Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 15th International Scientific and Practical Conference. Volume II, 317-323

Environmental Data and Digital Twins for Road Traffic Safety in Rural Areas

Roberts Volkovičs
Modelling of sociotechnical systems
Vidzeme University of Applied Sciences
Valmiera, Latvia
roberts.volkovics@va.lv

Viesturs Bambāns
Modelling of sociotechnical systems
Vidzeme University of Applied Sciences
Valmiera, Latvia
viesturs.bambans@va.lv

Ginta Majore
Modelling of sociotechnical systems
Vidzeme University of Applied Sciences
Valmiera, Latvia
ginta.majore@va.lv

Abstract. In this article, the critical factors of the road traffic safety in rural areas are reviewed. The critical factors and related information from open data and environmental data are analysed and the meta model is created for processing of road traffic safety data in rural areas. Later on the basis of meta model the Digital Twin is created that could be used for an improvement of the traffic safety on the rural area roads in Latvia. An open data analysis is performed that concludes the opportunities to improve the potential of Latvian traffic accident open data.

Keywords: Road Traffic safety, Digital Twins, Open data, Environmental data.

I. INTRODUCTION

In this article as critical factors (key factors) we will consider such factors that may raise probability of road traffic accidents or probability of higher severity of the accidents if accidents do happen. In the next paragraphs we will list several sources of the information about road traffic accidents in different countries of the world. In case a source of data or report about the traffic safety mentions some factor as an important factor and collects statistics on such factors then we can consider such factor as a critical factor.

In the article [1] critical factors of road traffic accidents are presented in hierarchical structure, see "Fig. 1". This study focuses on finding associations (correlations) from categorical variables which describe road traffic accidents to severity of the accidents and probability of accidents with higher severity.

Correlation rule is the study of correlation between different events. It is a data mining method. In simple terms, correlation rule can produce a type of "A=>B" effect, namely when A happens, B will occur as a result. Association analysis is with the goal of mining the hidden relationship between data, looking for the same event in the correlation of different items [1].

Similar approach with defining critical factors of road traffic accidents is used in many scientific articles, for example in the article [2], where authors formulate critical factors in the following way: "our research data included variables describing the conditions that contributed to the accident and injury severity: characteristics of the accident (month, time, day type, number of injuries, number of occupants, accident type, number of involved vehicles and cause); weather information (prevailing weather conditions and lighting); driver characteristics (age and gender); and road characteristics (pavement width, lane width, shoulder

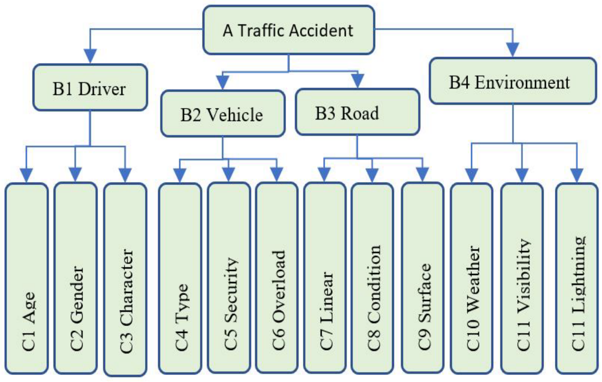

Fig. 1 RTA - critical factors [1].

width, paved shoulder, pavement markings and sight distance)".

We will process the critical factors using our meta model or digital twin with different methods available in the digital twin, for example, statistics and try to distinguish the factors which are more critical than the others based on the open-data we were able to find on the Internet.

II. MATERIALS AND METHODS

In the next table "Tab. 1" we have named the critical factors, that could cause road traffic accidents and raise the severity of road traffic accidents, found from various sources [1], [3], [4], [5], [6], [7], [8]. Most of the critical factors mentioned are traditional ones mentioned already in reports on traffic accidents since the very beginning when analysis of traffic accidents started in the fifties of previous century in the United States.

