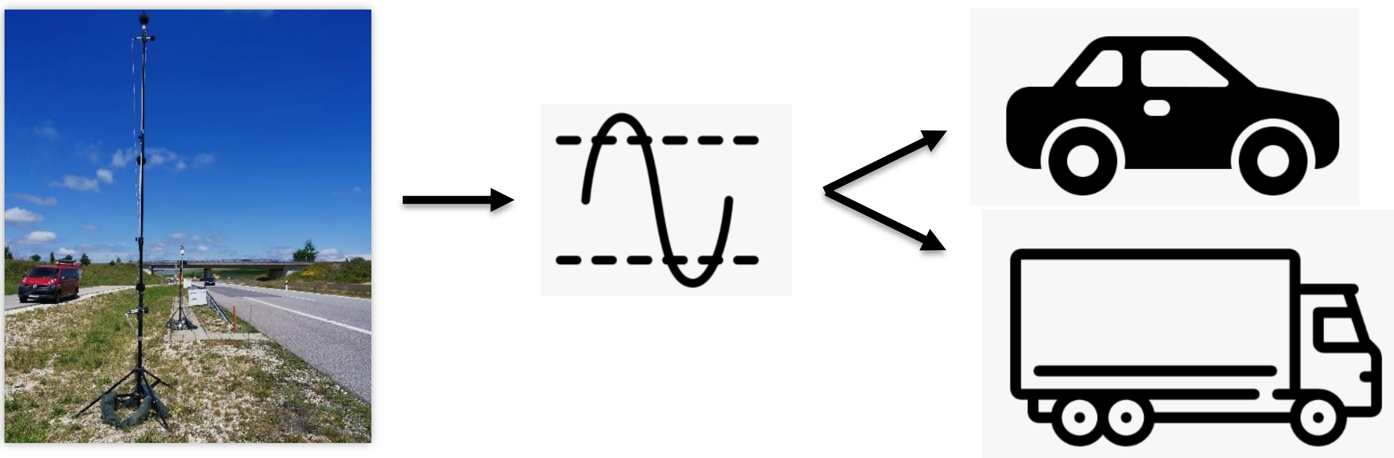
Determining the Noise Behaviour of the German Vehicle Fleet by Measurement

Project of the Federal Department of Environment

Determine the **noise behaviour** of the german vehicle fleet by making measurments in **30 different locations**



Project aim: develop a classification method for passenger cars and heavy vehicles in order to distinguish them

Timeline

- **Kick-Off**
- Introduction to problem
- Project management
- Data processing: analyze data distribution & use different data features
- Machine Learning: try different models

Day 1 - 08/01

Day 2 - 09/01

- Continue with data processing
- Machine Learning: build one big model

- Continue with data processing
- Model training
- Model testing
- Perfomance analysis
- Look into futher approaches
- **Reflection meeting**

Day 3 - 10/01

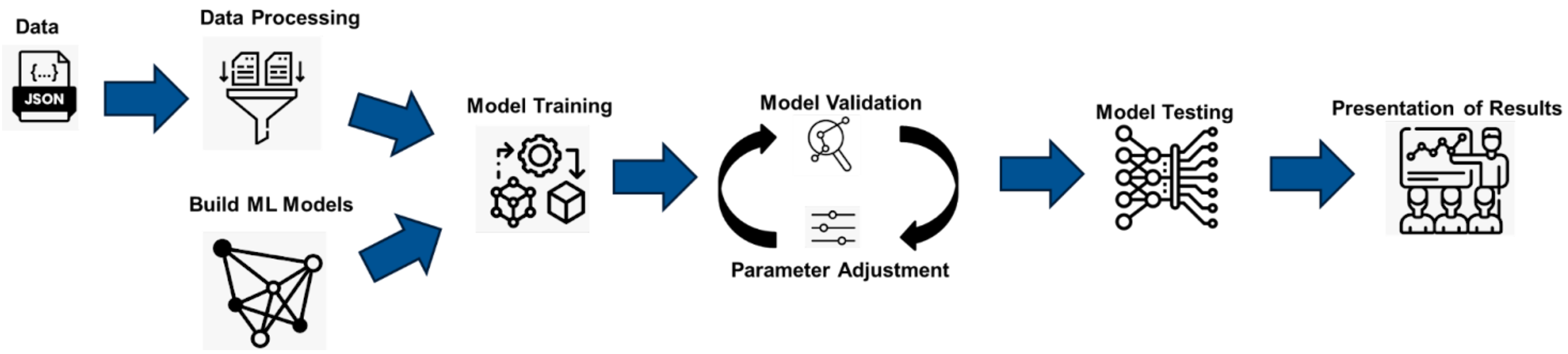
Day 4 - 11/01

- Look deeper into futher approaches
- Analyse results
- Documentation

- Presentation
- Poster creation

Day 5 - 12/01

Problem Approach



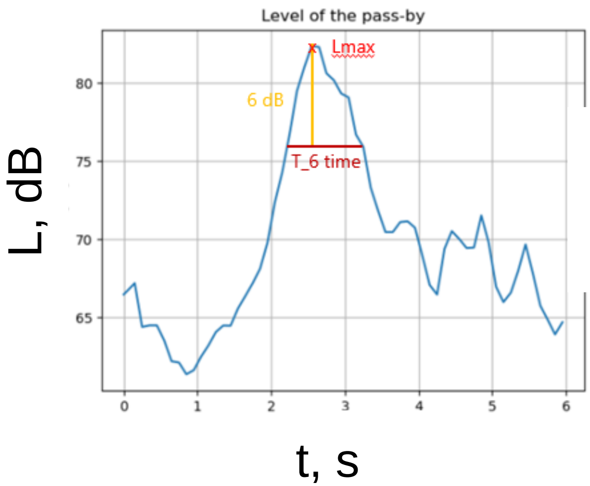
Captured Data

'ID':	unique ID
'MP':	measurement location number
'IDMP':	measurement location specific ID
'timestamp':	time stamp
'Lmax1':	max sound pressure level (SPL) channel 1
'Lmax2':	max SPL channel 2
'Lmax3':	max SPL channel 3
'levelTime1':	level-time-curve channel 1
'levelTime2':	level-time-curve channel 2
'levelTime3':	level-time-curve channel 3
'prominence1':	peak prominence of level-time-curve channel 1
'prominence2':	peak prominence of level-time-curve channel 2
'prominence3':	peak prominence of level-time-curve channel 3
'width1':	peak width of level-time-curve channel 1
'width2':	peak width of level-time-curve channel 2
'width3':	peak width of level-time-curve channel 3
'T6_1':	T6 time channel 1
'T6_2':	T6 time channel 2

