A Model for the Quantification of Running and Cycling Safety

JOÃO MIGUEL FREITAS PALMEIRA A73864 , JOSÉ LOPES RAMOS A73955 , RAFAEL MACHADO SILVA A74264, Universidade do Minho, Portugal

Indroduction
Objective
How we did it
Results e conclusoes
Application Talk

1 INTRODUCTION

With the development of information technologies, mainly with the increasing search for applications of Machine Learning in everyday life a wide range of areas previously closed to software developers have opened up with this advent, providing room to brew new and exciting ideas. One of such innovative projects is the one we are presenting in this paper, *A Model for the Quantification of Running and Cycling Safety* which integrates the capacity to learn that we can find in Artificial Intelligence and the need for a way to advise vulnerable road users about the current and future safety conditions of going outside to do sport activities.
This need was apparent to us when we started researching the number of incidents that occurred while someone was simply outside taking a stroll or riding a bicycle. As we can see in the PORDATA Database[1] in 2018 only, 5.652 people were run over on roads across Portugal. In order to reduce this alarming number, we decided to works toward creating a simple yet useful application that can inform it's users about how safe it is to practice sports outdoors at that moment.
The safety value provided to each user will be calculated using an algorithm developed with Artificial Intelligence, namely Neuronal Networks, which, after receiving different types of data, such as traffic flow and weather predictions, as well as pollution particle values collected hourly through sensors located in the city of Braga, will apply them as input values in a trained Neural Network returning its output as a calculated safety value based on the data provided.
With this in mind is easy to see that information and the way it is handled is paramount for a swift and prestigious approach to not only this but any Machine Learning problem. Understanding all the features regarding the data collected is, therefore, essential in order to understand how to tackle the problem, and with scientific research we were able, not only extract to the most pertinent information that could help us calculate the safety value of associated to the route the user will take, as well as the relationship between these variables which also affects said value and allows us to better prepare the data before inserting it to train the Neural Network.
The approach we took regarding this process was CRISP-DM (*Cross Industry Standard Process for Data Mining*). First, we dedicated to understand the process and build up as much knowledge as we thought needed about the problem of the project. We then reached the entities that could provide us with the data required to calculate safety levels and proceeded to analyze and prepare the information it contained. After having attained a sufficiently satisfying result we reevaluated it and decided which modeling and training technique should be used to deal with the information. After deliberating we settled down with Neural Networks and started the process of training and fine-tuning the data that was to be feed to the Networkuntil we reached a state where we had a performance we could be happy with.
In the scope of this project, we hope to develop and provide everyone in Braga an application that will be able to help them in deciding if going for a walk or a bicycle ride is safe or not. We plan to achieve this effect by quantification the safety of the route they would be planning to take for a given activity. Of course, we also hope that given a high enough request for this application in this small test site we could expand its area at a national level an create a business model based on this project.

2 DATA ANALYSIS

The data-sets provided for this project contained information about the city of Braga captured through the use of sensors owned by public entities who lent us for the development of this project. These data in each dataset were, respectively Trafic Flow, Pollution, Weather Conditions and Trafic Incidents.
We will take some time explaining each data-set, as well as the information it contained. Some of the information proved foreign to us so we had to investigate to better understand it and start processing the data into a configuration we agreed upon was practical and well defined. We also researched the implications of certain features to better ascertain which ranges of values had a certain danger level and possible connections between features of different data-sets. This way we can attribute a relative value of safety to an indicated feature which will then help us train our Neural Network.

The Pollution data-set offers us an overview of the volume concentration values of a few particles that, in sufficiently high concentrations, could represent a health risk for anyone going either for a run or a cycling trip. From this data set we realized that the atmospheric particle concentration conditions can not only create a dangerous environment, they may also be an aftereffect of another situation retreated in a different data set. With this in mind, we found necessary to invest time in dedicated research in this particular theme, focusing of course in the particles that we extracted from the provided source.