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2024vol2.8012

TAB. 1: CRITICAL FACTORS OF ROAD TRAFFIC ACCIDENTS

VAR. CAT.	CODE	NAME
ACT: Acc. type	ASC	Angle or side coll.
	HOC	Head-on coll.
	PUC	Pile-up coll.
	FOC	Fixed-object coll.
	ROR	Run off road
	CP	Coll. with pedestrian.
	RO	Roll over
	OT	Other
AGE: Age	TEE	$\leqslant 18$ or with $\leqslant 18$
	YOU	All $[18-25]$
	ADU	All $(25-64]$
	OLD	$\geqslant 64$ or with $\geqslant 64$
	YAA	$[18-25]$ and $(25-64]$
ATF: Atm. fact.	GW	Good weather
	LR	Light rain
	HR	Heavy rain
	OT	Other
CAU: Cause	DC	Driver characteristics
	RC	Road characteristics
	VC	Vehicle characteristics
	OT	Other
DAY: Day	BW	Beginning of week (Mon)
	EW	End of week (Friday)
	WD	Weekday
	WE	Weekend
GEN: Gender	M	Male
	F	Female
LAW: Lane width	THI	$< 3,25$m
	MED	$[3,25-3,75]$m
	WID	$> 3,75$m
LIG: Lighting	DAY	Daylight
	DU	Dusk
	IL	Insufficient
	SL	Sufficient
	WL	Without lighting
MON: Month	WIN	Winter
	SPR	Spring
	SUM	Summer
	AUT	Autumn
NOI: Num. of inj.	[1]	1 injury
	[2]	2 injuries
	[+2]	> 2 injuries
OI: Occup. Inv.	[1]	1 occupant
	[2]	2 occupants
	[+2]	> 2 occupants
PAS: Paved should	N	No
	Y	Yes
	NE	Does not exist or impractical
PAW: Pav. width	THI	< 6m
	MED	$[6-7]$m
	WID	> 7m
ROM: Pav. mark.	DME	Does not exist or was deleted
	DMR	Separate margins of roadway
	SLO	Separate lanes only
	SLD	Separate lanes define road
SHT: Should. type	NE	Does not exist or impractical
	THI	$< 1,5$m
	MED	$[1,5-2,5]$m
	WID	$> 2,5$m
SID: Sight distance	ATM	Atmospheric
	BUI	Building
	TOP	Topological
	VEG	Vegetation
	WR	Without restriction
	OT	Other
TIM: Time	$[0-6]$	$[0-6]$
	$(6-12]$	$(6-12]$
	$(12-18]$	$(12-18]$
	$(18-24]$	$(18-24]$
VI: Vehicles inv.	[1]	1 vehicle
	[2]	2 vehicles
	[3]	3 vehicles

There are many different options of improving human behaviour in drivers position by using modern technologies which are not mentioned in traditional sources of traffic accident data such as usage of mobile phones during the driving, usage of other modern technologies such as infrared automotive night vision of the road which could detect danger from animals or humans on the road during conditions with bad visibility. Nowadays infrared automotive night vision is included as a separate cost option in different models of cars from different car makers with the average price around 2500 EUR [9].

However, according to SAE classification level [10], most of the technologies available today in the cars available to public are released only with Level 1 (driver assistance) or Level 2 (partial automation) driving assistance technologies. This is because each of technology individually usually has several shortcomings, for example, infrared automotive vision does not function very well in heavy rain conditions and conditions with fog.

Lately new developments of self-driving cars are done with the help of LIDAR technology [11]. This technology uses lasers to detect the distance between the car and the nearest object. Or in more complicated case it can construct very accurate real-world 3D model from the laser scans.

However, such solutions are at the moment available only in concept models of the cars. There are predictions on the Internet that first models released with LIDAR technology will have a cost for this option which could exceed 30 000 EUR. LIDAR systems have been used since 1960-ties having initial use cases in space technologies and cartography due to the price of this technology only recently the evolution of technology allowed to start its usage in the industry of self-driving cars [11].

In "Fig. 2" we can see the work principles of LIDAR system [11].

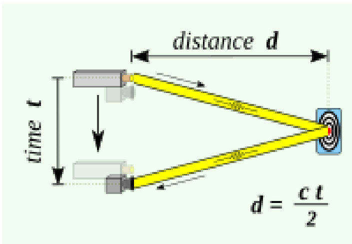

Fig. 2 Work principles of LIDAR system.

In the scope of this work we will focus on our meta model and construction of digital twin for our environment rather than automotive solutions for the cars, but it was worth to mention the options available as automotive solutions as the impact of those solution will only increase on different aspects of traffic safety.

It is important to review data about road traffic accidents on regular basis to find new solutions and options to reduce those events and reduce the severity of those events. We should remember that technology evolves and human behaviour could change during time which means that the topic of road traffic safety will be actual in all times and should be monitored closely.

