International Journal for Quality Research 16(2) 375–394
ISSN 1800-6450

Matej Babic
Cristiano Fragassa[1]
Dragan Marinkovic
Janez Povh

Article info:
Received 10.06.2021.
Accepted 05.01.2022.

UDC – 004.85
DOI – 10.24874/IJQR16.02-04

MACHINE LEARNING TOOLS IN THE ANALYZE OF A BIKE SHARING SYSTEM

Abstract: *Advanced models, based on artificial intelligence and machine learning, are used here to analyze a bike-sharing system. The specific target was to predict the number of rented bikes in the Nova Mesto (Slovenia) public bike share scheme. For this purpose, the topological properties of the transport network were determined and related to the weather conditions. Pajek software was used and the system behavior during a 30-week period was investigated. Open questions were, for instance: how many bikes are shared in different weather conditions? How the network topology impacts the bike sharing system? By providing a reasonable answer to these and similar questions, several accurate ways of modeling the bike sharing system which account for both topological properties and weather conditions, were developed and used for its optimization.*

Keywords: *Transportation Systems Engineering; Bike-Sharing System (PBS); Artificial Intelligence (AI); Machine Learning (ML); Hybrid Intelligent Systems; Weather Conditions.*

1. Introduction

1.1. Quality of Life

The efficiency of public transport operation has a relevant and direct impact on the individual *Quality of Life*, especially in the case of crowded cities, where, due to a combination of reasons (e.g., traffic, lack of parking), a private car no longer guarantees sufficient mobility to the owner (Berg et al., 2019; Ali et al., 2020).

Ensuring that anyone, regardless of the social status, can easily move between different destinations (e.g., to the places of work, care, feed, study, recreation, etc.) can be deemed as fundamental human right that every public transport policy must assure when addressed to support the development of the individual sphere of life.

Moreover, even if the crucial role of the public transport in the economic and social growth of cities and districts is evident, it has to be always considered including aspects of ecology and safety. Sustainability, in fact, is another important aspect in this equation (Durmić et al., 2020).

However, solving problems related to the public transport often represents a difficult task since the systems complexity and the potential impacts of any change. Hence, the interest in developing suitable approaches to model public transport systems and to predict the effect of each change emerges, including recent methods based on artificial intelligence (AI) & machine learning (ML).

The *Quality of Life* comes back again, finally, when, as in our case, mobility is achieved through healthy and sustainable means of transport, such as bicycles.

Cycling, like any other sport, has a positive effect on the state of the athlete's body, helping to keep it young and healthy.

[1] Corresponding author: Cristiano Fragassa
Email: cristiano.fragassa@unibo.it

Bicycling refers to a cyclic type of physical activity that develops the cardiovascular system, lungs, muscles, and increases the general healthiness of the human body. Medical evidence also exists that cycling can help prevent many serious illnesses (Celis-Morales, 2017) but it also improves overall strength, balance, and coordination.

Hence, emerging together with a more active and healthier lifestyle (Woodcock et al., 2014), even extreme cycling events as marathons (Menaspà et al., 2012) and mountain biking (Weiss et al., 2016) have also become quite popular in recent years.

In short, cycling is healthy, funny, cheap, and practical. It is therefore understandable why bike sales and rents are increasingly spreading everywhere (DeMaio, 2009).

1.2. Bike Rental Services

Bike rental services (BRS) have shown a real explosion in recent years, favored not only by such an increased attention to health and the environment benefits, but also by the larger use of technology in terms of traction (e.g., pedal assist) and control (e.g., geo-localized APPs to locate, book, unlock, pay).

BRS operate in accordance with the recognized technologies for managing rental points by ICT services and allow to optimize all the processes related to the customer service. In this way, modern BRS are able to offer an effective service response to an increasing number of users who prefer bikes to cars, especially during the warmer periods of year (Barbour et al., 2019).

Bike travelling can provide several logistic benefits to users, given the heavy traffic during rush hours, constant traffic jams, endless road works and parking limitations.

At the same time, a good quality bike can be quite expensive while renting usually does not cost much, even less than a bus ticket. Therefore, many city dwellers find it easier to rent bikes as needed instead of buying. Approximately 18 million public bicycles are available in 1608 cities around the world. Thus, Public Bicycle Sharing Programs (PBSPs) have become a prominent feature across city spaces worldwide. In less than a decade, PBSPs have grown from a small number of European cities to include five continents and more than 200 schemes. And these public bikes have already proven their positive effects in terms of

a) increasing cycling in many urban areas,
b) reducing traffic congestions of towns,
c) improving eco-sustainability in mobility,

especially in combination with a rational public transport (IPCC, 2017; IEA. 2014).

Hence, PBSPs' optimization, including bike rental services, can be considered as a priority for administrators of large and small cities. It represents, in fact, an effective way for reducing air pollution, congestion and carbon emissions.

At the same time, most public transport systems can also be seen as business processes to be analyzed, modified, and optimized. When their optimization has been performed, these systems not only become more efficient and profitable, but environmentally friendly and safe too.

1.3. Design parameters

Optimizing the level of quality for bike rental services (BRS) also means to find how comfortable on-call transport should be from the point of view of waiting for transport, travel time and density of entry and exit locations.

The waiting time for transport means the longest time allowed while the user is still ready to wait for transport. This time must be competitive with other existing transport options, such as public transport. According to the experience of systems abroad, the usual average waiting time for transport is somewhere between 10 and 20 minutes.

Another parameter that needs to be determined when designing a system is the longest time a passenger can spend on the journey. This means the difference in time if an individual travels the route from point A to point B on the usual (fastest) route and the time if other passengers join in while driving and the route is extended as a result of the vehicle making a certain detour.

The density of entry and exit stops is the third parameter. Stops are usually virtual locations that are marked in the app, but they do not have to exist physically (such as bus stops). Virtual points are usually set at various intersections, institutions, or other striking locations. It is important that they are recognizable, and that the vehicle can stop there safely. It is recommended that the walking time to the stops is not too long, but on the other hand, the excessive density of entry and exit stops can confuse people and put more strain on the system.

It is also important that bus stops cover areas where there are currently no existing bus stops and increase density where existing stops are placed too infrequently.

Some systems also provide door-to-door service. Such a solution enables higher accessibility of transport, which is especially suitable for the elderly and disabled people. On the other hand, for this level of service with the same number of vehicles, it is necessary to wait longer for transport, and the travel time is also extended (Bodini et al., 2013).

1.4. System modelling

The optimal design and management of BRS should take into consideration a huge array of variables and situations related to topology, population and behavior. In this way it should be possible, for instance, to recognize and optimize aspects as docking stations (e.g., positions, distances) on the territory and bikes (distributions).

However, a different optimization approach would also be possible, without entering in such a level of details: it is exclusively based on the overall number of bicycles and aims to provide an adequate number of bikes to the users accounting for expected situations. This is what is also reported in the present study where the investigation is not intended to modify the pre-existing network structure.

Whatever the approach chosen, and the level of detail used, every sort of optimization brings up the possibility to make use of engineering analysis in order to model the transport systems (van Wee, 2015). In this manner, it is possible to predict the impact of every change in the system and identify the most effective ones in terms of cost, safety, traffic capacity and other factors.