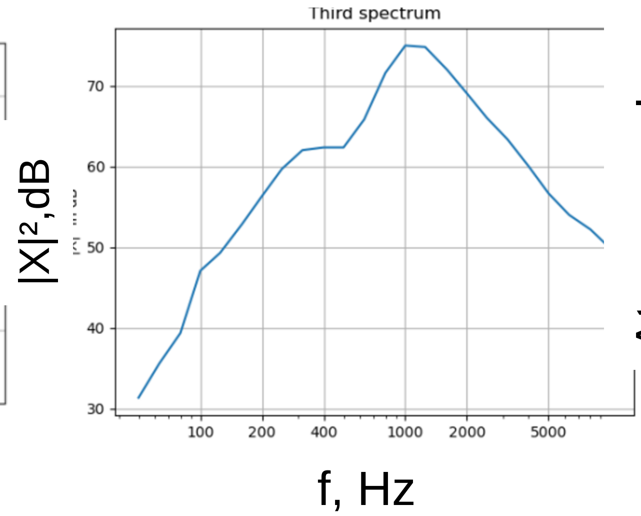
'Leq1':	equivalent SPL channel 1
'Leq2':	equivalent SPL channel 2
'SEL1':	sound exposure level channel 1
'SEL2':	sound exposure level channel 2
'thirdSpectrum1':	third-octave band spectrum channel 1
'thirdSpectrum2':	third-octave band spectrum channel 2
'thirdSpectrum3':	third-octave band spectrum channel 3
'trajectory':	passby localization angle over time
'distance':	distance between vehicle and microphone
'temperature':	air temperature
'relativeHumidity':	air humidity
'windSpeed':	wind speed
'velocity':	vehicle speed
'radarPulses':	vehicle length (estimated by radar)
'timeGap':	time to previous vehicle
'vehicleClass':	vehicle category

Captured Data - Examples

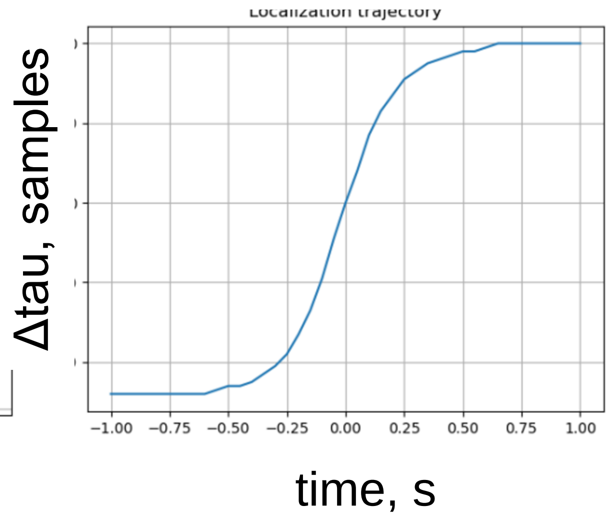
Level of the pass-by



Third spectrum



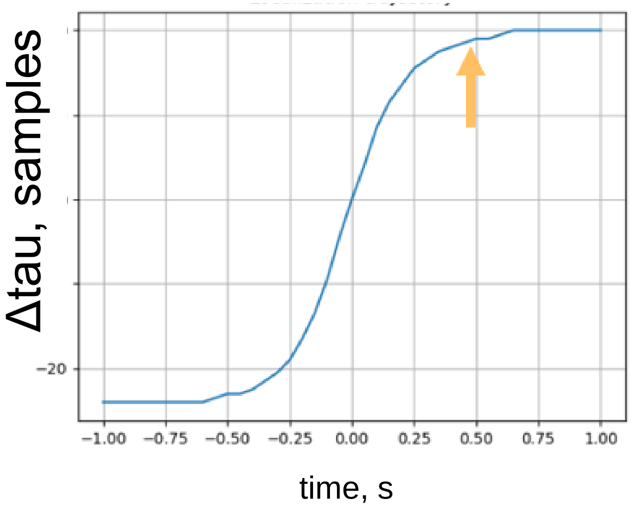
Localization trajectory



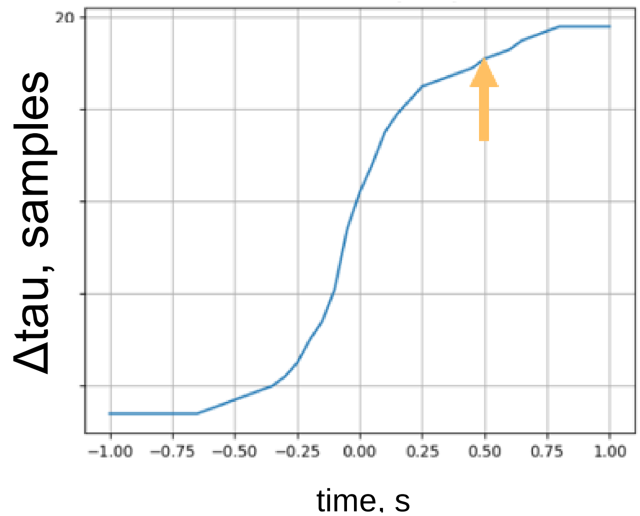
- ‘Lmax’: max sound pressure level (max in ‘levelTime’ curve)
  - T_6: width of the peak in ‘levelTime’ curve at (Lmax – 6) dB
- ‘thirdSpectrum’: third-octave spectrum measured at a short time interval around the time point of max sound level
- ‘trajectory’: measure of vehicle localization (angle between mic pair axis and direction of sound)

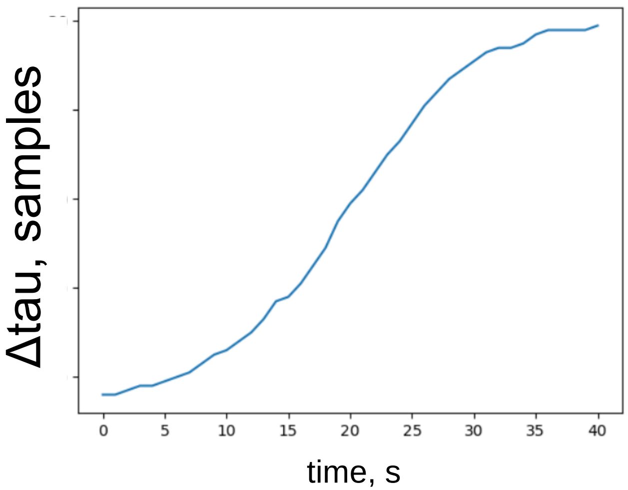
Data Processing – Slope of trajectory as an additional data feature