The first type of pollution researched was Ultra-Violet Radiation. Even though it belongs to common sense that this kind of pollution is in indeed very harmful, so much that at times where it is unusually strong it is customary for national TV to emit a safety warning

Author's address: João Miguel Freitas Palmeira A73864 , José Lopes Ramos A73955 , Rafael Machado Silva A74264, Universidade do Minho, Braga, Portugal, a73864@alunos.uminho.pt,a73855@alunos.uminho.pt,a774264@alunos.uminho.pt.

advising everyone to avoid sunlight as much as possible and apply sunscreen to protect themselves, we wanted to be more thorough. Excessive ultraviolet radiation can cause various diseases at skin level, such as skin cancer known as Melanoma. This type of cancer develops from melanocytes, which are melanin-producing cell, and has more prevalence on the skin, although it may be more rarely seen in the mouth, intestines or eyes. Also, it can also happen that the human body's immune system can suppress its proper functioning and the functioning of the skin's natural.[2] With such risks in line, we put hands at work and found the exact thresholds of safety for this particular pollution to prepare our data so that it can understand and better account for this health risk in its algorithm.

UV Range	Risk
0.0–2.9	Low
3.0–5.9	Moderate
6.0–7.9	High
8.0–10.9	Very High
11.0+	Extreme

Table 1. Risk of damage from exposure to Ultra-violet radiation for an average adult

Afterward, we researched Carbon Monoxide pollution. This is a colorless and odorless gas that, when inhaled, prevents our blood cells, mainly responsible for oxygen transportation, from linking with the oxygen molecule thus not allowing for the efficient oxygenation of vital organ tissue which may cause symptoms like headache, dizziness, weakness, vomiting, chest pain and confusion. If Carbon Monoxide inhalation occurs for a prolonged interval of time it can result in loss of consciousness, cardiac arrhythmia, epileptic seizures, possibly leading to sequels like tiredness, memory problems, and motor problems, and even ultimately resulting in death. In all practicality a small amount is not usually harmful, however, it results in intoxication when the concentrations of carbon monoxide in the blood surpass a certain threshold as exposed in the graphic XXX. It is also worthy to note that Carbon Monoxide disappears from the blood after several hours.

Concentration	Symptoms
0.0035%	Headaches and dizziness within 6 to 8 hours of continuous exposure
0.01%	Light headache within 2 a 3 hours of exposure.
0.02%	Mild headache within 2 to 3 hours of exposure. Loss of judgment.
0.04%	Frontal headache within 1 to 2 hours.
0.08%	Nausea and seizures in 45 minutes. In-sensitiveness within 2 hours.
0.16%	Headache, increased heart rate, dizziness and nausea in 20 minutes. Death in less than 2 hours.
0.32%	Headache, dizziness and nausea in 5 to 10 minutes. Death in 30 minutes.
0.64%	Headache and dizziness in 1 to 2 minutes. Convulsions, respiratory arrest and death in less than 20 minutes.
1.28%	Unconsciousness after 2–3 breaths. Death in less than 3 minutes.

Table 2. Effects of different amounts of CO on the body of an average adult human.

Most cases of intoxication originate from the inhalation of carbon monoxide produced in combustion heaters or kitchen equipment and motor vehicles powered by fossil fuels, which happens to be a factor of direct contact with the runners and cyclists we are hoping to safeguard.[3]

Regarding Nitrogen Dioxide, we found that it exists in the air mainly as a residue of fossil fuel consumption by cars being expelled with other gases. Besides, it can also be expelled to the atmosphere as a byproduct of processes used in certain types of power plants. The presence of this molecule on the environment can be prejudicial for the user's health. Its inhalation may not only irritate the respiratory tract it may also aggravate respiratory illnesses, particularly asthma, leading to respiratory symptoms (such as coughing or difficulty breathing). In situations where longer exposures to high concentrations of NO2 occur it may bring the development of asthma and potentially even increase our body's susceptibility to dangerous respiratory infections, being particularly alarming when affecting people who suffer from asthma, children or the elderly. In the table below we can see the values at which NO2 concentration becomes prejudicial to humans

Concentration (ppb)	Air Quality	Considerations
0-50	Good	No health impacts are expected when air quality is in this range.
51-100	Moderate	Individuals extraordinarily sensitive to nitrogen dioxide should consider limiting prolonged effort outdoors.
101-150	Unhealthy for sensitive groups	The following groups should limit prolonged effort outdoors: People with lung diseases, such as asthma. Children and older adults.
151-20	Unhealthy	The following groups should avoid prolonged efforts outdoors: People with lung diseases, such as asthma. Older children and adults. All others should limit prolonged effort outdoors.
201-300	Dangerous	The following groups should avoid all effort outdoors: People with lung diseases, such as asthma. Older children and adults. All others should limit the effort outdoors.