Several recent methods for transport system optimization are based on Information & Communication Technology. This trend is called digital transformation or digitalization of transport and involves several alternative approaches. Among them, Machine learning (ML) (Char et all, 2018) represents one of the most promising approaches, as it involves the concepts such as artificial intelligence (AI) and expert systems.

1.5. Weather conditions

Among the various external conditions to be included in every transport optimization, weather forecast is quite a relevant one, especially because of its direct influence. So, an open problem consists in the estimation of bikes demand with respect to different weather conditions and the answer to this problem is not straightforward. It calls for development of an expert system based on machine learning. And it is by no means a small problem given the potential effects on local transport.

In (Corcoran et al., 2014), the impact of local weather conditions and calendar events on the spatial-temporal dynamics of a PBSP was explored by using novel spatial analytical techniques. In (Rixey, 2013) the effects of demographic and environmental factors near 3 bike sharing stations on the bike sharing ridership levels in three operational US systems were investigated.

Results highlight how the ridership levels is strongly affected by the topological properties of the bike sharing station network, with a robust, statistically significant relationship between the systems and bike use, and independent from other variables, such as demography and age distribution.

Another research (Fuller et al., 2019) has also examined the potential of actions and changes on highly constrained transportation systems and their potential impact on cycling. In the period November 1-7th, 2016, Philadelphia's transit workers went on strike, stopping all transit services in the city. The authors used the strike event as a natural experiment to examine the impact of public transit strikes on the use of Philadelphia's bicycle share program. Two separate approaches were used for this investigation: interrupted time series and Bayesian structural time series models. However, the interdependencies between bicycle sharing and public transportation systems were not totally clear in that analysis.

In (Saberi at al., 2018), authors found that the disruption of public transportation in London increased the total number of bicycles sharing trips by 85% from an average 38,886 to 72,503 trips per day.

Considering similar situations, Lin et al. (2018) proposed a novel Graph Convolutional Neural Network with Data-driven Graph Filter (GCNN-DDGF) model that can learn hidden heterogeneous pairwise correlations between stations to predict station-level hourly demand in a large-scale bike-sharing network.

Partially in line with the mentioned works, this research presents an approach based on expert systems for modeling a bike sharing system, but it uses other machine learning (ML) *tools* with the aim at finding the best way for merging the topological properties of the network with the weather conditions.

2. Materials and Methods

2.1. The location

The present work aims at predicting the variability of bikes demand in the specific case of Nova Mesto (Slovenia) bike-sharing system (GoNM), including in this prediction environmental factors, in order to optimize offer and help reduce the use of cars.

According to the national Statistical Office, in 2019 in the Municipality of Novo Mesto, Slovenia, there were living slightly more than 37.000 inhabitants, on an area of 236 km^2, resulting in a density of 157 inhabitants/km^2. This population makes the Municipality of Novo Mesto the 6th largest one in the country, with a population density higher (+52%) with respect to the national average (103 inhabitants/km^2).

The City of Novo Mesto, approximately 33.3 km^2, is the urban center of the Municipality and the administrative, educational, health, economic and cultural center of the wider region of South-Eastern Slovenia. With its industry, this area is the carrier of the fastest economic development in the region. A strong automotive, pharmaceutical, and cosmetic industry has developed, as well as the insulation materials industry (Krka, Revoz, Adria Mobil, TPV), which also provides private labor silos from elsewhere. The old town is nestled between the Krka riverbeds. From the central part, the city extends outwards along the main city entrances. Settlement outside the city center is quite dispersed. Population density in the municipality is shown in Fig. 1.

As in other parts of Slovenia and EU, the population is rather elderly discouraging the use of bikes: 107 inhabitants are over 64 years old per 100 young people, under 15. This value is less compared to the Slovenian average (of 132) but the gap is closing fast.

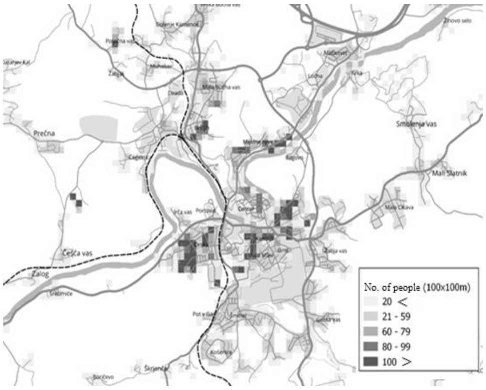

Figure 1. Population density (per 100m^2) in Novo Mesto, Slovenia

2.2. The system

GoNM bike-sharing system was established in 2017 when the city acquired an automatic bicycle rental system, making bike renting easy. It is enough to be registered and, by an appropriate card, it is possible to pick up the bike at 1 of the 14 docking stations and leave there or at another station (Figs. 2 and 3).

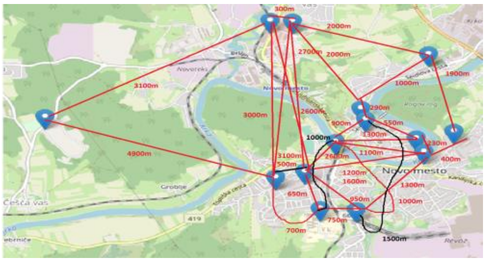

Figure 2. GoNM bike rental system with its 14 stations and distances

Figure 3. A GoNM bike docking station

2.3. Data and information

Anonymous information related to GoNM users (who rent a bike) during the 30-week period since 25th April 2020 till 20th November 2020, were provided by the Municipality of Novo Mesto. Data included approximately 7,000 items with user IDs, Renting & Returning stations, Renting & Returning Time and user's birth year (see Table 1).

Similarly, the total number of bicycles used each week were also provided (see Table 2).

As displayed in Fig. 4, the number of bikes changes a lot during the observation period, from 11 to 132 bikes, with an average value of 30 and standard deviation of 34. This variability in the service demand highlights the need to enable frequent adjustments of the bikes' availability.

Table 1. Sample data of GoNM public bicycle system.

User ID	Renting Station	Returning Station	Renting Time	Returning Time	Birth year
6*1**-20	Ragovska ulica	Avtobusna postaja Topliška cesta	26.05.2020 17:50:14	26.05.2020 17:56:24	1999
6*4**-20	Novi trg	Šolski center Novo mesto	31.03.2020 19:07:23	31.03.2020 19:15:21	1957
6*2**-20	Drska - Šegova ulica	Podbreznik	26.05.2020 16:58:44	26.05.2020 17:17:34	1989
6*3**-20	Ločna-Seidlova cesta	Kandijski most	02.06.2020 11:22:23	02.06.2020 14:05:39	1970

Table 2. Weekly rented bikes vs weather conditions as temperature (T); wind velocity (WV); cloudy (C); relative humidity (RH); air pressure (AP); rainfall (R); sun time (SD).