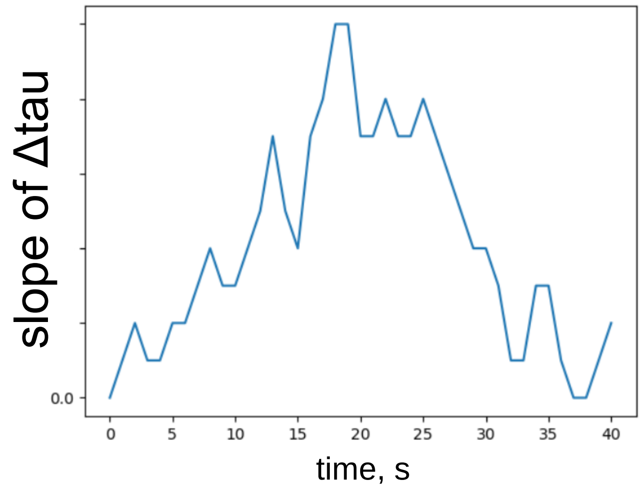
Localization trajectory of a PKW



Localization trajectory of LKW



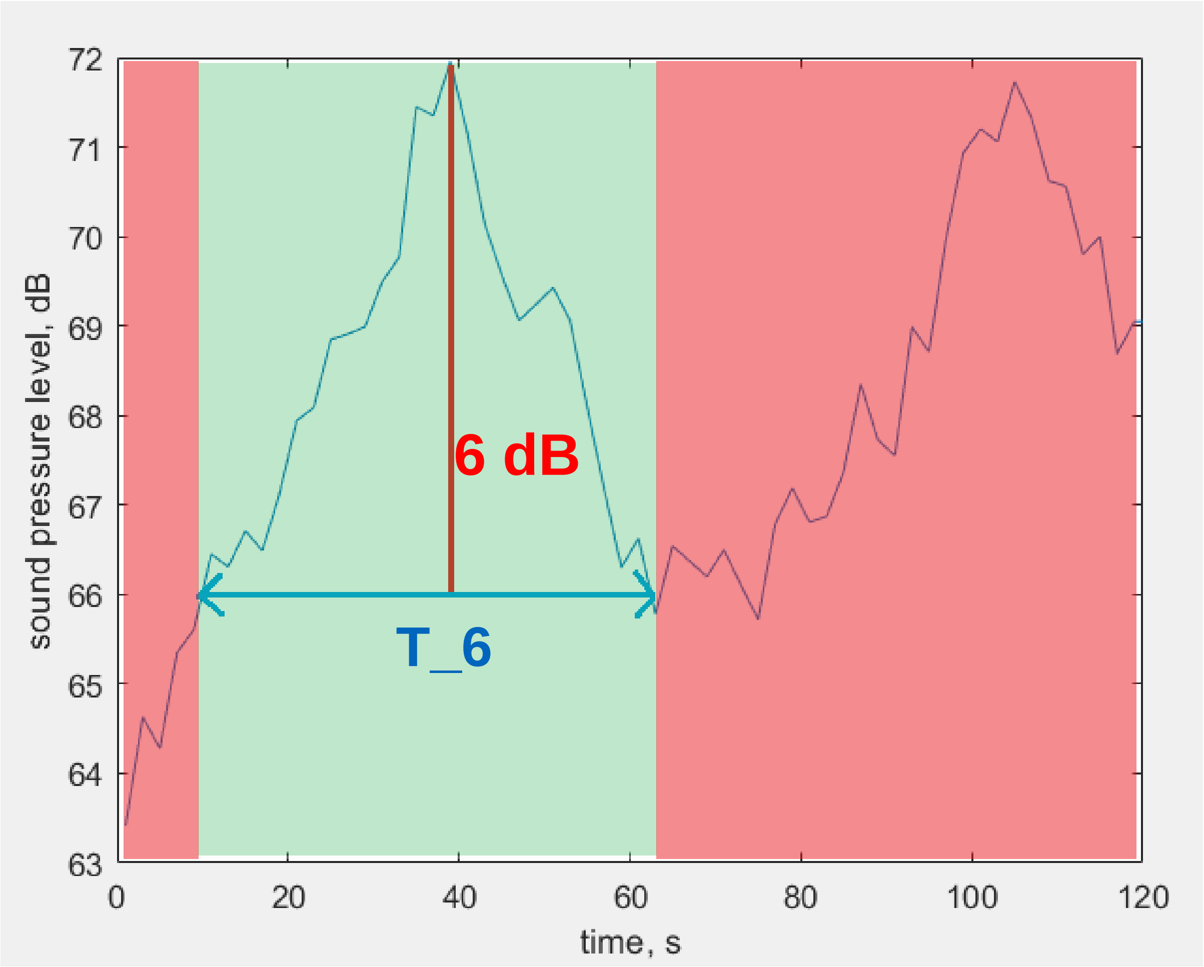




- LKWs are longer ⇒ slower increase of localization angle

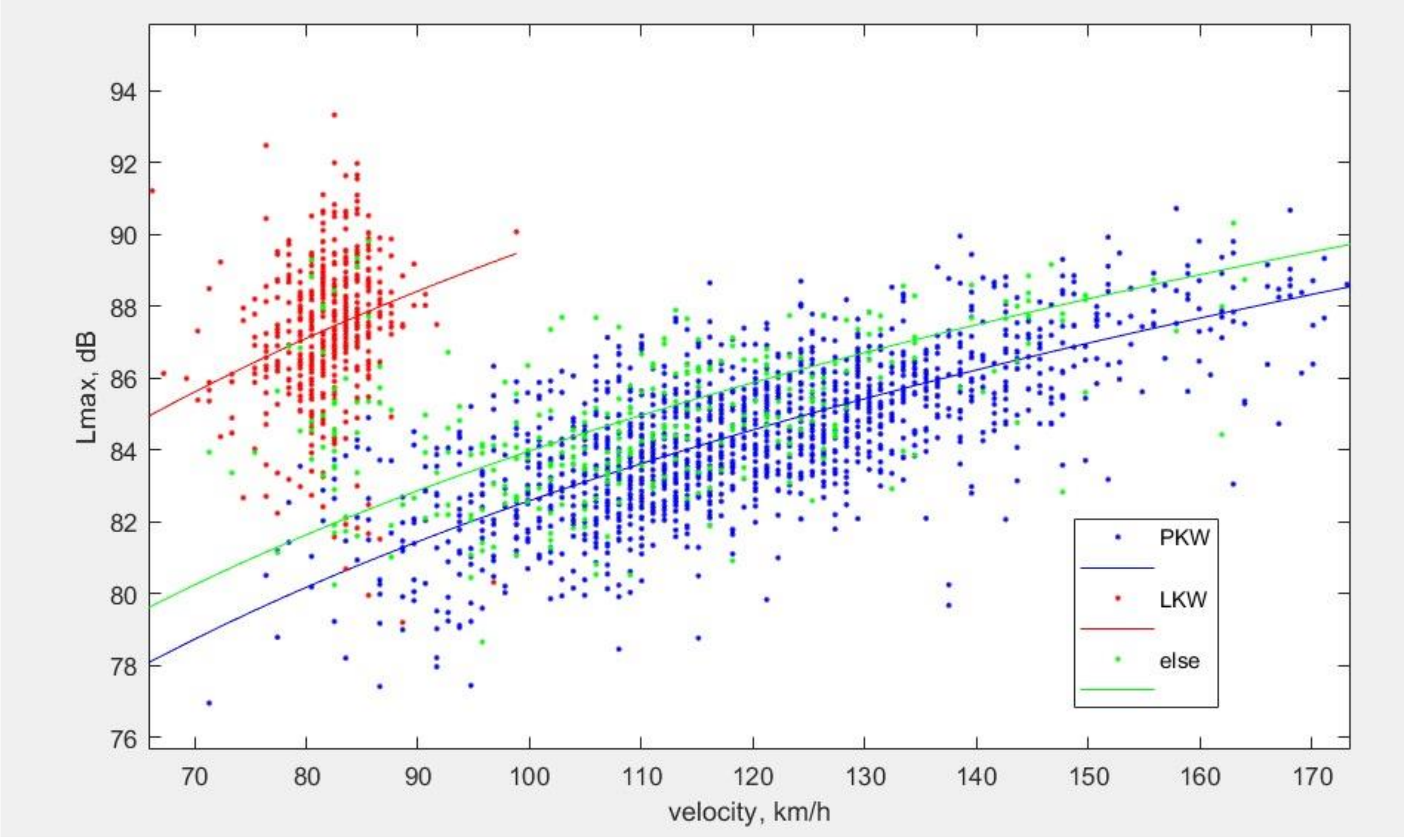
- Slope of ‘trajectory’ (calculated point-wise) can be used as an additional data feature