Knowing such factors of vehicle as geographical location we have technical possibilities to provide feedback to drivers via mobile devices, onboard navigation systems, for example, about dangerous situation on their way. Knowing combination of critical factors which caused the accident we can try to improve the future situation with adjustment of traffic signs and electric lighting in the particular "black spot" or providing the warning to the driver via mobile phone of "black spots" he or she is approaching.

In the "Fig. 3" [4] bellow we can see the data on road traffic accidents per million inhabitants per country during a period of one year. The most fatalities (approx. 130) per million inhabitants do happen in Kazakhstan and Georgia, the least (approx. 20) do happen in Iceland and Norway. Countries may differ by economic situation, geographic location, culture, traditions and many other aspects.

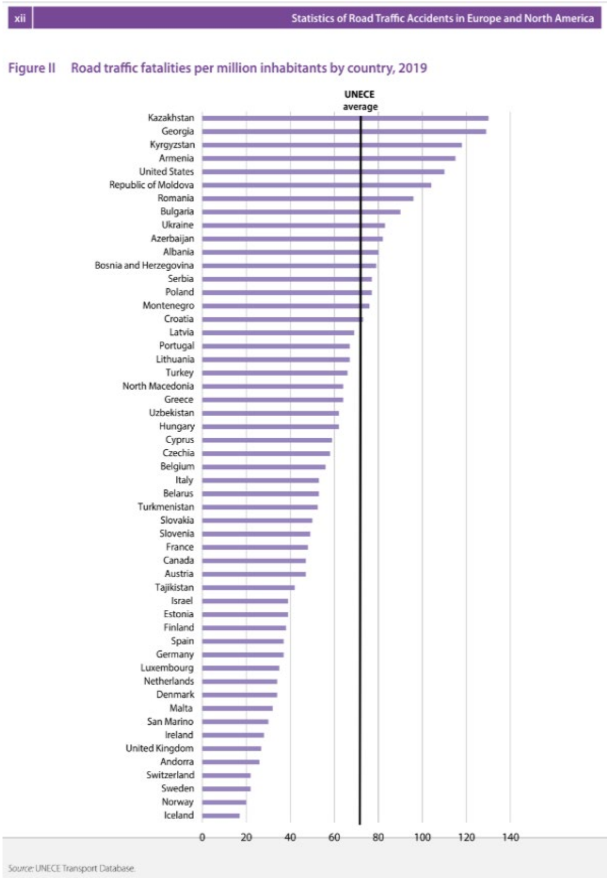

Fig. 3 RTA per million people per country – list.

The graph in "Fig. 3" clearly shows that it is possible to improve the situation with traffic safety on the roads by changing human behaviour with laws and regulations and by changing environment which directly impacts traffic safety.

One can see that neighbouring Baltic countries are in different positions in the graph however we can consider that countries are very similar by the size, the density of people, the rural vs the urban division and many other factors.

We can show the same data as in "Fig. 3" on the map, please, see "Fig. 4" [4] bellow.

In "Fig. 4" it is visible, that it is possible to achieve different results in different countries and human populations based on different set of laws, rules, regulations, culture, traditions. We should keep looking for solutions and continuously monitor the current situation to be able to react accordingly if situation changes towards negative side.

Open data about the Road Traffic safety

Nowadays government institutions over the world have decided to make some part of the data they collect to be available for everyone and place the data on the Internet so those data are available for everyone who has a connection to the Internet and we call those data an open data. Open data is data that is openly accessible, exploitable, editable and shared by anyone for any purpose. Open data is licensed under an open license [12].

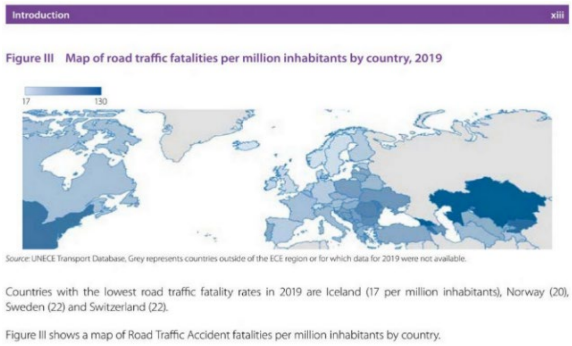

Fig. 4 RTA per million inhabitants per country – map.

One can find open data about the Latvian road traffic safety, accidents and other aspects related to traffic on the Internet resource [13]. This will be one of open data sets we are going to use in the scope of this article.