Week	Bikes	T	WV	C	RH	AP	R	ST
N.	**(n)**	**[°C]**	**[m/s]**	**[%]**	**[%]**	**[hPa]**	**[mm]**	**[h]**
1	28	14.27	1.56	64.86	73.57	984.57	4.49	4.57
2	43	14.30	1.77	25.71	60.00	992.57	1.67	10.54
3	132	14.76	2.49	73.86	72.00	988	3.03	4.66
4	98	15.63	1.78	70.08	74.62	992.15	2.94	4.7
5	108	14.96	1.44	66.63	73.84	995.68	3.83	5.29
6	72	17.54	1.55	52.43	66.86	984.71	1.19	7.59
7	68	18.11	1.73	34.14	73.00	985.29	5.04	8.47
8	39	18.80	1.33	75.14	80.00	986.25	7.63	5.59
9	40	20.74	1.29	42.29	66.29	992.29	2.13	10.2
10	56	21.73	1.39	43.29	75.86	988.43	2.09	8.79
11	112	21.24	1.37	16.71	65.29	990.57	3.06	12.19
12	101	17.60	1.63	48.63	72.63	991.88	4.68	10.05
13	80	21.30	1.00	42.57	77.57	990.14	6.93	8.64
14	58	25.03	1.20	21.71	73.43	990.29	1.33	11.41
15	52	20.36	1.49	67.71	84.71	989.00	6.99	5.21
16	71	23.16	1.03	29.86	79.14	990.14	2.64	8.89
17	90	19.58	1.36	46.8	73.85	990.46	3.90	8.27
18	78	21.91	1.59	49.43	70.43	987.43	0.13	8.33
19	91	17.44	1.27	39.57	80.29	991.00	11.13	6.23
20	77	18.83	1.36	29.57	76.00	994.57	0.00	8.57
21	131	10.72	1.35	66.54	85.03	992.26	4.97	3.98
22	119	15.71	1.30	79.00	83.57	983.83	6.53	4.79
23	78	13.4	1.56	51.57	83.86	982.71	5.51	5.57
24	106	12.79	1.56	53.71	83.43	989.43	9.79	5.66
25	70	8.64	1.50	87.29	86.71	986.43	11.69	1.96
26	47	11.39	1.71	41.57	81.71	995.00	0.07	5.13
27	12	12.13	1.27	64.71	83.71	990.71	4.19	3.96
28	22	10.00	1.09	56.57	80.14	1001.14	0.11	5.06
29	11	5.57	1.05	90.57	91.29	1000.71	0.13	1.13
30	21	5.72	1.03	82.83	93.17	999.67	7.78	1.48

In this regard, it is important to underline that those bikes are meant to be used several times during the same day.

On average, 30 bicycles were able to satisfy the request mobility coming from about 230 users per week, hence nearly 1 ride per day.

Therefore, even at a first glance, it is evident the GoNM bike-sharing system is subject to a very strong variability that, if not properly managed by mitigation actions, leads to a general underutilization of the bikes.

It follows that a system optimization can really provide tangible benefits, but it cannot be done without including further boundary conditions, such as the weather.

Weather data during the same 30 weeks were acquired (from Weather Society Zeus, Slovenia) including: Temperature [°C], Wind Velocity [m/s], Cloudy [%], Relative Humidity [%], air Pressure [hPa], Rainfall [mm] and Sun Duration [h] (see Table 2). A part of this information is shown in Fig. 4 where Temperature, Wind Velocity, Cloudy, Rainfall and Sun Duration trends are displayed against the rented bikes.

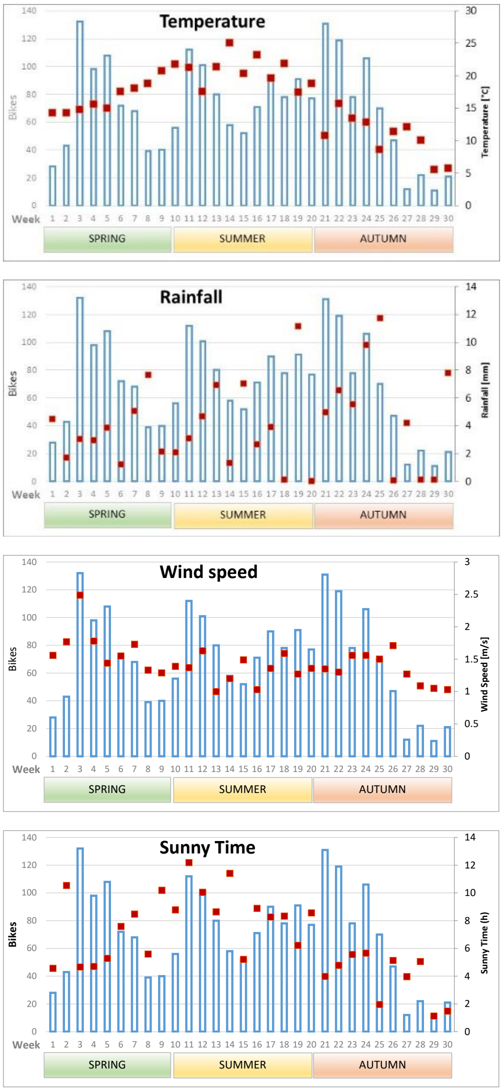

Figure 4. Rented bikes vs temperature, wind speed, cloudy, rainfall and sunny time.

Even from a preliminary data analysis, no clear relationship emerges between rented bikes and weather conditions. A rather weak correlation can be numerically estimated (by Pearson correlation coefficients) between the bikes rented and weather conditions: Temperature (0.25), Wind Velocity (0.4), Cloudy (-0.11), Relative Humidity (-0.21), air Pressure (-0.33), Rainfall (0.21) and Sun Duration (0.20). This is apparently illogical: for instance, it is clear that rain reduces the number of riders. On the other hand, these same correlation values, never negligible, suggest that a certain relationship still exists.

In correlating these data, it is necessary to consider that the effect of weather conditions onto the bicycle rental demand should be examined on a daily (rather than weekly) basis. However, the management of the service (i.e., change in the number of bicycles available in the area) does not allow for a sampling period shorter than a week.

This difference makes machine learning even more interesting. The analysis will provide data about its ability to find patterns among data available on weakly basis (in contrast to data available on daily basis).

2.4. Topological properties

In this research, we use the graph theory (Vecchio, 2017) as a mathematical tool. The graph theory (as a part of the network theory (Saleh et al., 2018)) is a section of discrete mathematics that examines the properties of finite sets and the relationships between their individual elements. In mathematics, graph theory or network theory, a graph or network is a structure amounting to a set of vertices (nodes or points) in which some pairs of the vertices are in relationship with edges. In the public transport system, we denote a station as a node. If a passenger rents a public bicycle at station X and returns it to station Y, there is a directed edge pointing from X to Y. On the other hand, the weight of this edge is equal to the number of cycling records from X to Y, which can show how close the relationship between X and Y is. Hence, a PBN is a directed weighted network. Graphs form the basis upon which semantic networks (Sun & Zhuge, 2018), cognitive maps (Warren et al., 2017), neural networks (Garrido et al., 2014) and economic models (Emmert-Streib et al., 2018) are constructed and transport problems (van Lierop et al., 2018) are solved. In Fig. 5 one example of an artificial network representing the bike renting system with its 14 stations is depicted. The present analysis is strictly in line with previous studies of the authors (Babic et al., 2018 and 2020).