Data Processing – Filtering out noise in level-time-curves



- Cropping ‘levelTime’ data at T_6 interval can help to get rid of noise produced by non-target cars in audio recordings
- Should be handled carefully in case of closely following cars (the peaks might not be resolved with 6 dB criterion)

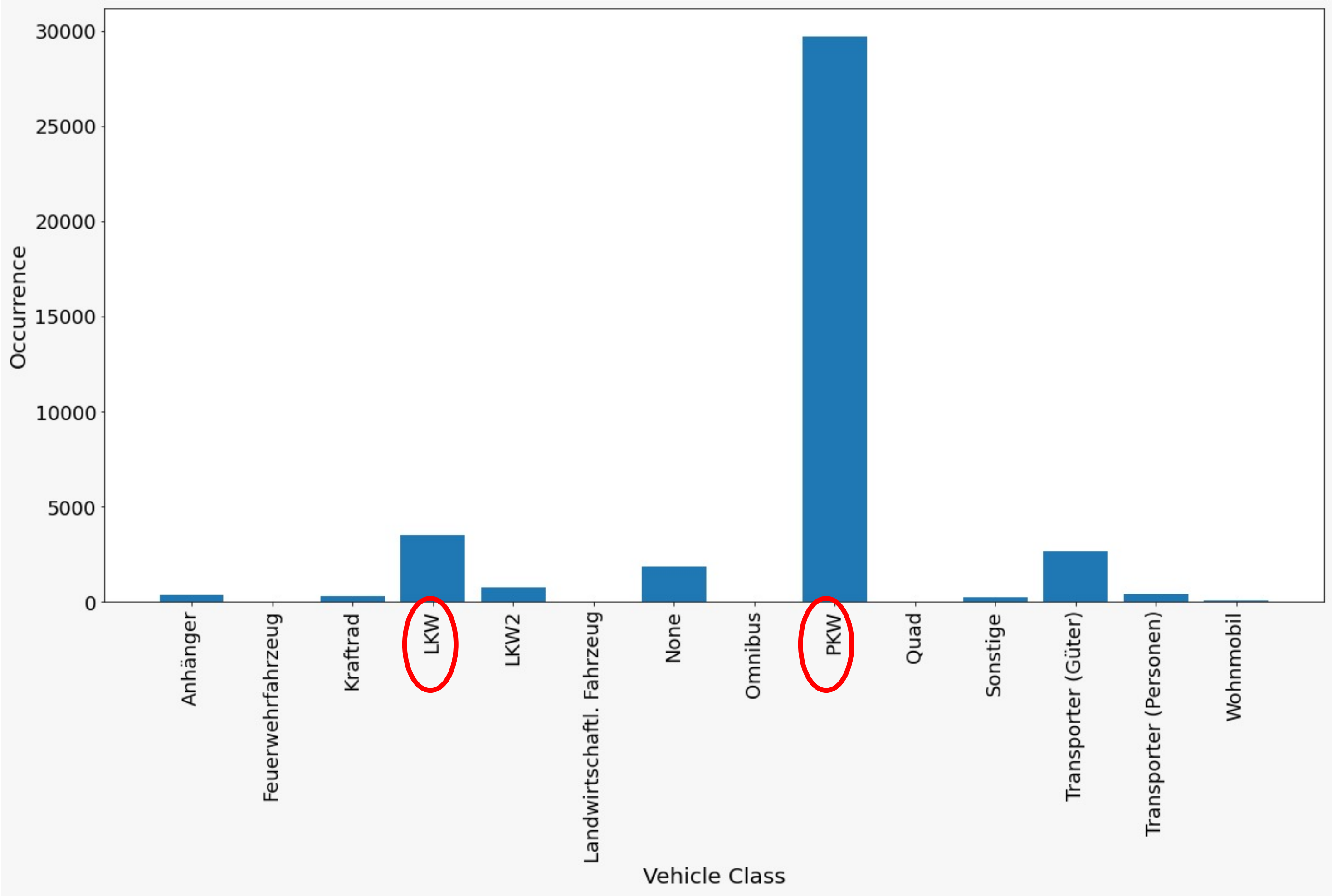
Data Processing – Deleting data features by normalization



Scatter plots of Lmax vs. velocity and respective regression curves

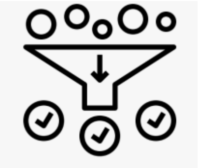
- Max sound pressure level (Lmax) linearly increases with lg(velocity)
- The feature ‘velocity’ can be discarded if the Lmax values are normalized by lg(velocity)

Data: Vehicle Class Distribution



Data Processing: Filter Features

'ID':	unique ID	'Leq1':	equivalent SPL channel 1
'MP':	measurement location number	'Leq2':	equivalent SPL channel 2
'IDMP':	measurement location specific ID	'SEL1':	sound exposure level channel 1
'timestamp':	time stamp	'SEL2':	sound exposure level channel 2
'Lmax1':	max sound pressure level (SPL) channel 1	'thirdSpectrum1':	third-octave band spectrum channel 1
'Lmax2':	max SPL channel 2	'thirdSpectrum2':	third-octave band spectrum channel 2
'Lmax3':	max SPL channel 3	'thirdSpectrum3':	third-octave band spectrum channel 3
'levelTime1':	level-time-curve channel 1	'trajectory':	passby localization angle over time
'levelTime2':	level-time-curve channel 2	'distance':	distance between vehicle and microphone
'levelTime3':	level-time-curve channel 3	'temperature':	air temperature
'prominence1':	peak prominence of level-time-curve channel 1	'relativeHumidity':	air humidity
'prominence2':	peak prominence of level-time-curve channel 2	'windSpeed':	wind speed
'prominence3':	peak prominence of level-time-curve channel 3	'velocity':	vehicle speed
'width1':	peak width of level-time-curve channel 1	'radarPulses':	vehicle length (estimated by radar)
'width2':	peak width of level-time-curve channel 2	'timeGap':	time to previous vehicle
'width3':	peak width of level-time-curve channel 3	'vehicleClass':	vehicle category
'T6_1':	T6 time channel 1		
'T6_2':	T6 time channel 2		



df: all datapoints, **all features**
(40114 x **34**)

filtered_df: all datapoints, **suggested features**
(40114 x **9**)

'ID'

'MP'

'Lmax1'

'levelTime1

'T6_1'

'thirdSpectrum1'

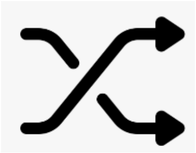
'trajectory'

'radarPulses'

'vehicleClass'

