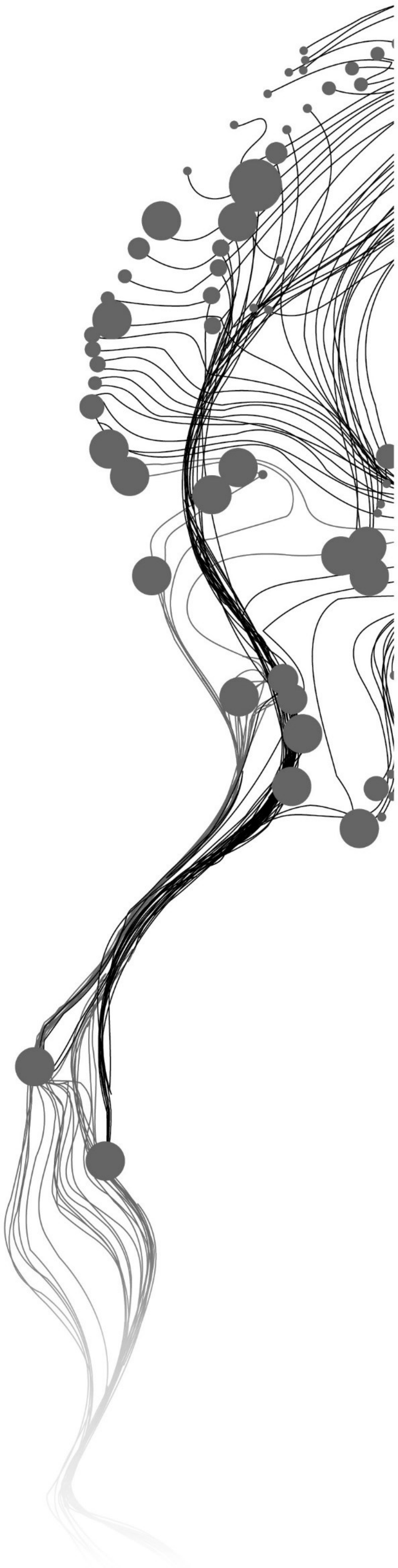
THE RELATION BETWEEN STREET PATTERN AND TRAFFIC CONGESTION, AN INVESTIGATION THROUGH MACHINE LEARNING APPROACH

M. TSAQIF WISMADI

June 2022

SUPERVISORS:
Dr. Jon Wang
Dr. Fran Meissner



THE RELATION BETWEEN STREET PATTERN AND TRAFFIC CONGESTION, AN INVESTIGATION THROUGH MACHINE LEARNING APPROACH

M. TSAQIF WISMADI
Enschede, The Netherlands, June 2022

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Urban Planning and Management

SUPERVISORS:
Dr. Jon Wang
Dr. Fran Meissner

THESIS ASSESSMENT BOARD:
Prof. Dr. Karin Pfeffer (Chair)

EXTERNAL EXAMINER:
Dr. Achilleas Psyllidis (Technische Universiteit Delft)

ABSTRACT

Traffic congestion is a problem that many cities face. To address this problem, policymakers frequently implement traffic interventions such as changing street directions or street access without building new roads. Such an approach may result in an unexpected distribution of traffic flows and the emergence of new congestion. This study attempts to understand this relationship between the street and traffic patterns. It is hypothesized that changes in traffic patterns, as well as the emergence of new congestion, are caused by traffic modifications which ultimately lead to changes in intricate street patterns. To test this hypothesis, a machine learning modelling approach is proposed in parallel to explore the potential of machine learning models in capturing the relationship between street morphology and traffic patterns. Using Barcelona and its Superblock policy as a case study, first, a traffic mapping comparison between before and after 'traffic modification' is conducted. Subsequently, change in street indices and the classification of street indices into street pattern types are explored. Finally, machine-learning traffic modelling is performed to compare linear regression, decision tree, and random forest algorithms. After estimating these without distinguishing by street type, the algorithm's evaluation is repeated by differentiating between street pattern types. Through the traffic congestion mapping, the study demonstrates the changes in traffic congestion following the traffic modification. Through the exploration and the classification of street indices, the study identified the change in the street pattern morphology that occurs post-traffic modification. Furthermore, by evaluating the performances of various machine learning-based models, the study identifies the best model for the study case and whether street patterns play important role in the change of congestion post the implementation of traffic modification.

Keywords: Traffic Congestion, Traffic Modification, Urban Street Pattern, Street Morphology, Machine Learning Modelling

ACKNOWLEDGEMENTS

First and foremost, this thesis research and my master's study would not have been possible without The Nuffic Neso, *Studeren in Nederland* (StuNed) Scholarship. I am grateful to The Embassy of the Kingdom of the Netherlands in Indonesia for selecting me as an awardee.

I owe my sincere gratitude to Dr. Jon Wang, Dr. Fran Meissner, and Prof. Karin Pfeffer, my supervisors, and my thesis assessor, for their valuable advice and assistance which greatly improve the quality of my master's thesis. Jon's expertise in spatial data science has undoubtedly improved my coding skills and expanded my knowledge of data exploration techniques. Fran's expertise in critical data studies has sharpened my intuition for organizing research flow and developing profound arguments. Professor Karin's suggestions helped me see the flaws in my research and provided me with a new perspective.

I would also like to thank Prof. Raul Zurita, Dr. Mahdi Khodadadzadeh, Dr. Rolf de By, and Srikumar Sastry for opening the door and guiding me into the world of computer science. I doubt I would have had the courage and skills to write a thesis related to machine learning and data science without their inspiring lectures and tutorials in the big geodata processing course.

I want to thank Meruyert Kenzhebay, Wiyanda Aflah, Gregorius Bryan, Keanu Herbowo, Dewi Kumalasari, Maulia Mahirani, Elva Aulia, Christian Sarungallo, Roberto Ramirez-Juarez, and Ikra Durmus for the late-night talks and the occasional BBQ party in the past two years. The challenging life of my master's study become bearable because of their companionship and support.

Special thanks to my parents, for which I will everlastingly be indebted. To my father, thank you for being my role model in life, and to my mother, thank you for the endless love and prayers. Also big thanks to my siblings Fathin, Nabila, and Dzaki for their subtle but perceived support in the past two years.

Finally, I would like to honour my late grandmother, Soefijah, by dedicating this thesis to her. She is the one who first teaches me intermediate math and algebra. Her teaching has contributed to my ability in coding and grasp programming scripts, which I then used extensively in this research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Justification and overview

Traffic congestion is a global issue that many cities face. This problem exists both in the developing and developed worlds. For example, in the Asia Pacific region, traffic congestion has intensified pollution levels and provoked an up to 35% increase in respiratory-related diseases (Climate & Clean Air Coalition, 2015). Meanwhile, in Western countries, traffic congestion is recognized as a daily burden on personal happiness that degrades life quality (Hays, Olds, & Spence, 2016). Responding to these situations, global cities have developed various traffic interventions to improve their traffic situations. These interventions can be divided into four categories: a) financial incentives/disincentives, b) road development, c) mode transition, and d) traffic modification. However, these interventions are not perfect and frequently cause additional issues.

a) In terms of financial disincentive interventions, London implemented a charging zone in February 2003 to reduce congestion in the city centre. The policy's concept was to charge motorists who enter the zone on a weekday afternoon (Badstuber, 2019). Although this policy reduced overall city centre congestion by 30% (Ambühl, Loder, Becker, Menendez, & Axhausen, 2018), it also reduced taxi and ride-hailing service coverage in that area (BBC News, 2018).

b) Regarding road development interventions, in 2003, the municipality of New Delhi decided to enact a roads development project that combines land-use policy and spatial distribution plan through the construction of arterial roads (Tiwari, 2003). A few years after the project was enacted, they found that the spatial distribution plan has not led to promising results. Instead of creating a spatial distribution, this project is alleged to increase people's dependence on motorized vehicles, which leads to an increase in car ownership (CSE India, 2017).

c) Another intervention that has been favoured lately is the transport mode transition. A transport mode transition can be defined as shifting the demand for private vehicles to other transport modes like cycling and public transport (SWARCO, 2018). In Copenhagen, the transportation transition to cycling manages to increase public satisfaction and is estimated to decrease the city's annual healthcare cost by up to 59 million euros (Gössling, 2013). However, it is also mentioned that their transportation transition is a long-term investment that caused societal resistance (Gössling, 2013).

d) Lastly, the fourth type of intervention is traffic modification. This type of intervention aims to reduce traffic congestion by changing street directions or street access without constructing new roads (Topirceanu et al., 2016). For example, at the end of August 2016, the Jakarta provincial government used this intervention by enacting an "odd-even" plate policy (scheduled access for private cars based on vehicle plate number) on some major arterial roads in Jakarta. Even though the policy successfully shifts traffic flows on most arterial roads, it also redistributes traffic flows into collector roads (The Jakarta Post, 2019). This increases collector road congestion in comparison to before the policy was implemented (Supriana, Siregar, Tangkudung, & Kusuma, 2020).

Based on the examples provided, one can conclude that each type of traffic intervention has limitations. However, given the concern that traffic modification may cause additional road congestion, this type of intervention requires particular scrutiny, and urban planners would benefit from being better able to estimate the impact of that kind of intervention. As a result, a study focusing on traffic modification and attempting to predict the congestion caused by it is required.

This paper takes up this challenge. In this first chapter, I will construct a hypothesis to comprehend why traffic modifications frequently cause congestion and describe why machine learning is a suitable approach to investigate it. Then, based on this hypothesis, I will form the research problems and objectives, as well as simplify the logical assumption into a conceptual framework. Following that, in Chapter 2, I will elaborate on the literature relevant to and capable of supporting this hypothesis. In Chapter 3, I will describe the Eixample district in Barcelona as the study area, followed by the reasoning for location selection, the dataset used in the study, and the research process. Subsequently, in chapter 4, I will present and explain the findings from each stage of the research process, as well as respond to the research objectives raised in the first chapter. In chapter 5, I will discuss and critically examine the analysis results from the previous chapter. Finally, in Chapter 6, I will summarize the research paper, conclude the key findings, and suggest future research directions.

1.2. Background and research problem

Modifying traffic often causes road congestion elsewhere. This is because traffic modification can result in incidental changes in street patterns, causing an unprecedented distribution of traffic flow and triggering the emergence of new congestion hotspots (Heidari Soureshjani, 2016). To give a general understanding, street patterns can be defined as the abstract form of street configurations in a city. The street configurations might form squares or odd shapes between them, depending on how their layout is designed. The combination of shapes formed by the streets is the street pattern (Louf & Barthelemy, 2014).