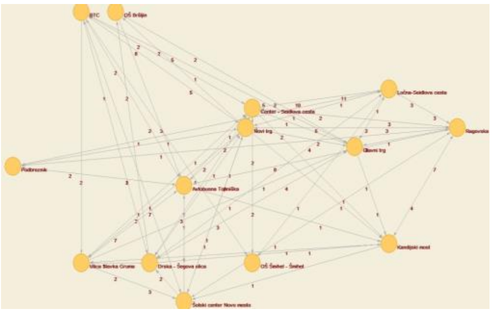

Figure 5. Artificial network representing the bike-sharing system with 14 stations

Four topological properties were considered in order to schematize the transportation systems network: *Degree, Network Density, Betweenness centrality* and *Clustering coefficient* of the network. They are defined in the following.

Degree (D)

The total degree of a node i (k_i^{total}) is equal to the sum of its in-degree k_i^{in} and out-degree k_i^{out}:

$$k_i^{in}=\sum_j a_{ij},\ k_i^{out}=\sum_j a_{ij}, \quad (1)$$

$$k_i^{total} = k_i^{in} + k_i^{out}, \quad (2)$$

where a_{ij} is the adjacency matrix element corresponding to its nodes. Out-degree k_i^{out} represents the number of bikes, rented from station i, that are returned to any destination station. In-degree k_i^{in} represents the number of bicycles, rented from any origin station, that are returned to station i.

Network Density (ND)

The ND measures the territorial occupation of a transport network in terms of km of links (L) per square kilometers of surface (S). The higher it is, the more a network is developed:

$$\text{N}\Delta=\frac{L}{S}. \quad (3)$$

Betweenness centrality (BC)

The BC is an important statistical property of a network, applied in many real-world problems, such as finding border-crossing points that have most extensive traffic or a trade flow. It measures the *accessibility* that is the number of times a node is crossed by the shortest paths in the graph. An anomalous value of centrality is detected when a node has a high betweenness centrality and a low order (degree centrality). The betweenness centrality of a node i is given by the expression:

$$\sum_{s\neq i\neq t}\frac{\sigma_{st}(i)}{\sigma_{st}}, \quad (4)$$

where σ_{st} is the total number of the shortest paths from node s to node t and $\sigma_{st}(i)$ is the number of those paths that pass through i.

Clustering coefficient (CC)

The CC, also known as the *network transitivity*, shows how well the neighbors of a node are connected to each other. For node i with degree k_i, e_{ij} is distinct from e_{ji}. The local clustering coefficient for a directed network is defined as

$$c_i = \frac{|\{e_{jk}: v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i-1)}. \quad (5)$$

Then, the clustering coefficient is the average value of network clustering coefficients, defined as:

$$\mathrm{CC} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{c_i}. \quad (6)$$

2.5. Standard models

Several conventional methods (Gomes et al., 2017) are considered first:

a) *Genetic Programming (GP)*
GP is a collection of methods for the automatic generation of computer programs that solve carefully specified problems by using highly abstracted principles of natural selection. At the beginning, we have some randomly written programs, which represent the initial population. In the next steps, by crossing and selection, we get the next generations. Additional details are available in (Kovačič et al., 2020; Pavlović et al., 2019).

Table 3 lists the used parameters for GP: the population size of organisms, the maximum number of generations, reproduction and crossover probability, the maximum permissible depth in the creation of population and after the operation of crossover of two organisms, the smallest permissible depth of organisms in generating new organisms and tournament size used for the selection of organisms.

Table 3. Input parameters of the Genetic Programming (GP)

Size of the population of organisms	500
Maximum number of generations	100
Reproduction probability	0.6
Crossover probability	0.5
Maximum permissible depth in the creation of the population	8
Maximum permissible depth after crossover between two organisms	10
Smallest permissible depth in generating new organisms	4
Tournament size used for selection of organisms	6

b) *Artificial neural network (NN)*
NN consists of a configurable stratification of nodes (input, hidden, and output layers), connected by artificial neurons, typified by developed weights that modulate signals' crossing. Additional details are available in (Le et al., 2020; Albu et al., 2019).
Table 4 lists the used parameters for NN: learning speed, inertial coefficient, learning set tolerance, test mass tolerance and number of layers.

Table 4. Input parameters of the Artificial Neural Network (NN)

Learning speed	0.7
Inertial coefficient	0.6
Test mass tolerance	0.03
Tolerance of the learning set	0.02
Number of layers	6

c) *Multiple Regression (MR)*
MR analyzes the relationship between one dependent variable and several independent variables by an equation of the following form:

$$Y = b_0 + b_1 \times X_1 + \ldots + b_6 \times X_6 + e, \quad (7)$$

where Y is the dependent variable, the b's are the regression coefficients for the corresponding X (independent) terms, b_0 is a constant (or intercept,) and e is the error term reflected in the residuals. Additional details are available in (Saldana-Perez et al., 2019).

2.6. Hybrid models

Hybrid machine learning models aim at combining the strengths offered by different AI models (Vadlamani et al., 2013). It makes sense as long as these methods offer independent predictions.

In the present research, several numerical combinations of predicted values coming from the three conventional models (MR, GP and NN) were considered and compared with the scope to search for an accurate combination. For instance, hybrid outputs were set, time by time, as minimum, mean or maximum between MR, GP, NN, or between only two of them. Rounding up, down and to the nearest integer operators were involved.

More advanced approaches were also taken in consideration such as, e.g., a 'two out of three' system logic able to identify which estimator was to be eliminated (given a prediction very far from the others and, therefore, probably incorrect), averaging the values of the remaining two.

The comparison (with respect to real values and to the other estimators) was carried out using various immediate criteria, such as:

- error in predicting the total number of rented bikes in the entire period;
- error in predicting the average weekly value of bikes used;
- linear correlation between predictions.

similarly to (Fragassa et al. 2019 & 2020).

3. Results and discussion

GoNM consists of a transport network conveniently representable by an artificial network with 14 nodes (one per station), as already shown in Figure 5. One can see there directed edges between the nodes (stations), and the weights are the number of public bicycles rented or returned. Table 5 reports the topological properties of the network (*D*, *ND*, *BC* and *CC*), also highlighting max. and min. values. For instance, the Network Density (ND) is maximum ($8.2 \cdot 10^{-2}$) at the 2nd week and minimum ($4.3 \cdot 10^{-2}$) at the 27th and 29th weeks.

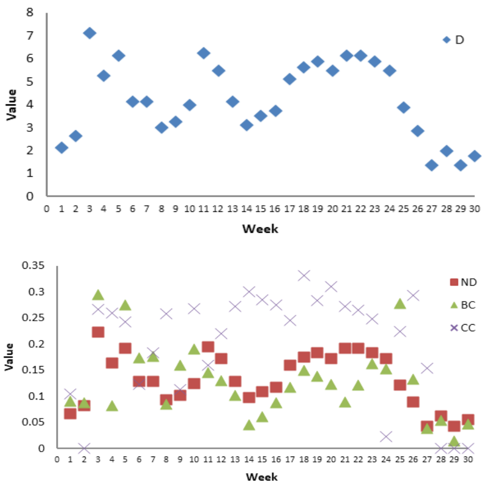

Figure 6. Topological properties.