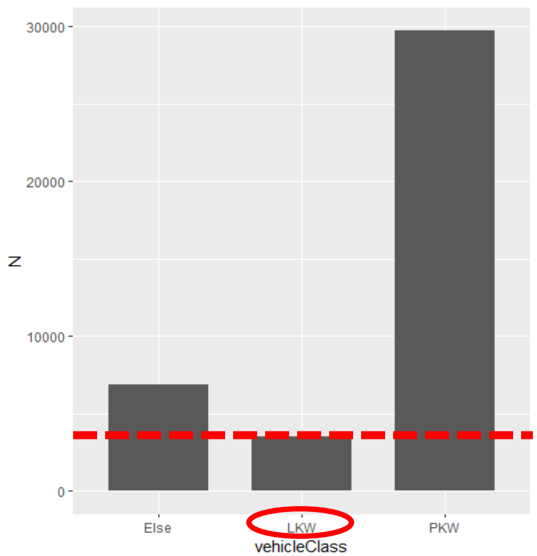
Data Processing: Adjust Class Distribution

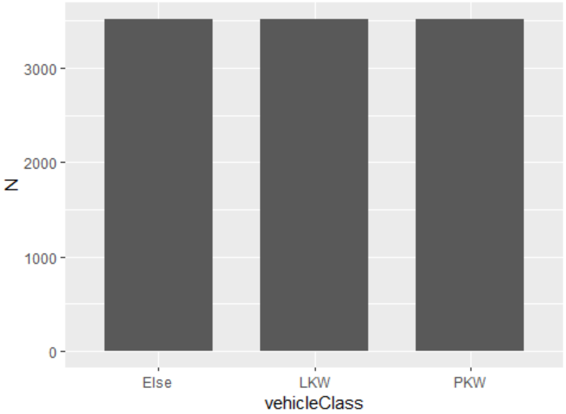
Shuffel data table



ID	MP	Lmax1	levelTime1	T6_1	thirdSpectrum1	trajectory	radarPulses	vehicleClass
1384	MP2	0,00909	[61.68, 62.05,	1,9	[32.82, 44.63, 36.6	[-20.0, -19.(	822	LKW
3942	MP4	0,00844	[64.0, 64.22, 6	2,9	[31.4, 24.98, 30.28	[-20.0, -20.(	412	PKW
39678	MP29	0,01015	[70.915, 71.49	1,3	[34.38, 42.45, 43.5	[-1.0, -5.5, -	374	PKW

Extract equal datapoints for each class





filtered_final_df: datapoints with **equal ratio** **`'Else':'LKW':'PKW'`** (**10563**x9)

Distribution over all locations of **filtered_df** (**40114** x 9)

Data Processing: Normalize Data

Normalization

Single values: find min and max values in column, normalize values as

new_x = (x-min)/(max-min)

Arrays: divide each value in every array with the max value found in the column

Remove locations with special conditions

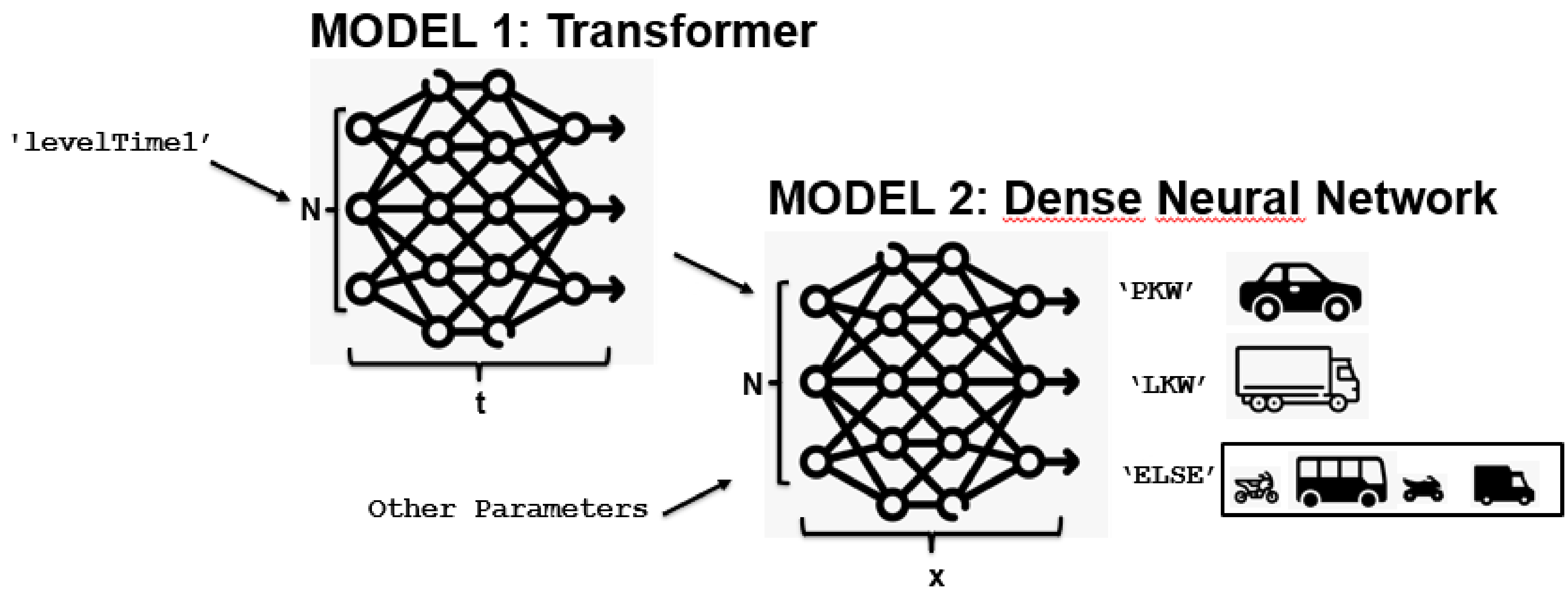
Location	Condition
MP 30, MP 5	Road surface
MP 12, MP 13	Winter