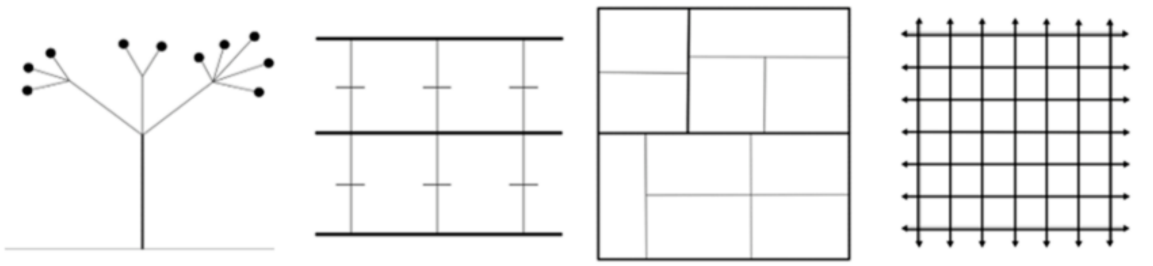


Figure 1. Structures of Four General Types of Street Patterns (Chan, Donner, & Lämmer, 2011b)

Four types of street patterns are commonly referred to in urban planning literature, namely a) the tree-like type, b) cul-de-sac type, c) T-type, and d) grid type (Figure 1). Each type of street pattern has its functions, objectives, and limitations: a) the tree-like street type tends to protect the privacy of the residents and is therefore mostly used in a residential area, b) the cul-de-sac type has good accessibility but has weaker interior connectivity than the others, c) the T-type tends to reduce domestic connectivity but has high efficiency for long-distance or freight transport, and d) the grid type has high connectivity but a low transportation efficiency due to its excessive number of intersections (Han, Sun, Yu, Song, & Ding, 2020). In short, these are the general type of street patterns that mainly exist in our cities, however, there are also many other types, or variations of the typical ones, such as the polycentric tree pattern (Deng, Liu, & Yun, 2019).

Corresponding to the initial discussion, these street patterns have a direct influence on traffic congestion. For example, it is found that street patterns with multiple junctions can improve traffic flow efficiency (Tsekeris & Geroliminis, 2013). Also, tortuous long streets without sufficient junctions are recorded to be constantly more congested than those without this characteristic (P. Zhao & Hu, 2019). Additionally, traffic congestion is positively associated with a compact grid pattern but negatively associated with a polycentric tree pattern (Y. Li, Xiong, & Wang, 2019). On top of that, a rigid grid pattern critically increases congestion density due to its hindrance on the junction's turning and U-turning (Wu, Hu, Jiang, & Hao, 2021). Realizing these correlations, street patterns must be considered as one of the factors that influence the outcome of traffic modification and thus should also be assessed to better evaluate the latter.

The discussion on the most practical way to assess traffic modification is ongoing. At the moment, there are two main approaches to assessing traffic modifications: a) the Coordination of Network Descriptors for Urban Intelligent Transportation Systems (CONDUITS) approach, and b) the stakeholder participatory policy evaluation approach. The CONDUITS approach excellently manages to quantify the effect of a traffic modification. However, it is not a suitable tool for spatial modelling contexts. This is because it is mainly designed to assess traffic policy performance based on certain key performance indicators such as the number of accidents, and pollution levels, which are all aggregated measurements with spatial information missing. Therefore, it cannot quite depict the localized spatial consequences that traffic modification has, such as new congestion hotspots (Kaparias et al., 2020).

Furthermore, in terms of assessing traffic modification based on community satisfaction, the participatory policy evaluation approach is noted to be the most comprehensive approach. Since it relies on community inputs and complaints, this approach answers the very base objective of the policy itself, which is improving the community travel experience. However, since it requires intensive social engagement and wide-scale surveys (consumer surveys and traffic counting), it comes with a high price tag and is undoubtedly time-consuming (Shi, Lee, Guo, & Zhu, 2020).

Although CONDUITS and participatory evaluation are appropriate methods for evaluating traffic policy performance, they are not the best way to understand the relationship between traffic congestion and street patterns. Alternatively, with the proliferation of traffic data, using data-driven models for modelling traffic flow, particularly those based on machine learning techniques, appears to be a viable option for assessing traffic modification (Peng et al., 2021). Not only can policy impact be measured using this approach, but spatial changes and the relationship between traffic factors can also be measured (Alqudah & Yaseen, 2020). Having said that, no one has explicitly studied a machine learning approach that is specifically designed to model the congestion outcome of traffic modification while simultaneously taking urban street patterns as an input factor.

1.3. General research objective

Using machine learning modelling to investigate the relationship between street pattern and traffic congestion that occurs after traffic modification implementation.

1.4. Specific research objectives

The general objective can be further elaborated into four specific objectives and several research questions as follows:

1. **Identifying the pattern of traffic congestion before and after the implementation of traffic modification.**
   a. Is there any significant shift of congestion hotspots post traffic modification implementation?
   b. How does the traffic modification affect the congestion pattern?

2. **Measure and classify urban street patterns in the study area based on quantitative street indices.**
   a. What are the suitable street indices for measuring and classifying urban street patterns?

b. How is the street pattern in the study area before and after the traffic modification?
c. How is the street pattern reflects the influence of the traffic modification?

3. **Applying and comparing machine learning model that can estimate the level of congestion, based on different street indices and street pattern inputs.**
    a. What are the appropriate street indices to use for modelling traffic congestion with machine learning?
    b. How accurately does the traffic congestion model perform?

4. **Decide on the most accurate machine learning model for predicting traffic congestion with urban street patterns as an input factor.**
    a. How does each machine learning model behave when provided with different street pattern data?
    b. Does the chosen model have consistent accuracy on all types of street patterns in the study area?

1.5. Conceptual framework

In summary, the study will include four key components: traffic congestion, traffic modification, urban street pattern, and machine learning modelling. This study is based on the hypothesis that traffic congestion is influenced by urban street patterns and that modifying traffic can alter urban street patterns. Thus, applying traffic modification can change the trajectory of traffic congestion. Furthermore, because traffic congestion is a dynamic phenomenon, a methodical approach capable of capturing its dynamic inputs must be used. In this case, machine learning modelling appears to be a viable option (Noorlander, 2021). In addition, deriving from the specific research objective, the component of 'traffic congestion' is represented by traffic pattern mapping, whilst the 'urban street pattern' is represented by the classification of street networks into street pattern types. To summarise, once these two components are generated, machine learning modelling will be used to investigate the relation between them.

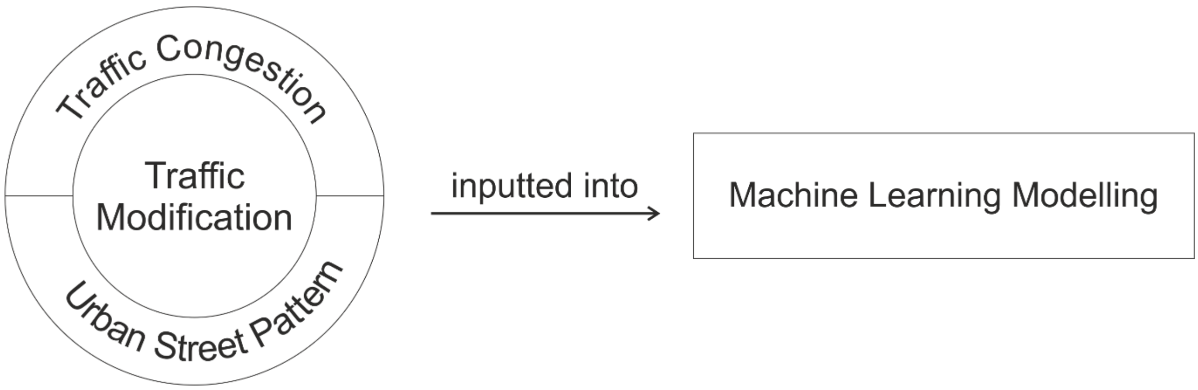


Figure 2. The Research Conceptual Framework

2. LITERATURE REVIEW

To assess the impact of traffic modification based on the proposed hypothesis, three major components must be reviewed: traffic congestion, street patterns, and machine learning-based traffic modelling. This chapter will go over those elements. The definition of traffic congestion, as well as how it is measured, will be explained. Furthermore, the street pattern will also be explained in terms of how it can be classified through street indices. Lastly, this chapter will also scrutinize the possibility of using machine learning modelling in traffic analysis and traffic prediction tasks.

2.1. Traffic congestion

Traffic congestion can be seen from various perspectives of traffic components. In terms of speed and time, it can be defined as an occurrence of longer vehicular queuing, slower speed, and longer trip time than usual (Ministry of Transport New Zealand, 2019). It can also be defined as an inability of a road to accommodate a safe and uninterrupted traffic flow due to environmental constraints (Knoop, 2017). In another study, it is defined as stagnation of traffic which results in connectivity disturbance within urban areas and intra-city interaction (Christodoulou & Christidis, 2020). All these definitions are valid in their own way, but to take into account the factor of 'change', in this study, traffic congestion is defined as the decrease in vehicular movement efficiency that is caused by a physical alteration in its surrounding (Laetz, 1990).

2.2. Measuring traffic movement

Measuring and quantifying traffic movement has been studied and applied through many kinds of instruments. At the moment, there are two main ways of generating traffic data, namely a) GPS user-generated and b) inductive-loop traffic sensors (Alqudah & Yaseen, 2020). On the one hand, user-generated traffic data has been frequently used for commercial purposes due to the abundance of smartphone GPS data (Lau, 2020). Despite the ease that comes with it, this approach has its context bias if it is used for governmental decision-making. In reality, the users are too randomly distributed and therefore do not qualify for systematic assessment (Ruktanonchai, Ruktanonchai, Floyd, & Tatem, 2018). On the other hand, the inductive-loop traffic sensor is more evenly distributed, has a better fit for traffic policy assessment (suitable for structured research as well as periodic monitoring), and has a lower potential for location bias (Aronsson & Bengtsson, 2019). However, even though it has a better data input accuracy, it is more expensive, and non-scalable for developing countries with numerous non-motorized vehicles and proliferated ride-sharing services (Pramanik et al., 2021).

Considering the advantages and disadvantages of each approach, in the study, inductive loop traffic data will be used as the main producer of congestion data. In brief, an inductive-loop works by detecting the presence of a conductive metal object that moves between the wire detector and induces an electrical current in the object. The induced current reduces the loop inductance, which is sensed by the inductive-loop electronics unit. The electronic unit (vehicle counter) interprets the decrease of induction as a moving object and determines the speed of the moving object (calculated by the distance of how far the wire detectors are installed from each other). Once the movement is appropriately estimated, the recorded traffic data will be sent through the transmitter into the controller (Soriguera & Robusté, 2011) (refers to Annex 1 in the Appendix for further technical detail).