3.1. Degree

The average out-degree is 64 with the maximum being 132 and the minimum 11. Fig. 7 presents, as an example, the distribution degree and log-degree for the 3rd week. In a log–log plot, the least square estimation was applied to estimate the out-degree distribution, the in-degree distribution, and the regression equations with the coefficient of determination R^2, which showed a fitting effect. All graphs have $R^2 = 1$.

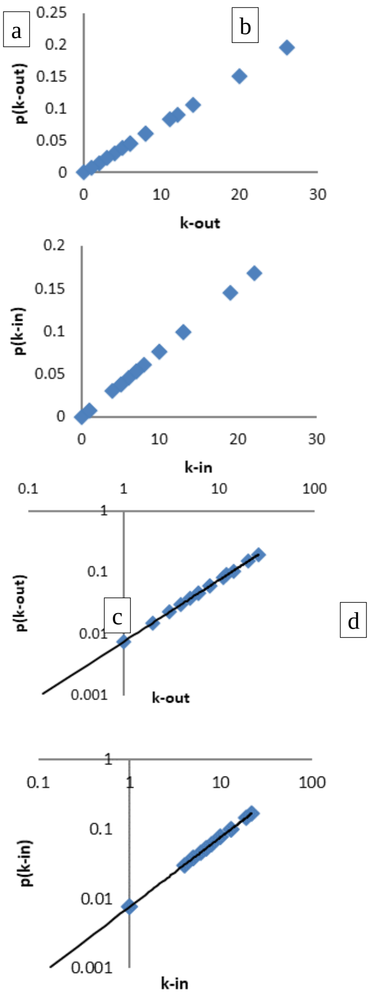

Figure 7. Degree/log-degree distribution of the network: (a) Out-degree; (b) In-degree; (c) log-Out-degree; (d) log-In-degree.

Table 5 also reports *k* coefficient of graph log-Out-degree distribution and log-In-degree distribution of the bike sharing network. The network has the maximum *k* in the 27th and 29th week and the minimum is in the 3rd week.

Table 5. Bike sharing network topological properties.

Week	Topological properties					*k* – coefficient	
N	D	ND	BC	CC		Out	In
1	2.125	0.0664	0.0908	0.1042		0.0345	0.0345
2	2.625	0.0820	0.0879	0.0000		0.0233	0.0238
3	7.125	0.2227	0.2954	0.2667		0.0076	0.0076
4	5.250	0.1641	0.0816	0.2593		0.0133	0.0103
5	6.125	0.1914	0.2759	0.2422		0.0096	0.0093
6	4.125	0.1289	0.1734	0.1228		0.0152	0.0152
7	4.125	0.1289	0.1765	0.1837		0.0147	0.0145
8	3.000	0.0938	0.0849	0.2581		0.0278	0.0278
9	3.250	0.1016	0.1597	0.1136		0.0263	0.0263
10	4.000	0.1250	0.1905	0.2688		0.0179	0.0179
11	6.250	0.1953	0.1452	0.1598		0.0090	0.0090
12	5.500	0.1719	0.1295	0.2204		0.0111	0.0108
13	4.125	0.1289	0.1013	0.2727		0.0133	0.0133
14	3.125	0.0977	0.0454	0.3000		0.0192	0.0192
15	3.500	0.1094	0.0615	0.2856		0.0192	0.0217
16	3.750	0.1172	0.0876	0.2750		0.0164	0.0164
17	5.125	0.1602	0.1172	0.2458		0.0116	0.0114
18	5.625	0.1758	0.1504	0.3311		0.0135	0.0135
19	5.875	0.1836	0.1379	0.2840		0.0115	0.0114
20	5.500	0.1719	0.1225	0.3100		0.0132	0.0132
21	6.125	0.1914	0.0893	0.2731		0.0526	0.0079
22	6.125	0.1914	0.1213	0.2655		0.0093	0.0087
23	5.875	0.1836	0.1619	0.2480		0.0128	0.0128
24	5.500	0.1719	0.1522	0.0235		0.0089	0.0091
25	3.875	0.1211	0.2785	0.2245		0.0143	0.0139
26	2.875	0.0898	0.1333	0.2931		0.0217	0.0208
27	1.375	0.0430	0.0384	0.1545		0.0909	0.0909
28	2.000	0.0625	0.0533	0.0000		0.0500	0.0556
29	1.375	0.0430	0.0137	0.0000		0.0909	0.0909
30	1.750	0.0547	0.0473	0.0000		0.0526	0.0476

3.2. Network Density

A network's density is the number of connections divided by the number of potential connections. The density of a graph is a measure of how many ties between actors exist compared to how many ties between actors are possible, given the graph size (number of nodes) and the graph order (number of links).

As such, the density of an undirected graph is quite simply calculated as the ratio of the observed number of edges (the cardinality of the edge set) to the graph maximum size. Another way to think about density is as giving the probability that, if we were to choose two random nodes in the network, this random dyad will have probability *p* of being connected (as opposed to null). To compute the density of a directed graph, there is no need to multiply the numerator by two, as each edge does single duty.

Fig. 8 shows the network density during the period of 30 weeks. The regression equation is:

$$y = -0.00014x + 0.1536, \qquad (8)$$

with the coefficient of determination of

$$R^2 = 0.0558. \qquad (9)$$

In statistics, a moving average is a calculation used to analyze data points by creating a series of averages of different subsets of the full data set. The reason for calculating the moving average is to help smoothing out the sharing data by creating a constantly updated average price. By calculating the moving average, the impacts of random, short-term fluctuations on the bike sharing of a stock over a specified period are mitigated. The line in Fig. 8 represents a moving average.

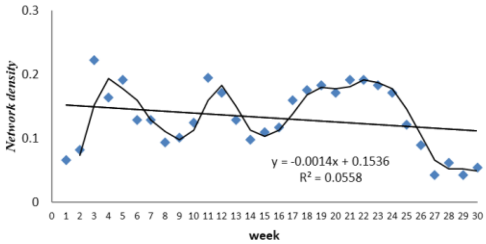

Figure 8. Network density.

3.3. Clustering coefficient

Clustering coefficient is, as previously mentioned, the overall probability for the network to have adjacent nodes interconnected, thus revealing the existence of tightly connected communities. Fig. 9 represents the clustering coefficient over the 30-week period. In this density plot, the regression equation was:

$$y = -0.00018x + 0.2272, \qquad (10)$$

with the coefficient of determination:

$$R^2 = 0.0222. \qquad (11)$$

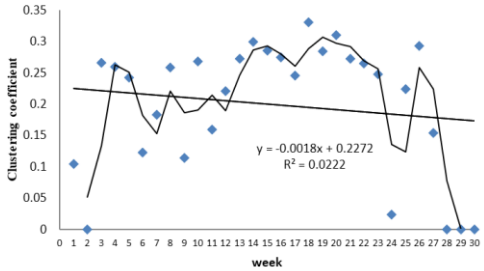

Figure 9. Clustering coefficient trend

3.4. Formula

The formulation of the Genetic programming (GP) and Multiple regression (MR) models in the specific case of the network under investigation is reported in, respectively, Eq. 8 and 9 (see *Appendix A*). For instance, between the four topological properties, the highest impact on the model is due to *ND* (since rather high coefficients as 639.378).