Data Processing: Datasets

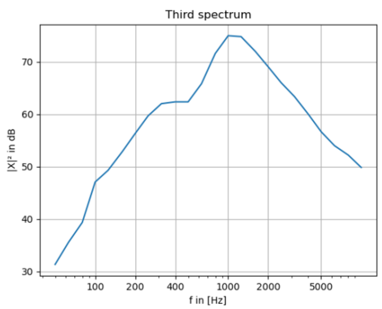
Color code: designed, created, used

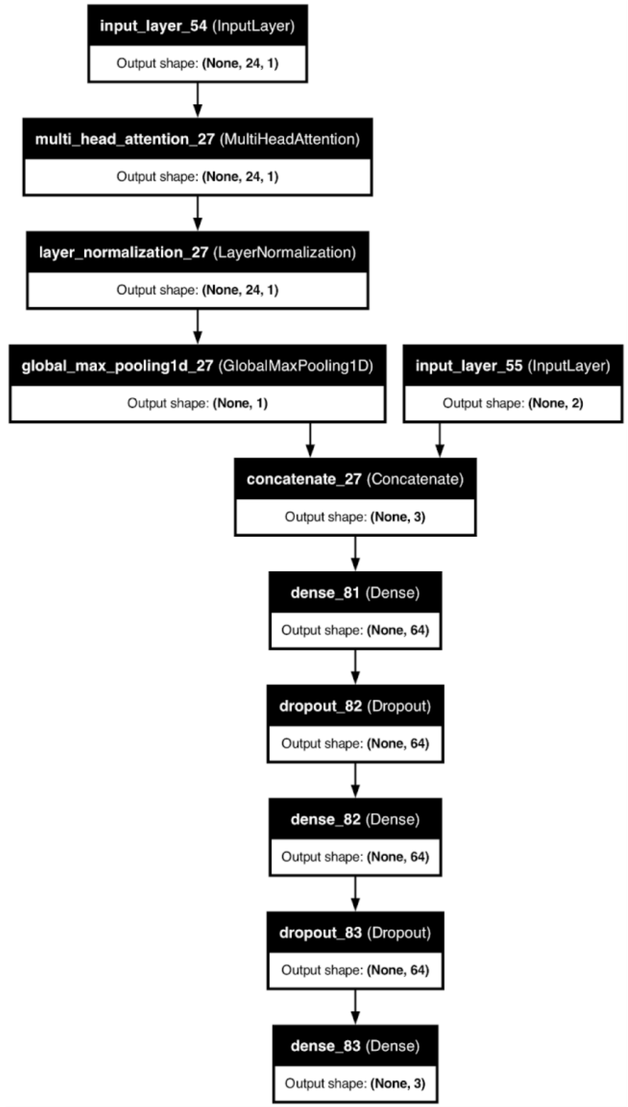
Name	Description	Size
ID1	all datapoints, all features	29882 x 34
ID2	all datapoints, selected features	39676 x 9
ID3	all datapoints, all features, 3 classes	
ID4	all datapoints, selected features, 3 vehicle classes	39676 x 9
ID5	all features, 3 equally distributed vehicle classes	
ID6	selected features, 3 equally distributed vehicle classes	10419 x 9
ID7	all features, 3 equally distributed classes, special locations in bottom rows	
ID8	selected features, 3 equally distributed vehicle classes, special locations in bottom rows	10419 x 9
ID9	selected features, 3 equally distributed vehicle classes, special locations excluded	9873 x 9
ID10	selected features, 3 equally distributed vehicle classes, additional features	