2.3. Defining street configuration and street indices

Street configuration can be defined as the orientation of the streets that determine a city's spatial logic and order (Boeing, 2019). In the sense of urban shapes and forms, it can also be described as the structure and geometric ordering that arise through street network orientation (Boeing, 2020). However, at its core, a street configuration is an arrangement of interconnecting streets and junctions that portray a system of roads for a given area (ESRI, 2018).

Street configurations are frequently quantified to assist policymakers in better understanding their cities (Peponis, Allen, Haynie, Scoppa, & Zhang, 2007). These quantitative measurements are commonly referred to as 'street indices' and can range from simple indices like street length, the number of road segments, the number of junctions, the ratio of junction/dead-ends, to more sophisticated indices depending on the condition and phenomenon that it attempts to capture (Wang & Debbage, 2021a; Cubukcu, 2015). For example, an area with a high value of the cyclomatic index has inefficient street network looping. As a result, this index is frequently used to identify areas with poor street orientation for disaster evacuation (Sharifi, 2019). In another case, when it comes to walkability studies, the street section density index is frequently used to determine the city's most walkable neighbourhood. This is because it is widely accepted that the denser the street networks, the more walkable the area (Lamíquiz & López-Domínguez, 2015).

2.4. Using street indices to classify street patterns

Since this study is focused on the urban street pattern as a street configuration, a collection of street indices that can measure and quantify street networks into groups of street pattern types is needed. A study by Han et al., 2020, developed the indices that can be used for the matter. These indices are called the cross-road ratio/X-type (measuring the ratio of the number of crossroads against Y-junctions), penetrating street ratio (measuring the ratio of the number of street sections that have no dead-ends), and the cul-de-sac ratio (measuring the ratio of the number of street sections that has dead-ends). According to the study, a street network can be classified as cul-de-sac type, tree type, T-type, or grid type, if the values of the street indices belong to a certain range on the pattern type spectrum. As a result, these indices will be an important indication for detecting any changes in street patterns that may result from traffic modifications.

2.5. Relationship between street indices and traffic congestion

With the hypothesis that "changes in urban street pattern may result in changes in congestion pattern," it is reasonable to assume that changes in street indices (which are the formulator of street patterns) may also affect congestion trend. Several studies have been commonly referred to on this matter. For example, The street flow index, which is used to indicate the presence of long roads, has a negative correlation with traffic congestion (Song & Knaap, 2007), the junction adjacency index, which is used to measure street junction closeness, also has a negative correlation to congestion (Wang & Debbage, 2021b), and a street connectivity index, which calculates the ratio of street sections to junctions, has also shown a negative correlation too (Lowry & Lowry, 2014). Furthermore, The length of a street section (Cubukcu, 2015), the number of junctions (Choi & Ewing, 2021), and the length of a one-way road (J. Zhang, Zhang, Yang, & Zhou, 2020), are among the simpler street indices that have been found to negatively correlate with congestion, which is in contrast with the blocked road index that is positively related to traffic congestion (Gavrilyuk, Vorob'Yova, & Shalagina, 2020). In conclusion, all of these studies show that there is a link between street indices and traffic congestion. Thus, in the further sections, these street indices are calculated and will be selected as independent variables for the machine learning traffic model.

2.6. Machine learning for traffic analysis and modelling

Machine learning (ML) is a branch of computer science derived from pattern recognition and computational learning theory. It focuses on the development of an algorithm capable of making predictions and learning from data (Alpaydın, 2010). Realizing its potential, machine learning is widely used to predict and unpack the hidden relationship between complicated data features (Helm et al., 2020). One of the capabilities of machine learning is to draw a relationship between street elements and formulate a prediction model of traffic patterns from their spatial tendency (Noorlander, 2021).

Many studies have attempted to address this possibility from various angles, such as identifying appropriate generic explanatory variables for a traffic model, comparing machine learning algorithms for real-time traffic analysis, and comparing how different machine learning algorithms performed on a traffic modelling task. For example, when looking for generic explanatory variables in an ML traffic model, it was discovered that the frequency of road accidents correlates with a significant increase in traffic jam occurrence (Sangare, Gupta, Bouzefrane, Banerjee, & Muhlethaler, 2021). Meanwhile, in another model, low traffic activity is associated with a compacted land use distribution (Abdulkareem, Adulaimi, Pradhan, Chakraborty, & Alamri, 2021). Furthermore, when it comes to real-time traffic data analysis, it was discovered that the neural network algorithm came out as the best option for machine learning algorithms (Lee & Min, 2018). However, because of its complexity in interpretation, the neural network is deemed unsuitable for investigating the relationship between variables (Bathaee, 2018).

Having said that, when comparing different algorithms for predictive traffic modelling, random forest, decision tree, and linear model outperformed other algorithms such as naïve bayes, support vector machine, and ada-boost (Malik, El Sayed, Khan, & Khan, 2021). Complementing that finding, a study conducted by Bratsas et al., 2020 found that the tree-based method and linear regression algorithm are more efficient than other methods for predicting traffic mean speed (TMS). Eventually, since the goal of this study is to build a predictive traffic model for the purpose of investigation, well-performing and interpretable algorithms such as decision tree, random forest and linear regression were chosen for the task.

3. MATERIALS AND METHODS

This chapter describes study materials such as 'the study area' and 'dataset'. In addition, it also explained 'the methodology overview' to depict the general workflow of the study. Lastly, 'technical steps' derived from the methodology overview are thoroughly explained to elaborate on what was done in the study.

3.1. Study area

The chosen study area is Barcelona. It was chosen because it meets the criteria of a) having an ongoing traffic modification, b) have a distinctive urban street pattern, and c) have adequate traffic sensor infrastructure with publicly available data. As for further detail, its ongoing traffic modification policy is called the 'superblock'.

Barcelona's superblock policy aims to minimize the presence of cars by modifying traffic flows. As can be seen in Figure 3, it is designed to limit motorized vehicle movement in some dedicated areas, meaning that inside the dedicated areas traffic will be eliminated or traffic calming interventions will be enacted. Those interventions can vary from road conversion into urban parks, creating pedestrian-friendly zones, to widening pedestrian lanes. Through this approach, the city aims to reclaim urban spaces from traffic and become more livable (Ajuntament de Barcelona, 2014).

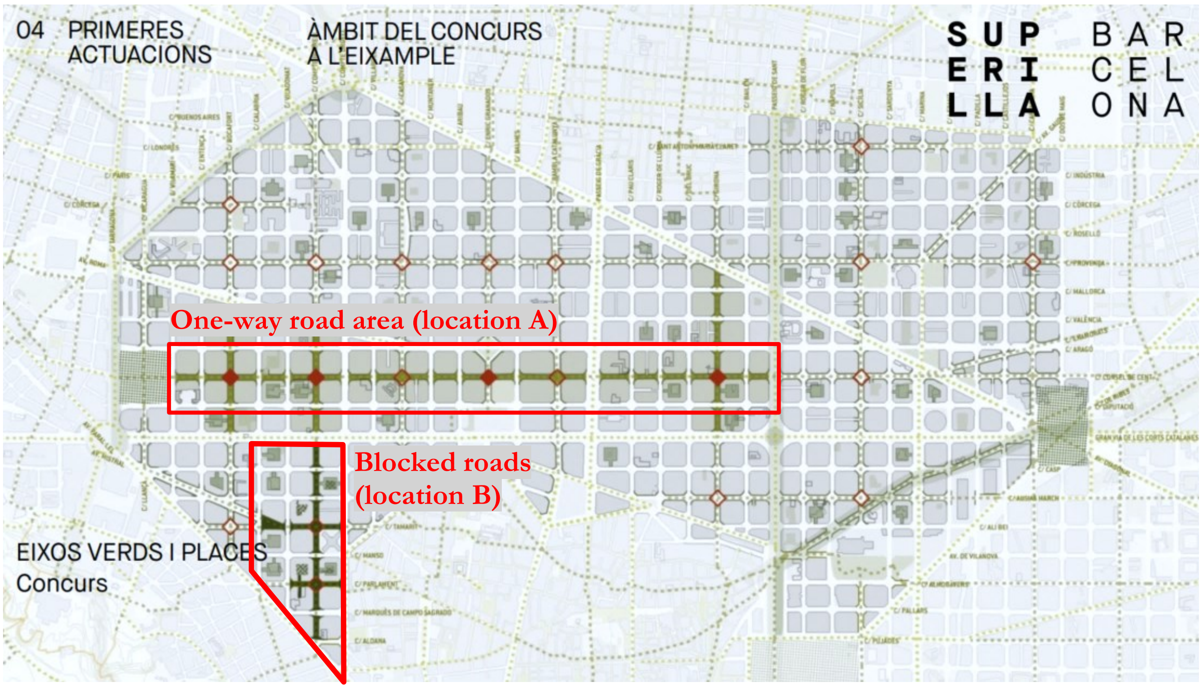


Figure 3. Illustration of Superblock Modifications in Eixample District (Ajuntament de Barcelona, 2014)

Although this policy will create more space for people and generally increase the environmental quality, it can also become a recipe for future traffic issues and make the city prone to the emergence of new congestion hotspots. This potentiality needs to be investigated further (Duchêne, 2019). Furthermore, at the moment, the superblock is mainly implemented in The District of Eixample. This is because, unlike other districts, The Eixample District has wide roads with high traffic volume, a suitable condition for policy trial and monitoring (Ajuntament de Barcelona, 2020). Correspondingly, this study will not cover the entire city of Barcelona but will concentrate on The District of Eixample as its study area. At the moment, two modification type are implemented in the district. For ease of identification, the one-way road implementation zone will be referred to as "location A," and the blocked road implementation zone will be referred to as "location B."

3.2. Datasets

The study will draw on two main datasets: the Barcelona traffic dataset taken directly from the municipality's open data catalogue and a street indices dataset calculated by the author of this study. The traffic data serves as the model's response variable, while the street indices serve as the collection of potential explanatory variables.

Dataset 1: Barcelona's traffic dataset

To generate a decent machine learning model, sufficient and reliable data is required to be fed into the algorithm (Alwosheel, van Cranenburgh, & Chorus, 2018). In this case, Barcelona has an adequate amount of traffic data that is generated continuously every 30 minutes. The data is produced by inductive-loop sensors installed under the asphalt, which are located all over the city (for the sensor locations map, please refer to Annex 2 in Appendix). The data is organized so that each row represents a traffic sensor unit, and the columns represent the attributes listed in Table 1.