3.5. Predictions

Table 6 reports predictions from different estimation models against real data. Results are also shown in Fig. 10. Moreover, Table 7 also reports predictions for weather forecast, topology properties and expected number of rented bikes during the different periods (spring, summer, and autumn) in terms of min, max and average seasonal values. One hybrid model, between several available, was also here included (explicitly defined as MIN (GP, NN)).

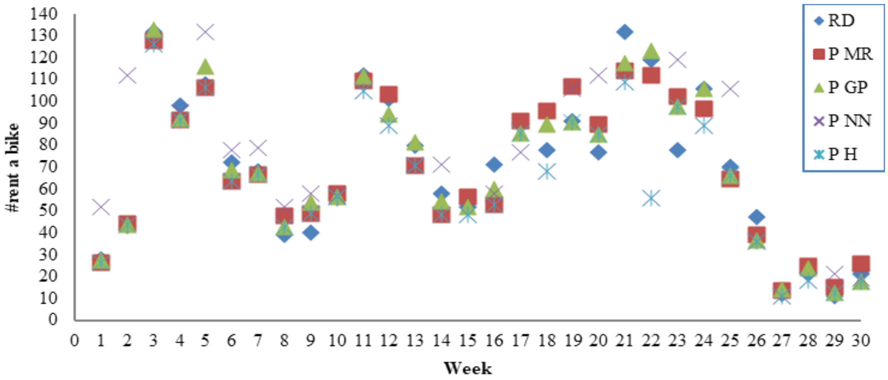

Figure 10. Real vs predicted data.

Table 6. Real data (RD) vs predictions by multiple regression (MR), genetic programming (GP), artificial neural network (NN) and hybrid machine learning models (H).

Week	RD	Predictions				Error			
N		MR	GP	NN	H	MR	GP	NN	H
1	28	26	27	52	26	-7.1%	-3.6%	85.7%	-7.1%
2	43	44	43	112	43	2.3%	0.0%	160.5%	0.0%
3	132	127	132	126	126	-3.8%	0.0%	-4.5%	-4.5%
4	98	91	92	95	91	-7.1%	-6.1%	-3.1%	-7.1%
5	108	106	115	132	106	-1.9%	6.5%	22.2%	-1.9%
6	72	63	68	78	63	-12.5%	-5.6%	8.3%	-12.5%
7	68	66	67	79	66	-2.9%	-1.5%	16.2%	-2.9%
8	39	47	42	52	42	20.5%	7.7%	33.3%	7.7%
9	40	48	53	58	48	20.0%	32.5%	45.0%	20.0%
10	56	57	56	57	56	1.8%	0.0%	1.8%	0.0%
11	112	109	111	105	105	-2.7%	-0.9%	-6.3%	-6.3%
12	101	103	94	89	89	2.0%	-6.9%	-11.9%	-11.9%
13	80	70	81	71	70	-12.5%	1.3%	-11.3%	-12.5%
14	58	48	54	71	48	-17.2%	-6.9%	22.4%	-17.2%
15	52	56	51	48	48	7.7%	-1.9%	-7.7%	-7.7%
16	71	53	60	58	53	-25.4%	-15.5%	-18.3%	-25.4%
17	90	90	85	77	85	0.0%	-5.6%	-14.4%	-5.6%
18	78	95	89	68	68	21.8%	14.1%	-12.8%	-12.8%
19	91	106	90	106	90	16.5%	-1.1%	16.5%	-1.1%
20	77	89	84	112	84	15.6%	9.1%	45.5%	9.1%
21	132	114	117	109	109	-13.6%	-11.4%	-17.4%	-17.4%
22	119	112	123	56	56	-5.9%	3.4%	-52.9%	-52.9%
23	78	102	97	119	97	30.8%	24.4%	52.6%	24.4%
24	106	96	105	89	89	-9.4%	-0.9%	-16.0%	-16.0%
25	70	64	66	106	66	-8.6%	-5.7%	51.4%	-5.7%
26	47	39	36	36	36	-17.0%	-23.4%	-23.4%	-23.4%
27	12	13	14	11	11	8.3%	16.7%	-8.3%	-8.3%
28	21	24	23	18	18	14.3%	9.5%	-14.3%	-14.3%
29	11	14	12	21	12	27.3%	9.1%	90.9%	9.1%
30	21	26	17	19	17	23.8%	-19.0%	-9.5%	-19.0%
Average	70	69	70	74	63	2.2%	0.6%	14.0%	-7.4%

Table 7. Seasonal predictions (in terms of average, maximum and minimal values) of weather conditions - temperature (T); wind velocity (WV); cloudy (C); relative humidity (RH); air pressure (AP); rainfall (R); sun time (SD) - and rented bikes.

		Weather conditions							**Topological properties**				***Bikes***
		T	**WV**	**C**	**RH**	**AP**	**R**	**SD**	**D**	**ND**	**BC**	**CC**	***N***
Spring	**Avg**	16.04	1.70	57.85	71.7	988.6	3.72	6.42	4.31	0.134	0.15	0.17	73
	Max	18.80	2.49	75.14	80.0	995.6	7.63	10.54	7.12	0.222	0.29	0.26	132
	Min	14.27	1.33	25.71	60.0	984.5	1.19	4.57	2.12	0.066	0.08	0.00	28
Summe	**Avg**	21.26	1.335	40.90	73.9	990.0	3.38	9.19	4.42	0.138	0.11	0.24	73
	Max	25.03	1.63	67.71	84.7	992.2	6.99	12.19	6.25	0.195	0.19	0.33	112
	Min	17.60	1.00	16.71	65.3	987.4	0.13	5.21	3.12	0.097	0.04	0.11	40
Autumn	**Avg**	10.61	1.34	67.43	85.2	992.1	5.07	3.87	3.68	0.115	0.10	0.14	61
	Max	15.71	1.71	90.57	93.1	1001.1	11.69	5.66	6.12	0.191	0.27	0.29	131
	Min	5.57	1.03	41.57	80.1	982.7	0.07	1.13	1.375	0.042	0.01	0.00	11

3.6. Accuracy

Several techniques were used to compare predicting models in terms of accuracy as: regression analysis, analysis of variance (ANOVA), Pearson correlation and so on.

For instance, Table 8 reports main statistical properties referring to the application of a regression analysis (*Multiple R, R Square*, etc.) in the case of the multiple regression (MR) model, but parallel evaluations were done for each model. Similarly, Table 9 reports main statistical properties of ANOVA on the same model.

Table 8. Regression statistical properties for the multiple regression model.

Regression Statistics	
Multiple R	0.959
R Square	0.921
Adjusted R Square	0.872
Standard Error	12.326
Observations	30

Table 9. ANOVA statistical properties for the multiple regression model.