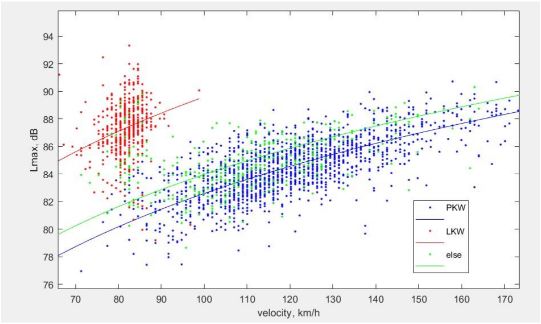
Machine Learning Model



Machine Learning Model

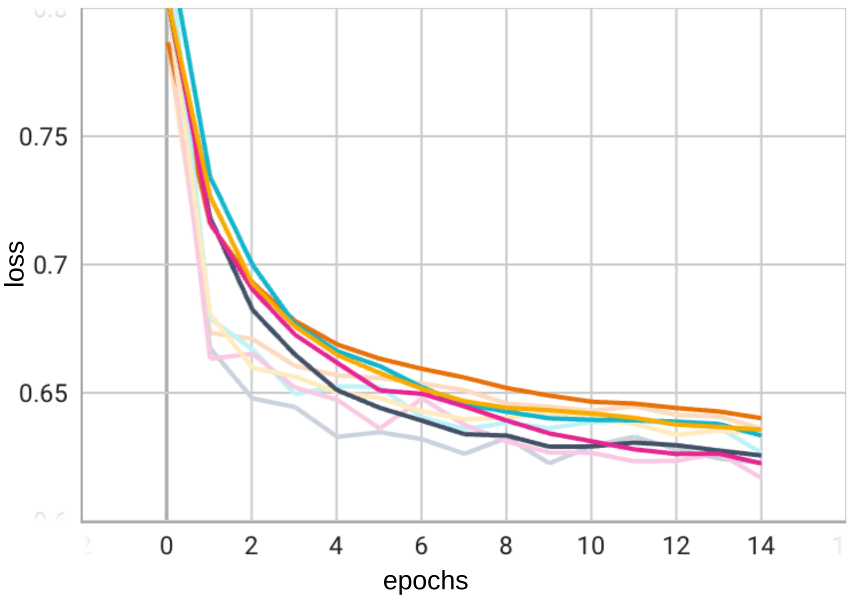






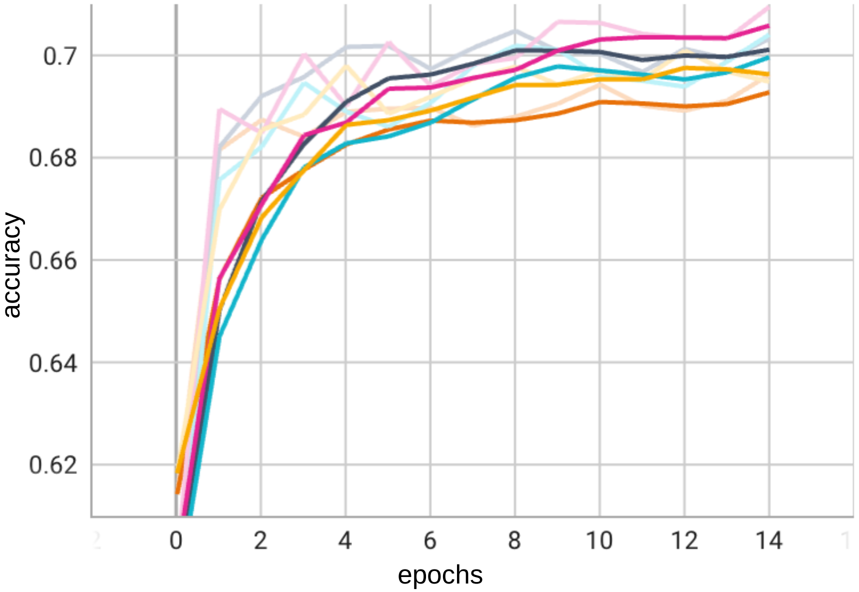
Model Testing

Epoch loss



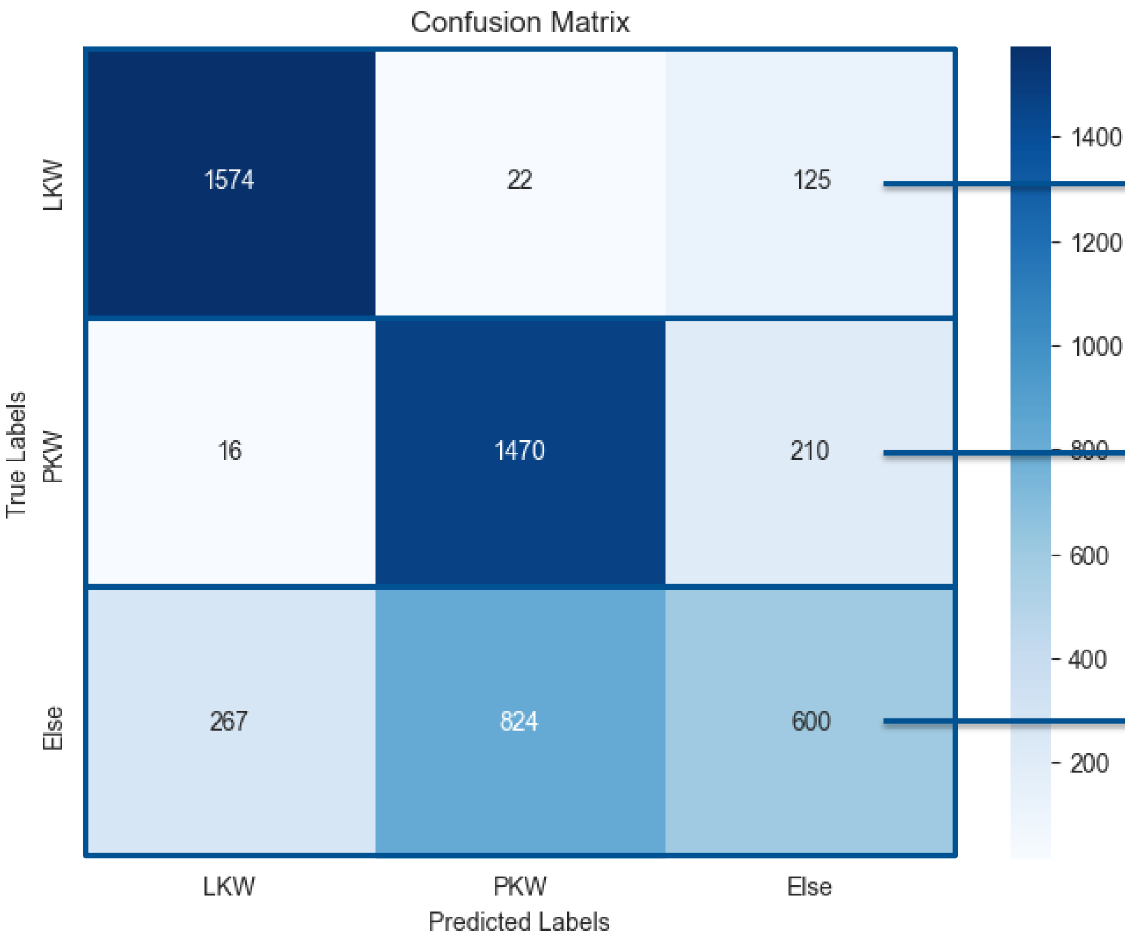
Epochs = 15, batch_size = 64, learning_rate = 1e-2

Epoch accuracy



Epochs = 15, batch_size = 64, learning_rate = 1e-2

Test Accuracy



Source: confusion matrix from test group ID 2

True Positive Rate

91,4%

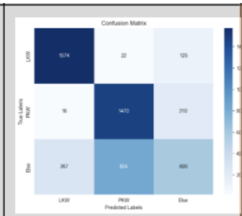
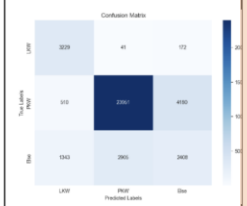
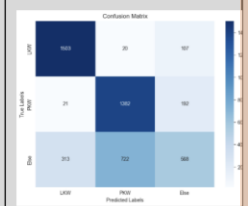
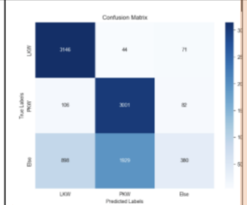
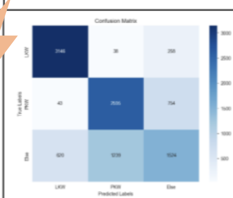
86,7%

35,5%

False Negative Rate

PKW	LKW
48,7%	15,8%

Results

Group description	Accuracy	Confusion Matrix	PKW	LKW	Else (TPR)
model validation Train and test with special roads	0.7091		0.867	0.914	0.355
train with special road and test on all datasets	0.7637		0.836	0.938	0.362
model validation Train and test without special roads	0.7062		0.866	0.922	0.354
train with special road and test without special road	0.6759		0.941	0.965	0.118
train without special road and test with special road	0.4532		0.914	0.765	0.45

Results

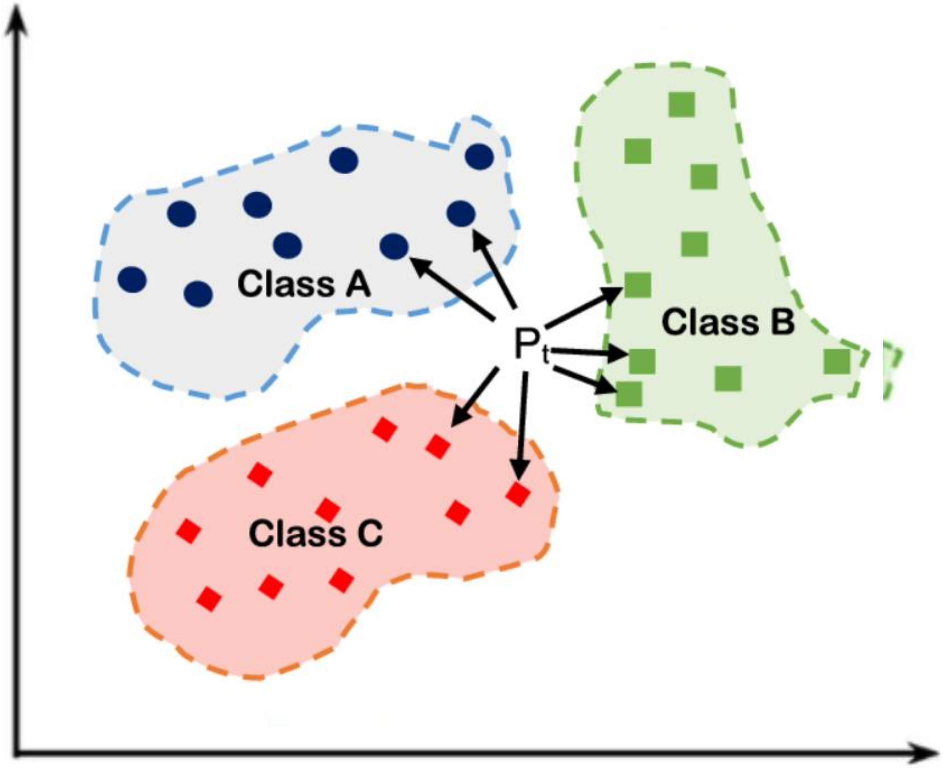
- a robust classification between LKW (trucks) and PKW (cars) based on the given datasets.

- There is a noticeable limitation when identifying the 'Else' group, evidenced by its lower true positive rate.