Table 1. Attribute Detail of Barcelona's Traffic Dataset (Ajuntament de Barcelona, 2021)

Attribute	Description	Rationale
Long	Longitude coordinates of the traffic sensor	Geospatial location and potential explanatory indices for spatial model (Banerjee, 2005)
Lat	Latitude coordinates of the traffic sensor	
Street_Name	Name of the street where the sensor belongs	Necessary labels to determine sensor locations
Id_Tram	Traffic sensor unit identifier	Unique identifier for data management
Time	The specific time where the traffic movement was recorded (YYYYMMDD format followed by time in HHMMSS format, e.g. 20180516103551)	As an indicator to distinguish between traffic movements recorded before and after 'the superblock'
Traffic	Traffic status (0 = no data, 1 = very fluid, 2 = fluid, 3 = dense, 4 = very dense, 5 = congestion)	The main response variable that the model will attempt to predict

Dataset 2: Street indices dataset

Some street indices have a significant correlation with traffic congestion, as discussed in Section 2.5. In this study, those street indices are used as potential explanatory variables for traffic congestion. As complementary information, the calculation of street indices is commonly conducted through item counting in the designated areas (Lowry & Lowry, 2014). The number of junctions, for example, is calculated by counting and normalizing the number of junctions in those areas (Cubukcu, 2015) (Figure 4). These areas can be administration borders, standardized rectangular borders (spatial fishnets), to buffer zones depending on the study interest (Patorniti, Stevens, & Salmon, 2020). In this manner, since the interest of the study is traffic congestion, the indices will be calculated with traffic propagation zones as the designated areas.

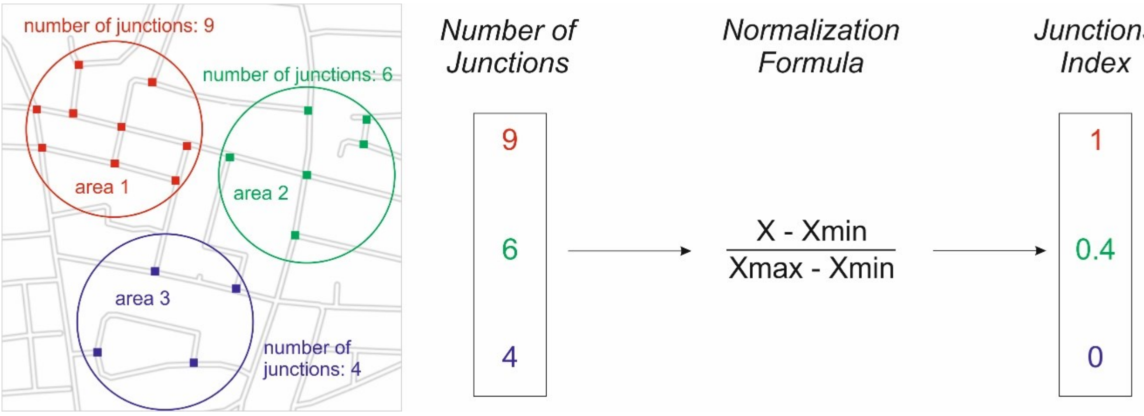


Figure 4. Illustration on Calculating Street Indices using Max-Min Normalization

A traffic propagation zone is a pre-determined buffer zone where vehicular queuing often reaches out to (Nagy & Simon, 2021). Calculating street indices with the traffic propagation zone as the designated areas is appropriate for the study because each traffic sensor's recorded congestion is influenced by the street configuration in its vicinity (Xiong, Vahedian, Zhou, Li, & Luo, 2018). Depending on the location, and trend, the propagation zone can range from 300 meters to 3 kilometres. (Jiang, Kang, Li, Guo, & Havlin, 2017). For example, in the United States, the minimum radius for a traffic propagation mitigation zone is 600 meters, in Germany and the United Kingdom, the standard is minimized to a 480 and 450 meters radius respectively. In France, the radius minimum is stretched to 1 kilometre (Wolhuter, 2015), and in Beijing, the radius is stretched up to 3 kilometres (Jiang et al., 2017). In this research, considering the average length of congestion and the average size of a block in Barcelona, the traffic propagation buffer is set to 600 meters radius from each traffic sensor location, and with that, the street indices calculation are based on these areas. Once each street index is calculated, the street indices dataset is built in a structure where the row represents each traffic sensor unit and the columns represent different street indices which can be seen in Table 2 (the calculation for all street indices are computed using "summarized within" tool in ArcGIS software).

Table 2. Attribute Detail of Calculated Street Indices Dataset

Attribute (Street Indices)	Description	Involved components	Rationale for including	Reference
Length	The total length of roads	Road length	The longer the road the less congested it will be	Cubukcu, 2015
Junctions	The total number of junctions	Number of junctions	More junctions indicate more options for mobility and hinder the possibility of congestion	
Blocked	The total length of blocked roads	Blocked roads	Road blockage can increase the probability of congestion	Gavrilyuk, Vorob'Yova, & Shalagina, 2020
Halven	The total length of one-way roads	One-way roads	One-way roads increase the efficiency of traffic flow and decrease the occurrence of congestion	J. Zhang, Zhang, Yang, & Zhou, 2020

NR (Node Ratio)	The ratio of the number of nodes (vertices with connections) to the total number of vertices	Existence of street nodes	Areas with a lot of street nodes tend to have less congestion	Tsekeris & Geroliminis, 2013
CR (Cul de Sac Ratio)	The ratio of the number of vertices without connections (cul de sac) to the total number of vertices	Existence of cul-de-sac	Areas with a lot of cul de sacs tend to have slower traffic and are prone to congestion	Cozens, Hillier, Cozens Ã, & Hillier Ãã, 2008
TR (T-junction Ratio)	The ratio of the number of T-junctions to the total number of junctions	Existence of T/Y-junctions	Areas with a lot of T-junctions tend to be more congested than areas that have a lot of cross-roads	G. Zhao et al., 2020
XR (Cross Road Ratio)	The ratio of the number of cross junctions to the total number of junctions	Existence of cross-road	Areas with a lot of cross-road tend to be less congested than areas that have a lot of T-junctions	Y. Li, Xiong, & Wang, 2019
Flow (Street Flow)	The length of street divided by the number of junctions	- Road length - Number of junctions	Long roads without junctions are usually less congested	Song & Knaap, 2007
Adjacent (Junction Adjacency)	The average distance between junctions	The proximity between road junctions	Adjacent junction indicates numerous options of turning which decrease the probability of traffic jam	Wang & Debbage, 2021
Connect (Connectivity)	The number of street sections divided by the number of junctions	- Number of street sections - Number of junctions	A balanced ratio between street sections and junctions represents good street connectivity which hinders the probability of congestion	Lowry & Lowry, 2014

***All indices are normalized with the max-min scaling equation (Illustration in Figure 4)*

3.3. Methodology overview

Once the traffic data and street indices data are produced, the study proceeds in four steps, (1) traffic congestion mapping; (2) data exploration and street pattern classification; (3) applying and comparing machine learning model; and (4) evaluating machine learning model based on the street pattern types (Figure 5). All of these steps are executed to answer the specific research objectives of the study.

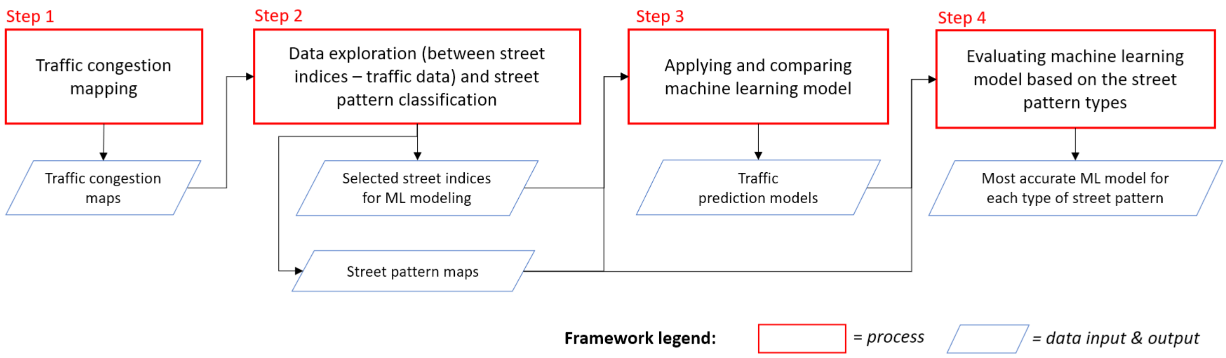


Figure 5. Methodological Framework of Study

The first step in the analysis, traffic congestion mapping, resulted in a comparison of traffic congestion maps before and after the superblock. This step is required to meet the research objective of "*identifying the pattern of traffic congestion before and after the implementation of traffic modification*". In the second step, data exploration and street pattern classification produced the selected street indices variables and the street pattern maps. These outputs are required to meet the research objective of "*measuring and classifying urban street patterns in the study area using quantitative street indices*". Following that, the selected street indices and the street pattern types are used in the third step of analysis, which is applying and comparing the machine learning model for predicting traffic congestion. From this step, the traffic prediction model is generated and the research objective of "*applying and comparing machine learning model that can estimate the level of congestion based on different street indices and street pattern inputs*" will be completed. Finally, in the last step, the performance of different machine learning models is evaluated in relation to the street pattern type. This last step provides insight into how different models behave when they are fed with data from different types of street patterns, and it will answer the last research objective of "*decide on the most accurate machine learning model for predicting traffic congestion with urban street patterns as an input factor*". Each step of analysis will be explained and elaborated on in the following sub-sections.

3.4. Technical steps

Step 1: Traffic congestion mapping

Since the available traffic data is available at a small temporal scale (every 30 minutes), a data aggregation process is required to rationalize the traffic data. This aggregation is needed to ensure that the congestion data is on the same temporal scale as the street indices so that an appropriate data exploration can be performed.

Initially, a period cut-off is defined to separate the before and after 'superblock' periods. Referring to the inauguration date, the superblock is implemented in The Eixample District on February 19th, 2019. (Ajuntament de Barcelona, 2020). As a result, the congestion data is divided based on that date and labelled as '**t1**' for all recorded traffic before February 19th, 2019, and '**t2**' for all recorded traffic between February 19th, 2019, and February 24th, 2020. (2 days before the first recorded COVID spreads). From this process, separated datasets of before and after the 'superblock' are generated, with no potential traffic anomaly caused by COVID. Following the creation of these datasets, the traffic attribute in each of them is grouped by the 'Id_Tram' and aggregated using the 'mean' method. Once the splitting and the aggregation are done, two simple datasets are produced: the first contains the mean value of recorded traffic for each sensor in t1, and the mean value of recorded traffic for each sensor in t2. These datasets are then mapped by coordinates, and the congestion status is compared using a heat map.