ANOVA	***df***	***SS***	***MS***	***F***	***SF***
Regression	11	31920	2901	19	1E-07
Residual	18	2734	152		
Total	29	34654			

Finally, Table 10 outlines the methods' accuracy where the best approximation is offered by GP, followed by MR.

Table 10. Comparing model accuracy.

MR	***GP***	***NN***	***H***
87.39%	91.49%	70.52%	87.88%

3.7. Hybrid vs Conventional Models

The accuracy of hybrid methods was finally investigated. In Table 11 it is possible to find a comparison between predictions offered by conventional (MR, GP, NN) and hybrid models where the last ones were a combination of predictions from the previous ones. In particular, these rules were used in the definition of the hybrid models:

- H_{min} = Min (MR, GP, NN)
- H_{max} = Max (MR, GP, NN)
- H_{mean} = Mean (MR, GP, NN)

Table 11. Comparing accuracy of prediction for conventional and hybrid models.

	Mean	**Total**	**Correl**
RD	**70**	**2111**	**1**
MR	69	2098	0.959
GP	70	2104	0.979
NN	74	2230	0.758
H_{min}	63	1908	0.921
H_{max}	80	2426	0.875
H_{mean}	71	2133	0.947

Other formulations were also considered, but not here reported since they do not add much information to the general discussion.

The comparison was done considering the accuracy in predicting the weekly use of rented bikes and the number of bikes rented inside the whole period. Linear correlation was also used.

In short it is possible to highlight that:

- The conventional methods (i.e., MM, GP, NN) already allow to obtain an excellent estimation. This accuracy can be observed by the mean and total values which differ slightly. The best model (i.e., GP) guesses the weekly average and slightly underestimates (-0.3%) the total number of bikes rented in the period. Even the worst of the three models (i.e., NN) makes an error of less than 6%.
- Since the good accuracy offered by these conventional methods respect to the mentioned reference values, their combination could easily provide a good estimation of the same values. It is therefore necessary to understand the possible added value of hybrid models. In these terms, hybrid models could add better overall correlation alongside the whole period. However, it does not seem to be happening in the present case.

The correlation coefficient of hybrid methods, in fact, does not suggest an improvement in the offered accuracy. However, the use of a hybrid model that averages the values predicted by the other conventional models (H_{mean}) makes it possible to save an excellent accuracy (error <1.5%), also avoiding the risk of selecting of an inappropriate predictive model. This is the reason why its adoption is suggested.

4. Conclusion

One of the main goals of every valid strategy aiming at the development of transport is to increase the sustainability of the transport system and, at the same time, provide a public transport solution that will be accessible to most of the population.

In this sense, transport systems involving the use of shared bikes appear very attractive.

Moreover, promoting active transportation is an important public health objective. Regular cycling stimulates the heart, improves the circulatory system, reduces the risk of cardiovascular disease and stroke, and lowers blood pressure.

Therefore, with the growing complexity of such systems (e.g., increase in the number of bikes, stations, popularity in using bikes), new network analysis methods and tools have to be considered for their optimization.

Application of the concepts of artificial intelligence and expert systems is expected to provide a deeper understanding of the public transport and organizational phenomena.

This paper elaborates the application of conventional methods, including Multiple Regression (MR), Genetic Programming (GP) and Artificial Neural Network, and an additional hybrid machine learning ensemble to modelling the bicycle rental system of the town of Novo mesto in Slovenia. The GP model gives the best results with a very high accuracy, but also the hybrid machine-learning ensemble offered quite solid estimations. In particular, as it was expected, the analysis measured the system as oversized in terms of number of bikes (and, as a consequence, with single little-used bikes). However, in accordance with the general goals of the Municipality for sustainable mobility, every action toward the implementation of a shared transport should be supported since they are much more environmentally friendly than private vehicles with petrol or diesel engines.

Thus, an initial oversizing was preferred with the scope to support spreading this new service among users. The idea is that the user should wait a minimum of time for an available bike, otherwise the service will no longer be attractive.

Funding: The investment is co-financed by European Regional Development Fund.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: Research supported by the Central European Initiative (CEI) within the 'ATC.EVO' transfer of technology project.

References

Albu, A., Precup, R.-E., & Teban, T.-A. (2019). Results and challenges of artificial neural networks used for decision-making and control in medical applications. *Facta Universitatis-Series: Mechanical Engineering, 17*(3), 285-308.

Ali, Y., Mehmood, B., Huzaifa, M., Yasir, U., & Khan, A. U. (2020). Development of a new hybrid multi criteria decision-making method for a car selection scenario. *Facta Universitatis-Series Mechanical Engineering, 18*(3), 357-373.

Babic, M., Calì, M., Nazarenko, I., Fragassa, C., Ekinovic, S., Mihaliková, M., Janjić, M., & Belič I. (2018): Surface Roughness Evaluation in Hardened Materials by Pattern Recognition Using Network Theory. *International Journal on Interactive Design and Manufacturing, 13*(1), 211-219. Doi: 10.1007/s12008-018-0507-3.

Babic, M., Fragassa, C., Lesiuk, G., & Marinkovic, D. (2020) New method for complexity determination by using fractals and its applications in material surface characteristic. *International Journal for Quality Research*, *14*(3) 705-716.

Barbour, N., Zhang, Y., & Mannering, F. (2019). A statistical analysis of bike sharing usage and its potential as an auto-trip substitute. *Journal of Transport & Health, 12*, 253-262.

Berg, J., & Ihlström, J. (2019). The importance of public transport for mobility and everyday activities among rural residents. *Social Sciences, 8*(2), 58.

Bodini, I., Lancini, M., Pasinetti, S., & Vetturi, D. (2013). Techniques for on-board vibrational passenger comfort monitoring in public transport. In *12th IMEKO TC10 Workshop on Tecnical Diagnostics*, (pp. 118-123).

Celis-Morales, C., Lyall, D. M., Welsh P., Anderson, J., Steell, L., Guo, Y., ... & Gill, J. M.R. (2017). Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study. *BMJ, 357* doi: https://doi.org/10.1136/bmj.j1456.

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in Health Care—Addressing Ethical Challenges. *New England Journal of Medicine, 378*(11), 981-983. doi:10.1056/nejmp1714229.

Corcoran, J., Li, T. B., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). Spatio-temporal patterns of a Public Bicycle Sharing Program: The effect of weather and calendar events. *Journal of Transport Geography, 41*, 292-305.

DeMaio, P. (2009). Bike sharing: History, impacts, models of provision, and future. *Journal of public transportation*, 12, 41-56.

Durmić, E., Stević, Ž., Chatterjee, P., Vasiljević, M., & Tomašević, M. (2020). Sustainable supplier selection using combined FUCOM–Rough SAW model. *Reports in Mechanical Engineering, 1* (1), 34-43. doi: 10.31181/rme200101034c

Emmert-Streib, F., Tripathi, S., Yli-Harja, O., & Dehmer M. (2009). Understanding the World Economy in Terms of Networks: A Survey of Data-Based Network Science Approaches on Economic Networks. *Frontiers in Applied Mathematics and Statistics, 4*(37). doi: 10.3389/fams.2018.00037.