Results from data analysis from Latvian data set about road traffic accidents could be compared to other data sets from other countries which could help in proving of null hypothesis and statistic's results. For example, data about road traffic accidents from United States are available on the Internet resource Kaggle [14]. This data set was collected as part of independent research by collecting data from all available open data set sources in the United States.

Other way to get data regarding traffic accidents from the United States is to use their government institution NHTSA official report data. The National Center for Statistics and Analysis (NCSA), an office of the National Highway Traffic Safety Administration, has been responsible for providing a wide range of analytical and statistical support to NHTSA and the highway safety community at large for over 45 years [15], [16], [3].

Information from three of NHTSA's primary data systems has been combined to create a single source for motor vehicle traffic crash statistics. The first data system, the Fatality Analysis Reporting System (FARS), is probably the better known of the three sources. Established in 1975, FARS contains data on the most severe traffic crashes, those in which someone was killed. The second source is the National Automotive Sampling System General Estimates System (NASS GES), which began operation in 1988 and ended in 2015. NASS GES contains data from a nationally representative sample of police-reported crashes of all severities, including those that resulted in death, injury, or property damage. The third source is the Crash Report Sampling System (CRSS), which replaced NASS GES in 2016. CRSS is the redesigned nationally representative sample of police-reported traffic crashes [3].

For several counties in Europe open data on traffic accidents could be found on European Commission's official portal [17].

Analysis of open data

Mostly open data are placed on the Internet in format of CSV. A comma-separated values (CSV) file is a delimited text file that uses a comma to separate values. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format. A CSV file typically stores tabular data (numbers and text) in plain text, in which case each line will have the same number of fields [18].

Nowadays very common method of processing open data in CSV files is Pandas library from Python programming language.

Python programming language allows tabular structures from CSV files to be loaded in the memory of personal computer or the server and to be processed there with many different mathematical methods, statistics methods and algorithms. In many cases Python serves as free and open source alternative to such commercial products as MATLAB, Excel, SPSS.

Pandas is an important Python library for data manipulation, wrangling and analysis. It functions as an intuitive and easy-to-use set of tools for performing operations on any kind of data. Initial works for Pandas was done by Wes McKinney in 2008 while he was developer at AQR Capital management. Since then, the scope of the Pandas project has increased a lot and it has become a popular library of choice for data scientists all over the world [19].

In the "Fig. 5" bellow we can see the meta model used in this work to process open data about traffic accidents.

In the "Fig. 5" as the first step we do see data collection from the Internet from the sources like [13]. The next step is data loading into relational database. From relational database data could be prepared to required format and shape using SQL language and later of-loaded to CSV files for further processing with the Pandas library from Python language which provides even further access to different types of algorithms from mathematics to artificial intelligence.

The meta model in "Fig. 5" is simplified and it does not include all the steps and decision points, but it clearly shows that from each step it is possible to go forward and backward in the processing as processing information from CSV files usually requires data adjustment and error corrections.

This meta model could be further extended connecting it to other models, for example, we could use anomaly detection model to exclude noisy data from the data set or we could use open data collected by weather stations to create analysis of traffic accidents considering weather conditions.

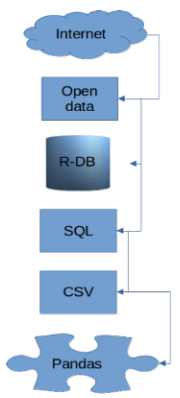

Fig. 5 Steps of the processing of open data.

Digital twins and Road traffic safety

A digital twin is a digital representation of an intended or actual real-world physical product, system, or process (a physical twin) that serves as the effectively indistinguishable digital counterpart of it for practical purposes, such as simulation, integration, testing, monitoring, and maintenance [20].

The first practical definition of a digital twin originated from NASA in an attempt to improve physical-model simulation of spacecraft in 2010 [20]. We can conclude that the digital twins can be utilised in other areas of transportation as well to perform calculations, simulations, integration, testing, monitoring and maintenance.

For the area of road traffic safety, we can consider that the representation of the real world is in the open data collected by the police on traffic accidents, measurements done by sensors of weather stations on the road, digital cartographic data available for the area and even open data from the space stations like Sentinel-2 or Sentinel-3.

The idea of the model of digital twin impacting the physical model (the real-world traffic system), could be implemented, for example, by performing analysis of situation on one road and measuring the effectiveness of different countermeasures for traffic safety improvements before implementing the most effective countermeasures in wider area. Here digital twins can bring benefits for both traffic safety and economical aspects of countermeasure implementation.