A spatial heat map is an appropriate type of visualization to demonstrate a change in spatial phenomenon from a different time period (Słomska-Przech, Panecki, & Pokojski, 2021). Furthermore, given how traffic sensors can be distributed unevenly throughout a city, a heat map is also less biased than other point-based mappings such as kernel density and thus more widely used for traffic congestion cases. This is because, unlike a kernel density map, a heat map does not calculate the density of features and focuses on the recorded value of the feature (Zambrano-Martinez et al., 2019). Finally, two pieces of information are extracted from this step: the detected shifts in congestion hotspots and the change in traffic pattern caused by the superblock.

Step 2: Data exploration and street pattern classification

To formulate an accurate and conformable machine learning model, a proper selection of explanatory variables for the response variable is essential (Jenkins-Smith et al., 2017). For that matter, this study will employ four kinds of analyses: the correlation test, scatter plotting, change detection mapping, and street pattern classification. The correlation test is completed to observe the colinearity between variables, the scatter plot is run to numerically see the trend of change in variables between two periods of time (t1 and t2), the change detection mapping is run to identify the location of changes between two periods of time, and finally, the street pattern classification is run to classify the street networks into a different type of street pattern.

A correlation test is essential to determine whether a street index is eligible to be considered a driving factor of traffic congestion. With that said, in this test, all street indices will be considered as a potential explanatory variable and each of them will be tested against 'congestion level' as the response variable. In terms of the calculation method, numerous correlation tests exist in the quantitative research field. However, considering the commonality and easiness of interpretation, Pearson r correlation will be used in this study. The Pearson *r*-correlation test works by statistically measuring the strength of a linear relationship between paired variables (Sedgwick, 2012). This measurement is denoted by *r,* where *r* is ranging between -1 and 1. Positive values of *r* denote positive linear correlation, while a negative value denotes negative linear correlation. When it comes to the degree of strength, the closer the value is to 1 or -1 the stronger the linear correlation is (Werner, Todorov Valev, & Lyubenov Danov, 2009). Furthermore, the *r*-value can also be qualitatively described into five ranges, where 0.00-0.19 is "very weak", 0.20-0.39 is "weak", 0.40-0.59 is "moderate", 0.60-0.79 is "strong", and 0.80-1.0 is "very strong". Therefore, as an example, a correlation value of $r = -0.42$ would be described as a "moderate negative correlation" (Evans, 1996).

Once the correlation result is produced, the study proceeds into the scatter plot analysis. The scatter plot used in this study is focusing on comparing data points from two different time series. This kind of scatter plot is also known as the "before-after scatter plot" (Friendly & Denis, 2005), it works by plotting the first data entries (t1) on the x-axis, while simultaneously plotting the second data entries (t2) on the y-axis. In such a way, an observation regarding whether a variable has been increasing, decreasing, or stagnant can be detected (Cox, 2009). Collectively, by doing this approach to all street indices variables, a preliminary selection of street indices that has a similar tendency to congestion data can be done. The tendency of changes is determined by the aggregated spread location in the scatter plot. A data spread located on the upper side of a diagonal line (identity line) refers to an increment, a spread on the lower side of the diagonal line refers to decrement, while a spread that accumulates along the diagonal line indicates a little to no change in the variable (Keim, Hao, Dayal, Janetzko, & Bak, 2010).

Change detection is applied after the scatter plots are generated. Change detection is a technique in GIS for determining how a specific area has changed over two or more time periods (Riccardo, 2012). It entails comparing changes in geospatial features (raster or vector) captured over different time periods covering the same geographic area (Xu, He, Zhang, & Guo, 2009). In short, it is a useful method for determining the precise location of where a feature is decreasing or increasing (Gaber, Ahmed, & Ali Farrag, 2021). In this study, change detection is specifically used to map the magnitude of change in street indices and traffic congestion. It generates the change value by subtracting t2 features from t1 features, which is also known as the subtract mapping (Wolz et al., 2014). As a result, the process produces a new feature in which a positive value is interpreted as an incremental change, a negative value as a decremental change, and a near-zero value as barely any change (ESRI, 2020). Subsequently, this new feature is mapped with a congruent colourization to facilitate visual interpretation (Lynch, Crosbie, Fagan, & Naveen, 2010).

To precisely classify the street networks into street pattern types, a combination of two clustering approaches, unsupervised and supervised, can be used. Unsupervised clustering is required to provide an overview of the optimal number of clusters that should be found in the dataset (Lynch et al., 2010), whereas supervised clustering is required to dictate the classification of the dataset into various types of street patterns. Implementing these two methods yields a middle ground in which the study area's streets are classified into an optimal number of groups of street pattern types. In addition, the street pattern type that is produced from this process will also be used in the ML modelling to evaluate different algorithms based on the type of street pattern (refers to the methodological framework in Figure 5).

For the unsupervised clustering, this study employs two techniques: fuzzy c-means clustering and dendrogram. Fuzzy c-means clustering is a clustering method that allows multidimensional data points to be assigned to one of two or more clusters based on their likelihood or probability score (Aman Gupta, 2021). In general, it works by assigning membership to data points associated with each cluster centre (c) based on their distance (μ) from the cluster centre (Yu Feng, 2021). The data is more likely to belong to the cluster centre if it is closer to it (Figure 6). The fuzzy clustering process is iterative; several options for cluster numbers must be entered and the algorithm will calibrate and visualize cluster spreads, for then the user can decide the optimal grouping (Nayak, Naik, & Behera, 2015). The optimal grouping or clustering is determined by Dunn's fuzzy partition coefficient (FPC) (Hung & Yang, 2001). The FPC reflects the 'goodness-of-fit' of clusters; it determines how close the cluster is to the best-fit grouping. It ranges from 0 to 1, with 0 indicating a completely distorted cluster and 1 indicating an optimal cluster (Liu, Zhang, Chen, & Chao, 2019). The FPC is a critical indicator for ensuring that the street pattern variations produced by supervised clustering fall into one of the groups with high FPC.

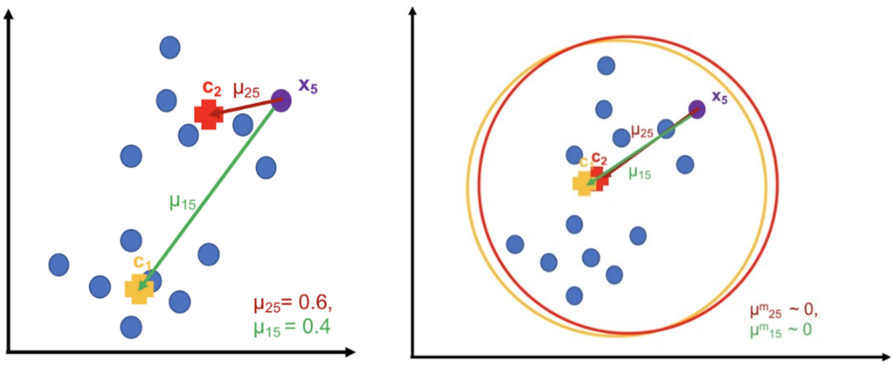


Figure 6. Visual Explanation of C-Means Clustering using Distance as Determinant (Yu Feng, 2021)

The dendrogram is another unsupervised clustering technique used in this study. It is primarily a tree diagram that represents relationships between data points with similar qualities (Prasad Pai, 2021). It is also known as hierarchical clustering because it depicts the hierarchical relationship between data (Nielsen, 2016). The primary application of a dendrogram is to determine the best way to allocate data clusters by tracking branches and observing their lengths and splits. This tracking is used to determine the optimal number of clusters in the dataset (Y. Zhao & Karypis, 2005). In more detail, the interpretation of the number of clusters is primarily determined by the length of its vertical branches. While the horizontal branches represent the data points and clusters, the vertical branches represent the distance or dissimilarity between clusters (Mesa & Restrepo, 2008). Having said that, the number of clusters within a dataset can be deduced through observation of the first occurrence of a large gap in vertical branches (Surhone, Timpledon, & Marseken, 2010) (Figure 7). However, it is worth noting that the hierarchical grouping of dendrogram works best when performed with an initial understanding of the dataset. This is because a large gap in hierarchical dissimilarity does not always represent empirical dissimilarity (Forina, Armanino, & Raggio, 2002). Like c-means fuzzy clustering, the result of dendrogram analysis is also used as a benchmark of optimal number of variation that can occur in the street pattern types.

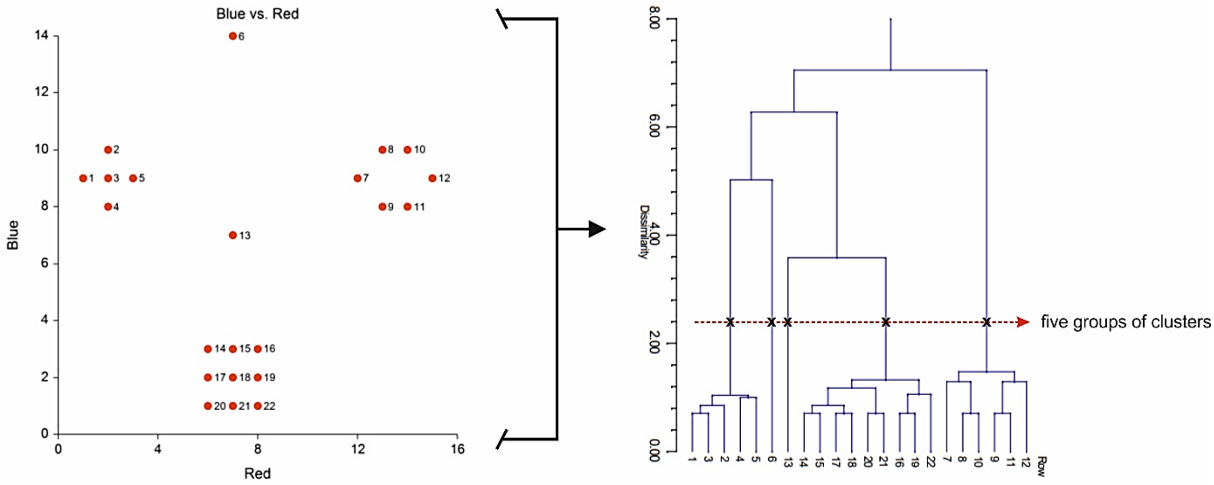


Figure 7. Example of Using Dendrogram to Clusters Several Data Points (Surhone et al., 2010)

After determining the optimal number of clusters that the dataset can produce, the study moves on to supervised clustering to classify street pattern types. In supervised clustering, the objective is to cluster the data into the desired existing classes. In this case, the task is to automatically cluster study areas' street networks into different urban street pattern types. Since the agreed-upon basic form of street patterns are consists of grid type, T type, tree type, and cul-de-sac (derived from pure-grid and pure-tree dichotomy) (Louf & Barthelemy, 2014; Chan, Donner, & Lämmer, 2011), the quantitative classification of street patterns by Han et al., 2020's (which is specifically designed to classify street into the above-mentioned types) is used in this study. In brief, the clustering process is carried out through numerical cut-off, such that several data points are classified into a particular pattern type if their network indices match one of the characteristic ratios that are enlisted in Table 3. These quantitative rules are ordered in a way that code "1" represents pure grid and code "5" represent pure cul de sac. This way, the higher the code, the higher the congestion probability is (Han et al., 2020). Furthermore, these characteristics are then translated into a base classifier, where data points are not only grouped but also visualized into the map of street pattern type in t1, and the map of street pattern type in t2. Although this approach (also known as supervised feature selection), is an excellent way to classify unlabeled data, it is important to note that it requires preliminary feature standardization, which is often unavailable for most types of datasets (Martínez Sotoca & Pla, 2010).