Fragassa, C., Babic, M., Bergmann, C. P., & Minak, G. (2019). Predicting the tensile behaviour of cast alloys by a pattern recognition analysis on experimental data. *Metals, 9*, 557.

Fragassa, C., Babic, M., Pavlovic, A., & do Santos, E. D. (2020). Machine Learning Approaches to Predict the Hardness of Cast Iron. *Tribology in Industry, 42*(1), 1-9.

Fuller, D., Luan, H., Buote, R., & Auchinclossd, A. H. (2019). Impact of a public transit strike on public bicycle share use: An interrupted time series natural experiment study. *Journal of Transport & Health, 13*, 137-142.

Garrido, C., De Oña, R., & De Oña, J. (2014). Neural networks for analyzing service quality in public transportation. *Expert Systems with Applications, 41*(15), 6830-6838.

Gomes, M., Silva, F., Ferraz, F., Silva, A., Analide, C., & Novais, P. (2017). Developing an Ambient Intelligent-Based Decision Support System for Production and Control Planning. In *Intelligent Systems Design and Applications*, Springer: Berlin, Germany, pp. 984-994.

IEA. (2017). *CO2 Emissions from Fuel Combustion 2017—Highlights*. IEA: Paris, France.

IPCC. (2014). *Summary for Policymakers.* IPCC: Geneva, Switzerland, ISBN 9789291691432.

Kovačič, T., Kovačič, M., Ovsenik, R., & Zurc, J. (2020). The impact of multicomponent programmes on balance and fall reduction in adults with intellectual disabilities: a randomised trial. *Journal of intellectual disability research, 64*(5), 381-394.

Le, L. T., Lee, G., Park, K.S., & Kim, H. (2020). Neural network-based fuel consumption estimation for container ships in Korea. *Maritime Policy & Management, 47*(5), 615-632.

Lin, L., He, Z., & Peeta, S. (2018). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transportation Research Part C: Emerging Technologies, 97*, 258-276.

Menaspà, P., Rampinini, E., Bosio, A., Carlomagno, D., Riggio., M., & Sassi, A. (2012). Physiological and anthropometric characteristics of junior cyclists of different specialties and performance levels. *Scandinavian Journal of medicine & science in sports, 22*(3), 392-398.

Pavlović, M., Nikolić, V., Simonović, M., Mitrović, V., & Ćirić, I. (2019). Edge detection parameter optimization based on the genetic algorithm for rail track detection. *Facta Universitatis-Series Mechanical Engineering, 17*(3), 333-344.

Rixey, R. (2013). Station-level forecasting of bike sharing ridership: Station network effects in three U.S. systems. *Transportation research record, 2387*, 46-55.

Saberi, M., Ghamami, M., Gu, Y., Shojaei, M.H., & Fishman, E. (2018). Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. *Journal of Transport Geography, 66*, 154-166.

Saldana-Perez M. (2019). Miguel Torres-Ruiz, & Marco Moreno-Ibarra, Geospatial Modeling of Road Traffic Using a Semi-Supervised Regression Algorithm. *IEEE Access, 7*, 177376-177386, doi: 10.1109/ACCESS.2019.2942586,

Saleh, M., Esa, Y., & Mohamed, A. (2018). Applications of Complex Network Analysis in Electric Power Systems. *Energies, 11*(6), 1381. doi:10.3390/en11061381.

Sun, X., & Zhuge, X. (2018). Summarization of Scientific Paper through Reinforcement Ranking on Semantic Link Network, *IEEE ACCESS*.

Vadlamani, R., Nekuri, N., & Mayank, P. (2013). Hybrid classification and regression models via particle swarm optimization auto associative neural network based nonlinear PCA. *International Journal of Hybrid Intelligent Systems, 10* (3).

van Lierop, D., Badami, M. G., & El-Geneidy, A. M. (2018). What influences satisfaction and loyalty in public transport? A review of the literature. *Transport Reviews, 38*(1), 52–72.

van Wee, Bert. (2015). Viewpoint: Toward a new generation of land use transport interaction models. *Journal of Transport and Land Use*, *8*(3).

Vecchio, F. (2017). Small World architecture in brain connectivity and hippocampal volume in Alzheimer's disease: a study via graph theory from EEG data. *Brain Imaging and Behavior*, *11*(2), 473-485. doi:10.1007/s11682-016-9528-3.

Warren, W. H., Rothman, D. B., Schnapp, B. H., & Ericson, J. D. (2017). An empirical articlethat outlines the cognitive-graph theory alternative to cognitive-map theory.

Weiss, F., Brummer, T., & Pufal, G. (2016). Mountain bikes as seed dispersers and their potential socio-ecological consequences. *Journal of Environmental Management, 181*, 326-332.

Woodcock, J., Tainio, M., Cheshire, J., O'Brie, O., & Goodman, A. (2014). Health effects of the London bicycle sharing system: Health impact modeling study. *British Medical Journal*, *348*, g425.

Matej Babic
Faculty of Information Science in Novo mesto,
Novo mesto, Slovenia
babicster@gmail.com
ORCID 0000-0002-5446-5087

Cristiano Fragassa
Department of Industrial Engineering, Alma Mater Studiorum University of Bologna, Bologna, Italy
cristiano.fragassa@unibo.it
ORCID 0000-0003-0046-8810

Dragan Marinkovic
Department of Structural Mechanics, Institute of Mechanics, Technical University of Berlin, Germany
dragan.marinkovic@tu-berlin.de
ORCID 0000-0002-3583-9434

Janez Povh
Faculty of Mechanical Engineering, University of Ljubljana, Novo mesto, Slovenia.
janez.povh@fs.uni-lj.si
ORCID 0000-0002-9856-1476

Appendix A

$$Y = 9.08502ND(-1.51133 + CC + ND + SD + TT + 2WV + ND(BC + C(ND + WV)) + ND\left(4.13212 + C + (BC + C)ND + \frac{SD + C\,ND\left(C + WV + \frac{0.110071\,AP}{TT + ND + CC - 4.13212}\right)}{TT}\right) + \frac{AP}{C + SD + Q + ND\,SD\left(ND + ND\,WV(C + ND(2C + ND + 3WV))\right)} - CC\left(TT + C\,ND\left(ND + WV\left(WV + ND^2(2C\,ND + ND(CC + TT + WV - ND))\right)\right)\right))$$

$$Q = \frac{0.110071\,AP}{(ND+WV)\left(ND\,(BC+CC\,ND)+ND\,RH\,(ND+TT+2WV+BC\,(ND+WV))-\frac{0.110071AP\left(BC+C+WV+\frac{AP}{(9.08502+BC)(CC+ND+TT+WV-4.13212)}\right)}{CC+TT+WV}\right)} \quad (8)$$

Y = - 406.6192311+ 0.00076234×D + 639.3785287×ND - 46.72088091×BC + 9.015791733×CC -0.747814175×T + 4.569795467×WV + 0.236921533×C - 0.043525382×RH + 0.372085427×AP + 1.012797827×R + 2.716867132×SD (9)