- The model trained with data on special roads performs well on datasets, but reversely it doesn’t. It is suggested using the dataset with the special road for the training process

Conventional Machine Learning Model

K Nearest Neighbors(KNN)



https://medium.com/@sachinsoni600517/k-nearest-neighbours-introduction-to-machine-learning-algorithms-9dbc9d9fb3b2

Three Classes:

PKW; LWK; Else

Input:

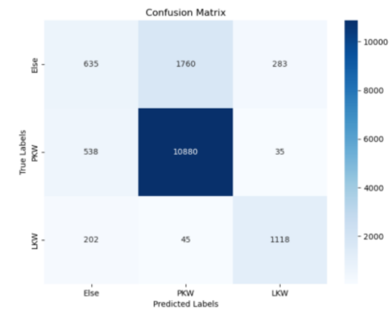
float array

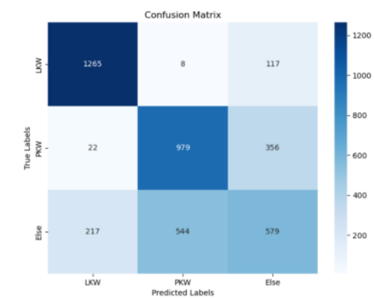
Output:

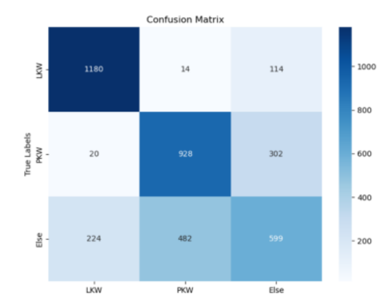
classification index

Testing Results

Training size	Dataset	Input	Accuracy
(23243, 188)	ID4_normalized. csv	levelTime1, trajectory, thirdSpectrum1 T6_1, Lmax1, RadarPulse	all: 0.81524 LKW: 0.81905 PKW: 0.94997 Else: 0.23712
(6130, 188) three classes equally distributed	ID6_normalized. csv	levelTime1, trajectory, thirdSpectrum1 T6_1, Lmax1, RadarPulse	all: 0.69073 LKW: 0.91007 PKW: 0.72144 Else: 0.43209
(5794, 188) remove special location	ID9_normalized. csv	levelTime1, trajectory, thirdSpectrum1 T6_1, Lmax1, RadarPulse	all: 0.70075 LKW: 0.90214 PKW: 0.74240 Else: 0.45900

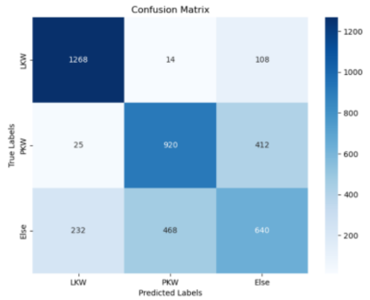


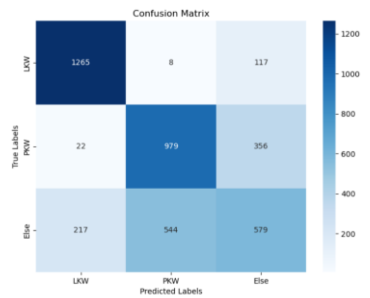


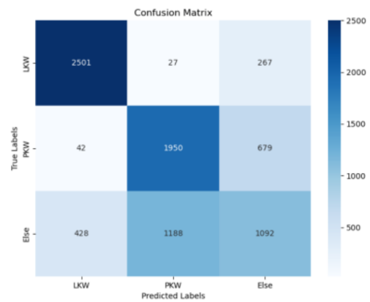


Testing Results

Training size	Dataset	Input	Accuracy
(6130, 24)	ID6_normalized. csv	thirdSpectrum1	all: 0.69195 LKW: 0.91223 PKW: 0.67797 Else: 0.47761
(6130, 188) three classes equally distributed	ID6_normalized. csv	levelTime1, trajectory, thirdSpectrum1 T6_1, Lmax1, RadarPulse	all: 0.69073 LKW: 0.91007 PKW: 0.72144 Else: 0.43209
(2043, 188) reduced training size	ID6_normalized. csv	—	all: 0.67813 LKW: 0.90214 PKW: 0.74240 Else: 0.45900
(8173, 188) increase training size	ID6_normalized. csv	—	all: 0.69814 LKW: 0.91079 PKW: 0.73252 Else: 0.45152







Results II

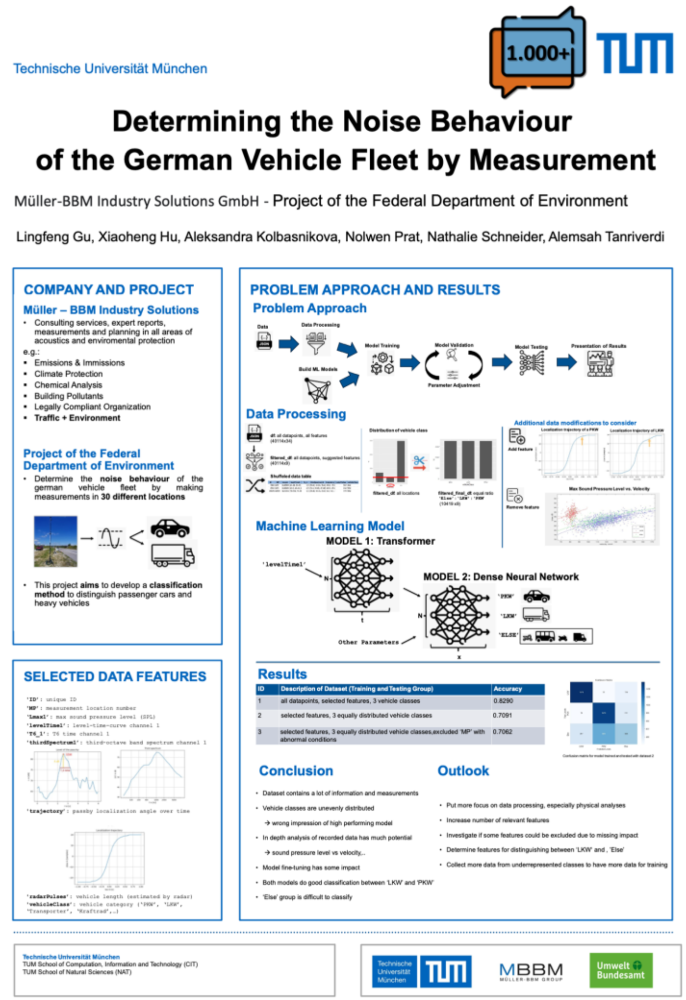
- Commendable performance with a small dataset. As the volume of data increases, the model shows a slight improvement

- The model is good at distinguishing between 'PKW' and 'LKW' categories. However, its ability to differentiate the 'else' category is relatively weak

- The 'thirdSpectrum' feature plays a crucial role in enhancing the model's performance

Final Results

- **2 Machine Learning approaches to classify the vehicles**
- **Different Data Processing steps to improve classification process**
- **Improvement of softskills**
- **Poster with overview of project week**



Conclusion

- Dataset contains a lot of information and measurements
- Vehicle classes are unevenly distributed

  → wrong impression of high performing model

- In depth analysis of recorded data has much potential (sound pressure level vs velocity,..)
- Model fine-tuning has some impact
- Both models do good classification between ‘LKW’ and ‘PKW’
- ‘Else’ group is difficult to classify

Outlook

- Put more focus on data processing, especially physical analyses
- Increase number of relevant features
- Investigate if some features could be excluded due to missing impact
- Determine features for distinguishing between ‘LKW‘ and ‚Else‘
- Collect more data from underrepresented classes to have more data for training