Table 3. Effects on humans of nitrogen dioxide in air. [4]

Particle matter 10 represents a serious and somewhat difficult to prevent health hazard as its is indeed their diameter, represented by the number in their name as micrometers, that enables them to bypass some of our body's defenses, such as the hairs in our noses, and reach our lung's alveoli, even penetrating our bloodstream.[?] As such exposure to these particles can affect the lungs and the heart. After researching several scientific studies we have managed to link it to a variety of problems as cardiovascular diseases, aggravated asthma, and lung cancer. Below we indicate the safety threshold values researched for this substance.

Reference Period	Threshold	Tolerance
1 day	50 mg/m^3	50%
1 civil year	40 mg/m^3	20%

Table 4. The referenced limit values of PM10 established.

Sulfur dioxide is a colorless gas with a pungent odor that can be identified by a tingle sensation and a sharp smell. The main source of sulfur dioxide in the air is an industrial activity that processes materials that contain sulfur, eg: the generation of electricity from coal, oil or gas that contains sulfur. Also, it is believed that around 80% of sulfur dioxide is due to the incomplete burning of fossil fuels.[?] Sulfur dioxide affects human health when it is breathed in. As it is an irritating gas it affects the mucous membranes found on the nose, throat, and airways causing coughing, wheezing and shortness of breath. The effects of sulfur dioxide are felt very quickly and in high enough concentrations can cause acute and chronic effects on human health. The simultaneous presence of SO2 and particles in the atmosphere can aggravate cardiovascular problems. Below we indicate the safety threshold values researched for this substance.

Concentration (mg/m^3)	**Air Quality**
0-19	Good
20-39	Moderate
40-364	Slightly unhealthy
365-800	Unhealthy
800+	Dangerous

Table 5. Air quality depending on the level of sulfur dioxide.

Ground-level ozone is a colorless gas and has a characteristic spicy odor, which is generally detectable by humans at concentrations of 0.02 and 0.05 ppm. Ground-level ozone is created by chemical reactions between oxides of nitrogen and volatile organic compounds in the presence of sunlight. Cars and gasoline-burning engines are large sources of these compounds. As such, it is normally on hot, sunny days, that ozone levels can reach very high concentrations. Inhaling this gas can be harmful causing inflammation and damage to the cells that line your lungs. Even at relatively low levels can cause negative effects on health. The table below presents values that are found limits widely accepted in the community:

Reference Period	**Threshold**
8 hours a day 5 days a week	0,1 ppm
15 minutes (short time exposure)	0,3 ppm

Table 6. The referenced limit values of Ground-level Ozone established.

The most simple data-set was composed only of three columns, city_name, speed_diff, and creation_date. While the city name and creation date were self-explanatory, being the city where each measurement was taken and the date of said measurement respectively. The speed_diff is a parameter that should be taken into account when processing data due to the information it provides in the face of traffic at the time the value was captured. This parameter is the difference in speeds between the maximum speed that cars can reach in a traffic-free scenario and the speed at which cars are driving at the moment. The higher the value, the greater the difference between both indicating, at high values of that vehicles are moving more slowly.

The next data-set being studied recorded information related to traffic incidents, recorded hourly, in the city of Braga. As we started taking note of each feature present we realised that this one would prove to be complex to analyse for previous treatment as it contained a vast amount of information, as well as a very diverse set of features, each one with its allotted quantity of different possibilities. As we have studied, it's often that an overwhelming amount of unneeded information can deviate the learning process of any Synthetic Intelligence algorithm, so we started by reviewing and set aside a few features with little to none importance, such as city_name, in which all rows presented Braga and seq_num, which remained 1 in all cases. In addition there were some other features, which although presented some variance were promptly unconsidered because, either they did not present reasonable attractiveness in the scope of the project we pretended to develop or another feature already covered their information values. They were from_road, to_road, affected_roads and cause_of_incident respectively. The features from_road, to_road, affected_roads presented information regarding the particular streets in Braga were the the incident occurred, were it ended and the roads that were affected. As we pretended only to capture a general idea of the danger any VRU's would be putting themselves on by going outside at any given moment in the city of Braga they were not needed. Regarding cause_of_incident the information it provided would be deemed irrelevant in face of the features that we decided were important for the analysis of this particular information. This said features, which were to compose our selection , proved to allow for an efficient and in context treatment of the data this data-set contained. Description, indicates the type of incident presenting high importance for risk assessment for the VRUs, as descriptions may indicate different situations, and by consequence different levels of danger for the VRU.