For example, we can find out from the data that many traffic accidents in rural areas do happen due to the collisions with wild animals. And if we see that these types of accidents do happen especially frequently on several roads, we can decide to test different countermeasures for preventing the incidents, like building the bridges for animals to cross the roads, implementing more road signs, implementing sensor systems to scare off wild animals. When we notice by digital twin which is the most effective form of countermeasures we can implement that in wider scope. Thus, decisions are made based on data and solutions are tested on digital twin before further investments.

Several implementations of such ultra sound wild life deterring devices where done in Latvia in 2020. In total 1275 such devices are installed along 34.1 kilometres of state roads in Latvia [21], [22].

Another example would be to make weather condition-based information provisioning to drivers on particular road via electric signs or radio in case we know that weather conditions will get worth after some time and drivers could benefit from this information by driving more carefully.

Since NASA implemented the first digital twin model in 2010, the concept of digital twin model has evolved further in science literature frequently mentioning that the feedback from digital twin to physical twin should be provided almost in real time. However, for the scope of this work we will base our understanding of digital twin assuming that the feedback provided to physical twin could be delayed in time due to the country wide scope and nature of this particular topic.

We can take the idea of time relativity here from physics and astronomy where nowadays scientists try to include the idea of human consciousness in understanding of time relativity, so we can consider that there are several time flows and there are "before and after" time flow for each event, the time spent performing all the steps of each event is out of scope of our interests and could be considered as a moment.

In explaining this relativity (the relativity of time), we shall be able to confirm the idea of the two different flows of time, namely "before to after" and "after to before", and finally explain how the idea of "consciousness" fits perfectly in that presentation of the idea of the relativity of time [23].

III. Results and Discussion

Usage of Mann Whitney U test for calculation of key factor criticality

In statistics, the Mann–Whitney U test (also called the Mann–Whitney–Wilcoxon (MWW/MWU), Wilcoxon rank-sum test, or Wilcoxon–Mann–Whitney test) is a non-parametric test of the null hypothesis that, for randomly selected values X and Y from two populations, the probability of X being greater than Y is equal to the probability of Y being greater than X [24].

Some key assumptions for Mann-Whitney U Test are detailed below [24]:

- The variable being compared between the two groups must be continuous (able to take any number in a range – for example age, weight, height or heart rate). This is because the test is based on ranking the observations in each group.

- The data are assumed to take a non-Normal, or skewed, distribution. If your data are normally distributed, the unpaired Student's t-test should be used to compare the two groups instead.
- While the data in both groups are not assumed to be Normal, the data are assumed to be similar in shape across the two groups.
- The data should be two randomly selected independent samples, meaning the groups have no relationship to each other. If samples are paired (for example, two measurements from the same group of participants), then a paired samples t-test should be used instead.
- Sufficient sample size is needed for a valid test, usually more than 5 observations in each group.

Further in the paper we will provide the results of calculations for the Mann-Whitney U Test where we will use the variable of traffic accidents death rate and compare it from two groups to calculate if death rate is higher during night or during day.

Data processing case

The majority of road fatalities in Latvia occurred on rural roads (70%). This percentage is much higher than in the European Union as a whole. The share of fatalities on urban roads on the other hand is lower than the EU average. There are no motorways in Latvia. Over the past ten years, fatalities show a downward trend on both road types in Latvia, the decrease on urban roads was considerably larger than in the European Union [25].

In the "Fig. 6" we do see territory of Latvia divided into Urban and Rural areas [25].

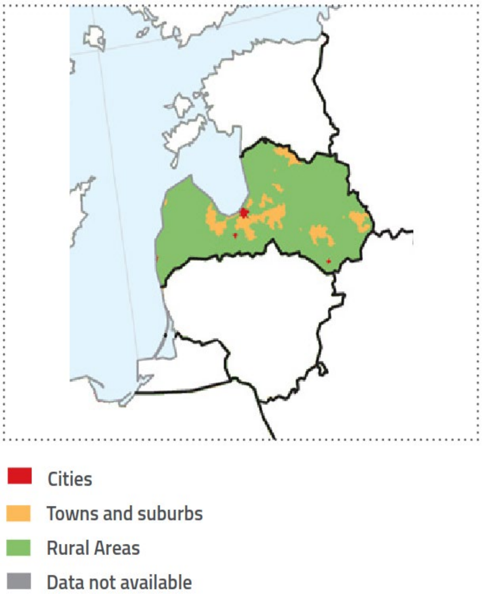

Fig. 6 Degree of urbanization – Latvia.