Table 3. Street Pattern Type and Their Characteristic Ratio (Han et al., 2020)

Pattern Type	**Code**	**Characteristic Ratio in Street Indices**
Grid-like	1	XR >= 0.6 ∩ TR <= 0.4 ∩ CR <= 0.1 ∩ NR >= 0.9
T-type	2	XR <= 0.6 ∩ TR >= 0.4 ∩ CR <= 0.25 ∩ NR >= 0.75
Transition	3	Does not belong to any end of characteristics ratio
Tree-like	4	XR <= 0.25 ∩ TR >= 0.75 ∩ CR = 0.4 – 1 ∩ NR <= 0.6
Cul-de-Sac	5	XR <= 0.1 ∩ TR >= 0.9 ∩ CR = 0.2 – 0.7 ∩ NR <= 0.8

Step 3: Applying the machine learning model

After selecting the appropriate street indices and generating street pattern types, the next step is to apply and compare the ML algorithm, so that the best traffic model can be selected. In this step, the street indices and the street pattern types are treated as driving factors of traffic congestion, serving as the model's explanatory variables. The congestion level, on the other hand, will be used as the response variable in the model. Furthermore, as discussed in Section 2.6 (machine learning for traffic analysis and modelling), the model will be run in linear regression, decision tree, and random forest algorithm, using the general rule of thumb of 70% and 30% train-test data split. The reasoning behind these algorithm selections is that when comparing different algorithms for predictive traffic modelling, random forest, decision tree, and linear model outperformed other algorithms such as naïve bayes, support vector machine, and ada-boost (Malik et al., 2021). It is also found that linear models have a constant lower rate of RMSE (root mean square deviation) for traffic mean speed modelling, compared to convoluted algorithms such as neural network and support vector regression (Bratsas et al., 2020). In addition, since this study aims to understand the relationship between variables, a convoluted algorithm such as the neural network can be difficult to interpret compared to more simple algorithms such as linear regression and random forest (Bathaee, 2018).

To give more detail, linear regression is a parameterized method and a supervised algorithm that uses a linear approach for a prediction task. The prediction function is handled as a linear combination of the data variables. It is designed to fit a line that explains the correlation between the response variable and the explanatory variables (Seber & Lee, 2003).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{i7} + \varepsilon$$

Where, for $i = n$ observations, $xi1\text{-}xi7$ = explanatory variables (selected street indices and street pattern types), Yi = response variable (traffic congestion level), $\beta 0$ = constant term, $\beta 1\text{-}7$ = constant term for each explanatory variable, and ε = the model's error term.

Furthermore, the decision tree is an algorithm that belongs to the family of tree-based supervised learning algorithms. It works by creating a training model that could be used to predict the value of the response variable by learning decision rules on the explanatory variables deduced from training data (Chauhan, 2022). In decision trees, the prediction of a value starts from the root of the tree (root node). It will then classify the input data by sorting them down the tree from the root to some terminal node providing the classification of the inputted data (Figure 8) (for a further explanation of the decision tree please refer to Annex 3 in Appendix).

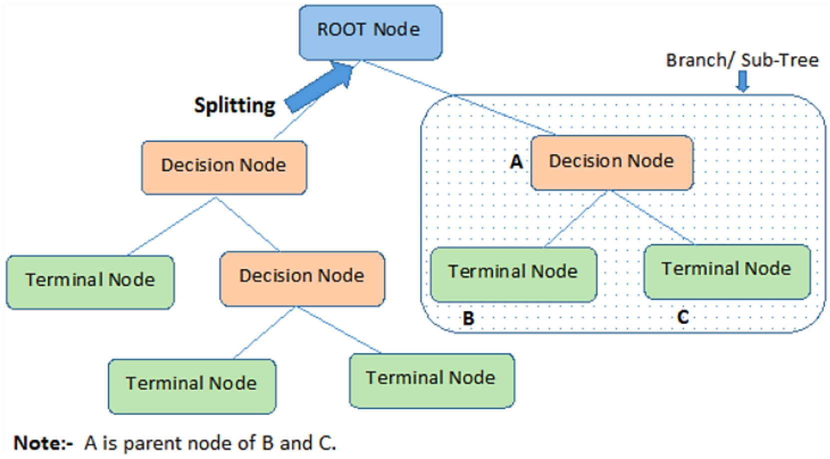


Figure 8. Schema of Decision Tree Workflow (Chauhan, 2022)

In addition, the random forest algorithm is principally an ensemble learning algorithm that is designed to handle the same data input differently (Cutler, Cutler, & Stevens, 2012). A random forest algorithm is several decision trees that use the same set of explanatory variables, yet each of them uses the variables in a different order as their root node and decision nodes (Breiman, 2001). As stated, the algorithm of random forest work as an ensemble. Meaning, that once each tress has its prediction, the value or the class with the most votes (the result of the majority) will be chosen as the model's prediction (Yiu, 2019) (Figure 9).

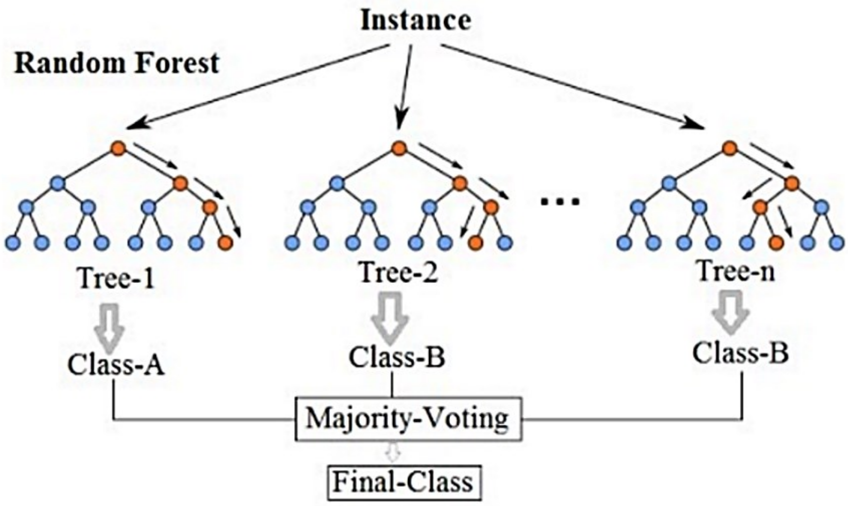


Figure 9. Schema of Random Forest Workflow (Tran, 2019)

Once the training and testing for each algorithm are done, three measures are calculated to evaluate their performance, namely R-square, normalized-RMSE (NRSME), and the validation data scatter graph. The R-square accounts for how much variation of a response variable is explained by the explanatory variable(s), the NRMSE measures the normalized distance between the predicted values and the actual observed values in a euclidean graph, while the scatter graph provides visual confirmation of how the predicted values performed compared to the actual value (Lundberg, Johnson, & Stewart, 2021). Once the best machine learning algorithm is chosen, the model proceeds into feature importance analysis. The feature importance analysis is performed to rank the independent variables based on their relative importance to the model (X. Li, Wang, Basu, Kumbier, & Yu, 2019). Overall, this step produced two outputs for the study: the predictive machine learning model of traffic congestion and an understanding of each variable's importance for the model.

Step 4: Evaluating the machine learning model

The final step in this study is to divide the dataset based on the street pattern type that was generated from the clustering and classification process. Once it is divided, each dataset is fed into several machine learning algorithms and evaluated by various measures just like in the previous step. The main difference is that feature importance analysis is not performed in this step. This is due to the fact that the goal of this step is not to build a predictive machine learning model, but rather to evaluate the algorithms and determine which algorithm is best suited for different types of street patterns. Eventually, through this step, the most suitable machine learning algorithm for each street pattern is determined.

4. RESULTS

Each section that is narrated in this chapter is in relation to five outputs that were depicted in the methodological framework (Figure 5). First, it will describe the traffic congestion maps. Next, it will report on the selected street indices for ML modelling, followed by the generated street pattern maps. After that, it will describe the machine learning traffic prediction model, and lastly, it will narrate the evaluation of different machine learning models based on the street pattern type.

4.1. Traffic congestion maps: before and after superblock

Traffic mapping in this study is performed through an aggregation process that involves averaging the traffic movement data into two attributes: 'traffic status before traffic modifications (t1)" and "traffic status after traffic modifications (t2)". Due to this process, the discrete traffic values as described in Table 1 (1= very fluid, 2= fluid, 3= dense, 4= very dense, and 5= congestion) are transformed into a continuous decimal value that is visualized in a gradation of colour as can be seen in Figure 10. In addition, for the purpose of this study, a traffic hotspot is defined as a traffic spot with a value greater than 2.8. (identified on the map with orange-yellow spots). This is based on the data distribution, which shows a large gap between values >2.8 and other values. By comparing the t1 and t2 maps, it is possible to conclude that there is indeed a change in congestion patterns that occurs post the implementation of traffic modification in Barcelona. For example, in t1, there are three traffic congestion hotspots detected, whereas, in t2, there are five congestion hotspots detected. Not only there are two additional hotspots, but the former hotspots are also becoming more intense. Furthermore, it can be seen that the traffic spots in the centre of the district have become more visible and dense. By considering the place of traffic modification (location A= one-way intervention, location B= blocked roads intervention), it is clear that the majority of increases in traffic congestion are located nearby these areas. For instance, one of the new congestion hotspots is located within 250 meters of location B's southwest border, and the new accumulation of traffic spots is detected on location A's southeast border. To say the least, the existing hotspot in the top northwest and east corner of location A is also becoming more congested (turning yellowish). All these findings suggest that the overall recorded traffic in the study area has increased following the enaction of the superblock.