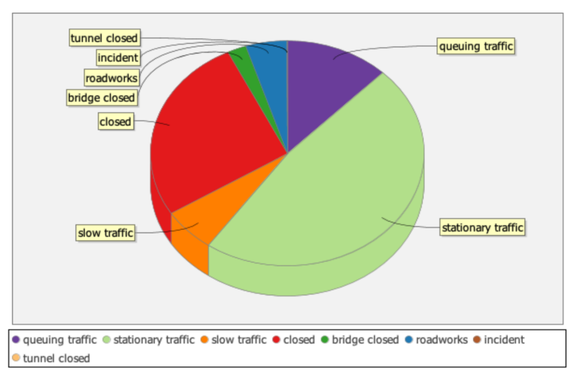

Fig. 1. Incident Description

Incident_category_desc, further enables us to understand the influence the incident had on the road, a road jam is not as meaning full as an incident that may have caused the road to be closed.

Magnitude_of_delay_desc, indicates the time delay that an accident causes. It's easy to understand that in most cases a very small incident will be promptly resolved causing very little delay, while a more severe one, which can lead to our VRU's being unsafe, can understandably mean higher risk.

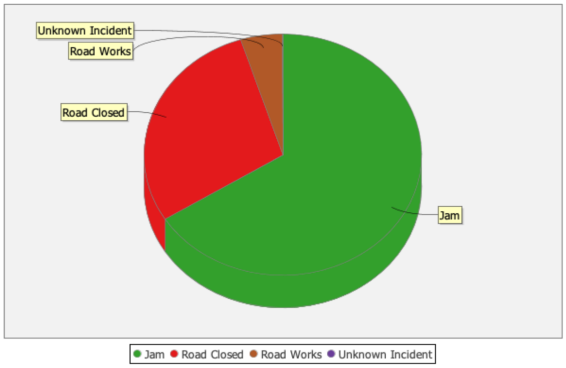

Fig. 2. Incident category

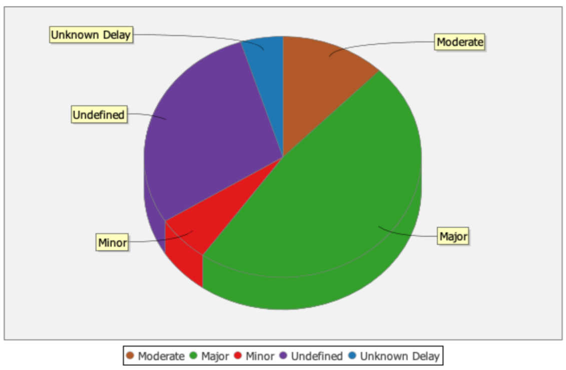

Fig. 3. Incident Magnitude

Length_in_m, stands for the distance of the line the respective incident causes. As a large line means a lot of traffic, and generally very eager drivers to get faster to their objective, even sometimes neglecting road rules, directly endangering everyone of our VRUs. Delay_in_s Should i put or should i not XP

3 DATA EXPLORATION

4 DEVELOPMENT AND TRAINING

5 PLATFORM

6 CONCLUSION

7 REFERENCES

REFERENCES

[1] https://www.pordata.pt/Municipios/Peões+atropelados+total+e+mortos-233?fbclid=IwAR0ruIf_coIbTJhT8l9aYMi39Y_zWn1D4OnkNCr1aKGWcx2ZnIi0163-T54
[2] Health problems of Overexposure to the Sun. URL https://www.epa.gov/sites/production/files/documents/healtheffects_1.pdf
[3] Carbon monoxide poisoning. URL https://en.wikipedia.org/wiki/Carbon_monoxide_poisoning
[4] Air Quality Guide for Nitrogen Dioxide. URL https://www3.epa.gov/airnow/no2.pdf