We can assume that some of the most critical factors causing traffic accidents affect both rural and urban roads in similar way. For example, daylight conditions which affect whole territory of the country at the same time as Latvia has only one time zone and it gets dark almost simultaneously everywhere.

We have collection of open data for traffic accidents in territory of Latvia for almost twenty of years and daylight conditions are one of the factors in the data set "Tab. 2" [13].

ID	ROAD_1	ROAD_KM	ROAD_2	COLUMNn	WEATHER	LIGHT	COLUMNm
0	A0001	67	NaN	...	clear	night	NaN
1	A0001	87	NaN	...	clear	day	NaN
2	A0001	90	NaN	...	cloudy	night	NaN
3	A0001	1	NaN	...	cloudy	day	3.0
4	A0001	0	NaN	...	cloudy	night	3.0
...	...	...	...	...	...	...	...
40983	V1484	1	NaN	...	cloudy	day	NaN
40984	V1484	6	NaN	...	cloudy	day	NaN
40985	V1484	1	NaN	...	cloudy	day	NaN
40986	V1488	4	NaN	...	clear	nightfall	NaN
40987	V1489	5	NaN	...	sunny	day	NaN
[40988 rows x 14 columns]							

Tab. 2. Traffic accidents – Latvia.

Bellow descriptive statistics are provided for the data set visible in "Tab. 2". "Tab. 3" shows descriptive statistics for traffic accidents with victims.

Descriptive statistics : Traffic accidents with victims						
Variable	N	Mean	SD	SE	95% Conf. Interval	
victims	40988.0	0.3008	0.4586	0.0023	0.2964	0.3052

Tab. 3 Traffic accidents – victims.

"Tab. 4" shows descriptive statistics for traffic accidents causing death of victims.

Descriptive statistics : Traffic accidents causing - death						
Variable	N	Mean	SD	SE	95% Conf. Interval	
deaths	40988.0	0.0364	0.2111	0.001	0.0344	0.0384

Tab. 4 Traffic accidents – deaths.

"Tab. 5" shows descriptive statistics for traffic accidents in different daylight conditions.

Descriptive statistics : Daylight			
Variable	Outcome	Count	Percent
daylight	day	25063	61.15
1	night	13417	32.73
2	nightfall	2135	5.21
3	NaN	373	0.91

Tab. 5 Traffic accidents - daylight

Using the data set described previously we calculated Mann Whitney U test to detect if probability P(DEATHS(day)) < P(DEATHS(night)) and we got the following result of Mann Whitney U test:

- (statistic$= $9106797.0, p-value$= $0.002120460965743069)

This proves that accidents which happen during the night time could be more severe in average due to visibility conditions and drivers need to pay attention to this factor as very important one for the safety.

Knowing such data government could decide to change rules for driving, for example, decrease the maximum allowed driving speed on the rural roads with the most accidents and compare the results achieved later. It is worth to mention that this test was performed periodically and the results of Mann Whitney U test returned less p-value with the growth of data set which means that Latvian data set of traffic accidents for such type of calculations does not provide general rules for this type of calculation and the test could be repeated for similar data sets of other countries to gain more evidence to prove the importance of this critical factor.

"Fig. 7" shows graphical representation of frequency of deaths caused by traffic accidents in different day light conditions.

"Fig. 8" a kernel density estimate (KDE) plot. KDE plot is a method for visualising the distribution of observations in a data set, analogous to a histogram. KDE represents the data using a continuous probability density curve in one or more dimensions [27].

And in "Fig. 8" we see the KDE plot of death distribution in different daylight conditions.

As KDE plot has characteristics as histogram we can conclude that data shape for variable deaths is similar in different daylight conditions which as well full-fills the requirement of having data of similar shape to use Mann-Whitney U test for analysis of the data and calculation of probability of impact on one numeric, continuous and dependent variable from other independent categorical variables.

IV. Conclusions

In the scope of writing this article we have achieved several results. We have created the meta model on how to create a digital twin to process the data about traffic accidents. We have created a digital twin which allows us to process the data of

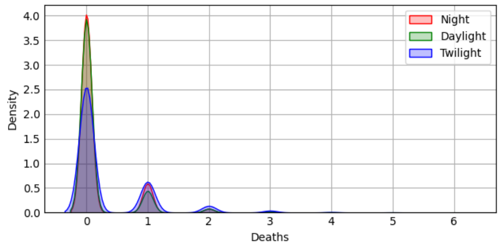

Fig. 8. Deaths KDE distribution in different daylight conditions.

traffic accidents in Latvia and we can use this approach in general as a system which helps to make decisions on how to improve traffic safety in general. The main components of our Digital Twin are:

- CVS files with open data;
- Python scripts and Python Pandas library used for statistical analysis of open data;
- Python scripts used for calculation of moon phase;
- Oracle database in the cloud used for data storage, shared access of data and data processing;
- Oracle SQL Developer used for access of Oracle database in the cloud;
- Oracle SQL loader used for loading large CSV data sets in the database;
- Digital Twin has very good potential for extension with other programming languages and tools.