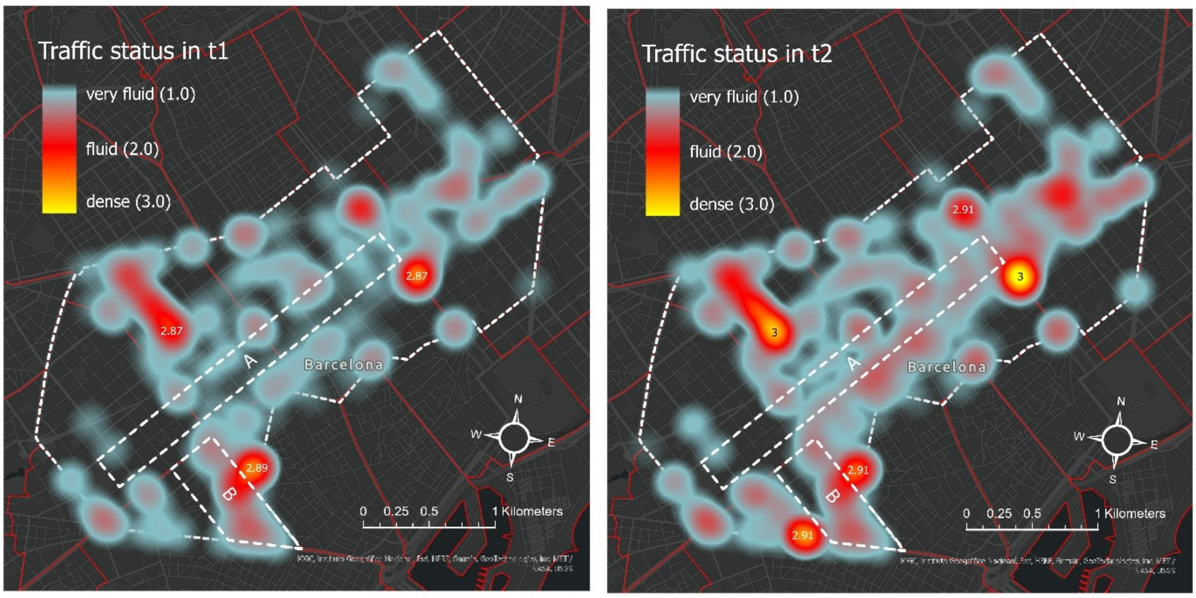


Figure 10. Comparison of The Traffic Congestion, Before (t1) and After (t2) Superblock

4.2. The selected street indices for the machine learning modelling

The selected street indices for traffic modelling are derived from three analyses: the correlation test, the before-after scatter plot, and the change detection mapping. Firstly, the Pearson correlation test (Figure 11) shows that there is some strong collinearity that exists between street indexes. In this case, the most prominent collinearities occur between the street form-related indices (TR, XR, length, junctions) and the street performance-related indices (flow, adjacent, and connect). It can be observed that the existence of crossroads is aligned with a good flow of traffic (flow), good street connectivity (connect), and high junctions adjacency (adjacent). However, the existence of T-junction (TR), long streets (length), and the number of junctions (junctions) are strongly non-linear with good flow, good connectivity, and high adjacency.

None of the potential street indices has a strong correlation with the averaged congestion data (Dashed Red Box, Figure 11). In fact, the highest correlation detected by the score is the adjacency index, which only provides a very weak negative correlation of 0.16. Nonetheless, it is also worth noting that the majority of the indices are consistent with the index rationale outlined in Section 2.3. For instance, the CR, TR, and blocked indices do have a positive correlation with congestion, whilst the NR, XR, halven, flow, adjacent, and connect indices have a negative correlation as predicted. Unexpectedly, the study's most basic indices, length and junction, are calculated to be positively correlated with congestion. This result contradicts the rationale derived from the cited research (Cubukcu, 2015). In summary, although significant collinearity occurs between street indexes, in this study case, there is no significant correlation between street indices and traffic congestion. Therefore, none of the indices can be chosen or ruled out without further investigation.

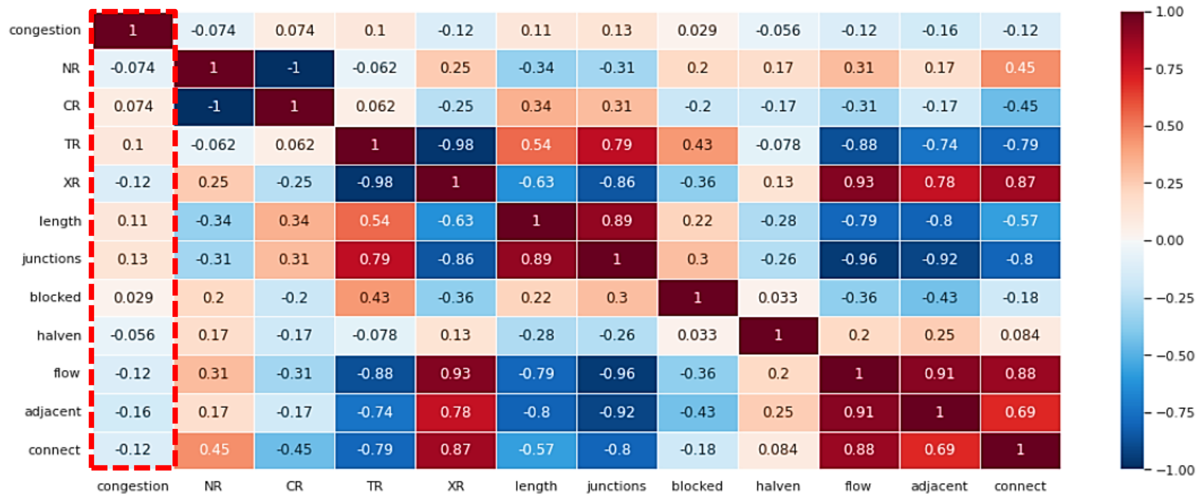


Figure 11. Correlation Test Result for All Variables (Dashed Red Box: Congestion Data)

To complement the correlation test result, the before-and-after scatter plots are created. The congestion data scatter plot in Figure 122 shows that the changes in congestion between t1 and t2 do not have a clear tendency. This means that, while the majority of data spreads are on the upper side of the identity line (indicating an increasing trend), there are also many data spreads below the identity line. This phenomenon does not correspond to any of the street indices trends in Figure 13. As can be seen, none of the indices data spreads resembles the congestion data's square-like spreads. This implies that, while traffic modification alters the value of street indices through its intervention, the tendency of the change in street indices does not resemble the change in congestion data.

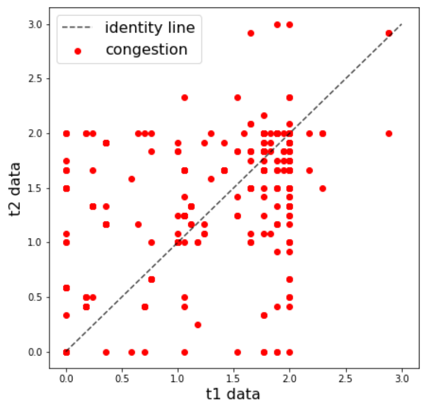


Figure 12. Before-After Scatter Plotting for Traffic Congestion Data

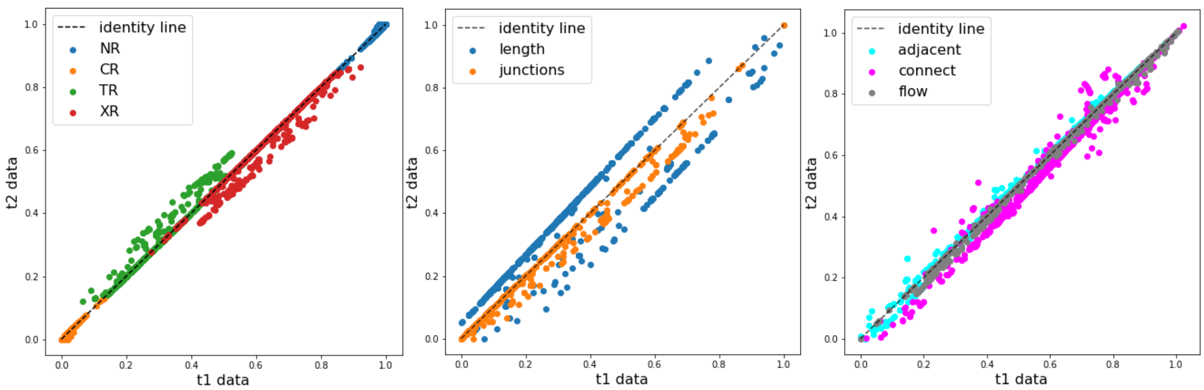


Figure 13. Before-After Scatter Plotting for Street Indices Data

Change detection mapping is the final selection analysis performed in this step. The colour pattern in the change detection map for congestion data (Figure 14), shows that congestion is primarily increased along the location A and on the southwest side of location B. When compared to the indices map, only the adjacent index prominently reflects this pattern. The distinction is that in the adjacent index, these are the areas where a decrease is detected. To put it another way, the location of adjacent index change is inversely related to the location of congestion change. Perhaps, this explains why the adjacent index got the highest Pearson correlation score. Furthermore, while the rest of the indices do not have the same change pattern as the congestion map, the location of the change in all of them is similar to the location of the change on the congestion map. The halven index, for example, detected an increase in location A, which is also true for the congestion map. In addition, the location of change is all placed inside and around location B in the NR, CR, XR, TR, junctions, blocked, length, and flow maps. This could explain why there is more congestion on the southwest side of location B. Moreover, the connect map shows that a significant decrease is detected exactly on the southwest side of location B, where congestion is increased. This phenomenon makes sense because, according to Lowry & Lowry (2014), the lower the connectivity, the greater the congestion.

Taken together, these three analyses do not yield any obvious result that can eliminate potential indices variables. It rather shows that, while the numerical correlation between traffic congestion data and street indices is weak, the locational correlation between them is substantial. With that said, to add more potential variables to this dataset, the longitude and the latitude coordinates of each traffic sensor will be included as explanatory variables. It is assumed since the street indices dataset as ratio data has no strong collinearity, with the integration of these nominal data (coordinate variables), the algorithm can better recognize the relationship tendency and generate a better traffic prediction with higher accuracy.