Using Mann Whitney U test, we have gained statistical evidence that daylight conditions is one of the most critical factors raising probability of traffic accidents and raising severity of traffic accidents on rural area roads in Latvia. Most of the Latvia territory could be considered as rural, as it is visible in “Fig. 6” and most of the accidents with high severity do happen on rural roads so in case we focus on improvement of road traffic safety and especially on reducing accidents with high severity then rural areas could gain the potential benefit out of it.

Using mathematical methods and algorithms we tried to verify if new critical factors could be introduced to classify traffic accidents. We calculated based on exiting algorithms the astronomical phase of the moon for each particular accident and tried to verify if moon phase has some impact on traffic accidents with high severity [28], [29].

In this case we noticed no evidence of moon phase affecting traffic accidents. However, the approach proves that new critical factors could be introduced or not introduced based on the results of calculations gained from our digital twin.

After performing open data analysis of traffic accident data from different countries and available sources. We have noticed opportunities to improve the potential of Latvian traffic accident open data. In case traffic accidents would be registered with geographical location coordinates we could use spatial data analysis and gain better understanding, for example, on "black spot" data and we could perform further calculations in finding out the reasons of "black spots" forming in each particular case.

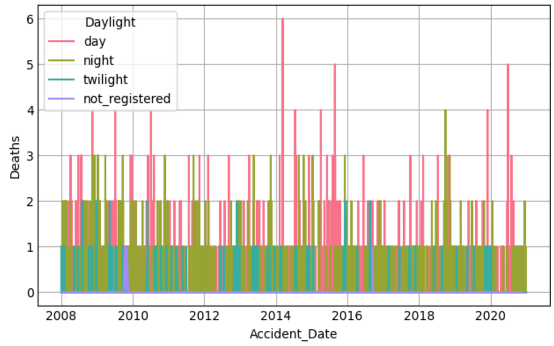

Fig. 7 Deaths different daylight conditions.

V. Acknowledgments

This research was funded by the European Commission, Research Executive Agency grant number 101079206, “Twinning in Environmental Data and Dynamical Systems Modelling for Latvia” (TED4LAT).

VI. References

[1] J. Xi, Z. Zhao, W. Li, Q. Wang, “A Traffic Accident Causation Analysis Method Based on AHP-Apriori,” Procedia Engineering, vol. 137, pp. 680-687, 2016. [Online] Available: https://doi.org/10.1016/j.proeng.2016.01.305. [Accessed: Dec. 2, 2023].

[2] J. Oña, G. López, R. Mujalli, F. J. Calvo, “Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks, Accident Analysis & Prevention,” Procedia Engineering, vol. 51, pp. 1-10, 2013. [Online] Available: https://doi.org/10.1016/j.aap.2012.10.016. [Accessed: Dec. 2, 2023].

[3] National Highway Traffic Safety Administration, National Center for Statistics and Analysis, Traffic Safety Facts 2020. Washington, DC 20590, 2020.

[4] United Nations, Statistics of Road Traffic Accidents in Europe and North America: vol. LVI, 2021.

[5] Inter-American Development Bank, Road safety strategy - IDB, Closing the Gap: Reducing road traffic deaths in Latin America and the Caribean, Action Plan 2010 – 2015, 2010.

[6] OECD, International Traffic Safety Data and Analysis Group, Road Safety Annual Report 2014, 2014.

[7] CSDD, Statistics of Road Traffic Accidents, (Latvian), 2021.

[8] PhD H. Haghpanahan, PhD Prof J. Lewsey, PhD D. F. Mackay, PhD Prof E. McIntosh, PhD Prof J. Pell, PhD Prof A. Jones, PhD N. Fitzgerald, PhD M. Robinson, “An evaluation of the effects of lowering blood alcohol concentration limits for drivers on the rates of road traffic accidents and alcohol consumption: a natural experiment,” 2018. [Online] Available: http://dx.doi.org/10.1016/S0140-6736(18)32850-2 [Accessed: Dec. 2, 2023].

[9] “Automotive night vision,” in A Dictionary of the Internet. Wkipedia, [online document], 2023. [Online]. Available: https://en.wikipedia.org/wiki/Automotive_night_vision. [Accessed Dec. 2, 2023].

[10] “Self-driving car,” in A Dictionary of the Internet. Wkipedia, [online document], 2023. [Online]. Available: https://en.wikipedia.org/wiki/Self-driving_car. [Accessed Dec. 2, 2023].

[11] “Lidar,” in A Dictionary of the Internet. Wkipedia, [online document], 2023. [Online]. Available: https://en.wikipedia.org/wiki/Lidar. [Accessed Dec. 2, 2023].

[12] "Open data," in A Dictionary of the Internet. Wkipedia, [online document], 2023. [Online]. Available: https://en.wikipedia.org/wiki/Open_data. [Accessed Dec. 2, 2023].

[13] CSDD, "Historical data of Road traffic accidents, Latvia," CSDD, 2023. [Online]. Available: https://www.csdd.lv/celu-satiksmes-negadijumi/celu-satiksmes-negadijumu-skaits. [Accessed: Dec. 2, 2023].

[14] S. Moosavi, "US Accidents (2016 - 2021) [Data set]," 2021. [Online] Available: https://doi.org/10.34740/KAGGLE/DSV/3286750 [Accessed: Dec. 2, 2023].

[15] National Center for Statistics and Analysis, 2023. [Online]. Available: https://www.nhtsa.gov/data [Accessed: Dec. 2, 2023].

[16] NCSA Tools, Publications, and Data, 2023. [Online]. Available: https://crashstats.nhtsa.dot.gov/\#!/DocumentTypeList/12 [Accessed: Dec. 2, 2023].

[17] The official portal for European data, 2023. [Online]. Available: https://data.europa.eu/en [Accessed: Dec. 2, 2023].

[18] "CSV (comma-separated values)," in A Dictionary of the Internet. Wkipedia, [online document], 2023. [Online]. Available: https://en.wikipedia.org/wiki/Comma-separated_values [Accessed Dec. 2, 2023].

[19] D. Sarkar, R. Bali and T. Sharma, Practical Machine Learning with Python. AI, Apress Berkeley, CA, 2017.

[20] "Digital twin," in A Dictionary of the Internet. Wkipedia, [online document], 2023. [Online]. Available: https://en.wikipedia.org/wiki/Digital_twin [Accessed Dec. 2, 2023].

[21] "News article (TV 3)," November 22, 2019. [Online], Available: https://zinas.tv3.lv/auto/arvien-biezak-uz-latvijas-celiem-tiek-notriekti-savvalas-zveri-sogad-to-skaits-pieaudzis-lidz-1500/ [Accessed Dec. 2, 2023].

[22] LVC. State road network statistics (2020), (Latvian, English), 2020.

[23] S. Jarvis, "The Relativity of Time," 2019. [Online] Available: https://www.researchgate.net/publication/328738389_The_Relativity_of_Time [Accessed: Dec. 2, 2023].

[24] "Mann-Whitney U Test: Assumptions and Example," in A Dictionary of the Internet. Wkipedia, [online document], 2023. [Online]. Available: https://www.technologynetworks.com/informatics/articles/mann-whitney-u-test-assumptions-and-example-363425 [Accessed Dec. 2, 2023].

[25] European Road Safety Observatory, National Road Safety Profile – Latvia, 2021. [Online]. Available: https://road-safety.transport.ec.europa.eu/system/files/2021-09/erso-country-overview-2021-latvia_en.pdf [Accessed Dec. 2, 2023].

[26] B. Finn, "Latvia - insight paper," 2019. [Online]. Available: https://ruralsharedmobility.eu/wp-content/uploads/2019/08/SMARTA-IP-Latvia.pdf [Accessed Dec. 3, 2023].

[27] Seaborn, "Python package seaborn kdeplot," 2023. [Online]. Available: https://seaborn.pydata.org/generated/seaborn.kdeplot.html [Accessed: Dec. 3, 2023].

[28] J. Meeus, Astronomical Algorithms. Willmann-Bell, 1991.

[29] "Moon / Lunar Phase Algorithm," 2023. [Online]. Available: https://stackoverflow.com/questions/2526815/moon-lunar-phase-algorithm [Accessed: Dec. 3, 2023].