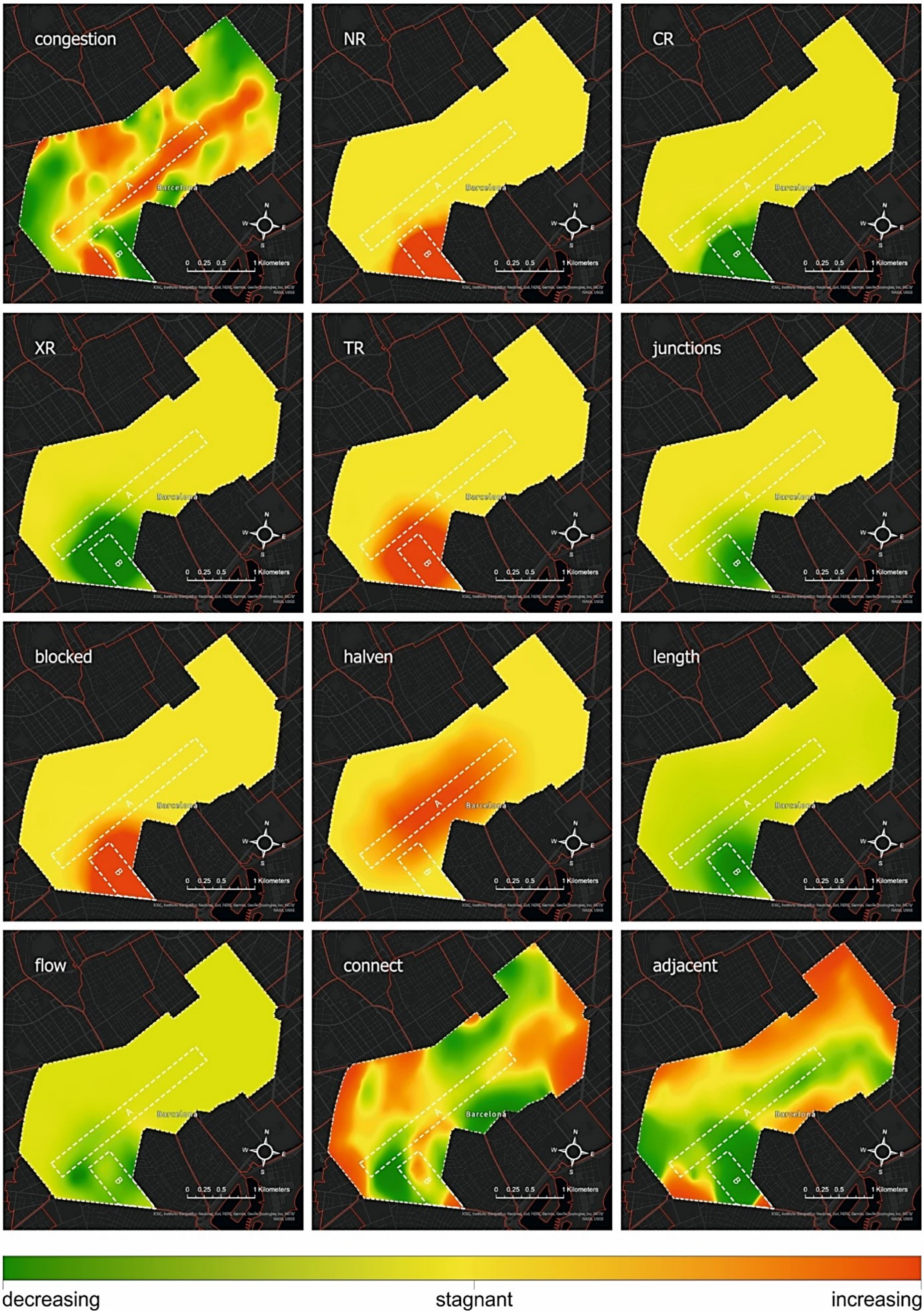


Figure 14. Change Detection Maps for All Study Variables

4.3. Street pattern maps: before and after superblock

Before classifying data into pattern types, unsupervised clustering analyses are performed to determine the optimal number of street clusters that should exist in the dataset. In this case, fuzzy c-means clustering reveals that the optimal cluster for the t1 and t2 datasets are formed in n=2 with FPC (fuzzy partition coefficient) of 0.426 (t1) and 0.381 (t2). Furthermore, as shown in Table 4, the second-highest FPC is measured in n=3 at 0.271 and 0.202 (for the clustering visualization please refer to Annex 4 in The Appendix).

Table 4. Fuzzy Clustering FPC Scores Result for t1 and t2 Dataset

Number of Clusters (n)	Dataset (t1)	Dataset (t2)
	FPC	FPC
2	0.426	0.381
3	0.271	0.202
4	0.178	0.154
5	0.115	0.110

In addition to the fuzzy clustering result, the dendrogram analysis for both the t1 and t2 datasets reveals that the first significant dissimilarity occurs when the dataset divides into three groups (Figure 15). Given how similar their hierarchical structures are, it can be concluded that, while traffic modification alters the value of the street indices, it does not significantly disrupt the hierarchical structure. Overall, the results of these two unsupervised analyses indicate that the optimal number of street pattern types in the dataset should be limited to two or three groups. Two as in the most optimal grouping (high FPC in fuzzy clustering and large dissimilarity gap in dendrogram) and three as the maximum number of allowed groups (acceptable FPC score compared to n>3 and the first occurrence of significant dissimilarity gap in dendrogram).

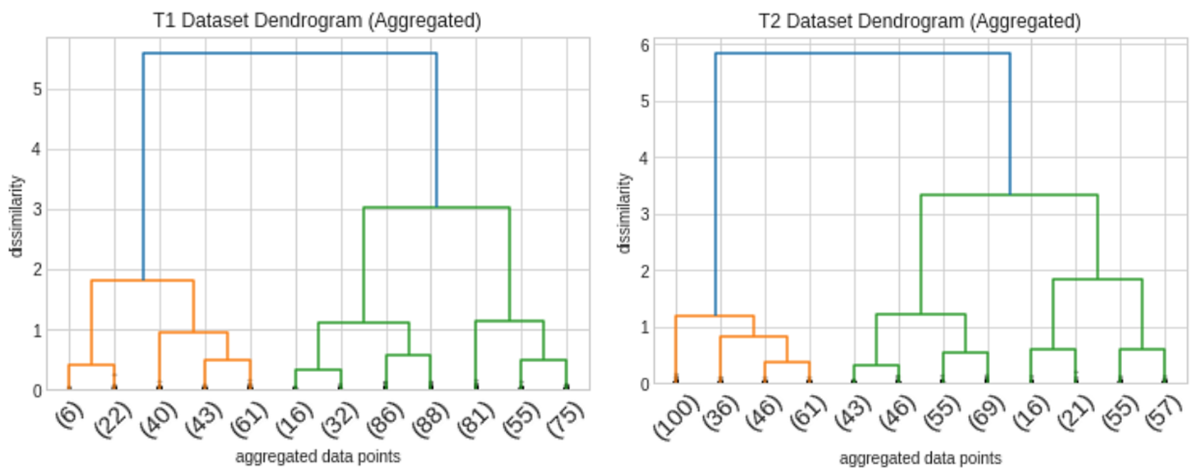


Figure 15. Dendrogram Graphs for T1 and T2 Dataset

In line with that result, the supervised classification of street patterns identified that three types of patterns exist in the t1 and t2 datasets (Figure 16). The noticeable difference is that in location B, the street pattern has changed from 'transition type' to 'T-type'. Aside from that, it should be noted that there is also an expansion of 'transition type' in the middle of location A (Figure 16, wider green networks in t2). This condition could explain why there is increasing congestion inside location A, but decreasing congestion inside location B. According to Chan et al., 2011b's study, the more grid streets there are, the less congested they should be. In this study case, the 'transition type' is a pattern type that is closest to a cul-de-sac and less similar to a grid. Thus it explains why the expansion of 'transition type' increases congestion in location A, while the disappearance of 'transition type' decreases congestion in location B. In summary, this step was successful in producing street pattern types for both the t1 and t2 datasets, and although there is a change in the street pattern inside locations A and B, the dominant street pattern type in both datasets is still the grid type. This pattern type data will then be used in conjunction with street indices to build a traffic congestion predictive model in the following section.

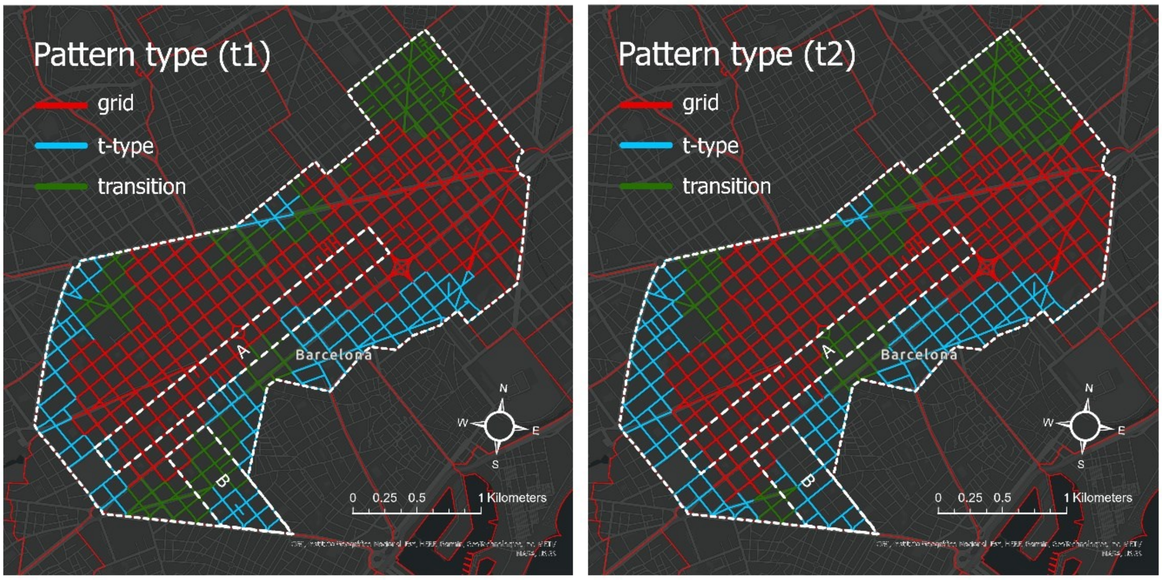


Figure 16. Comparison of The Detected Street Pattern Types, Before and After Superblock

4.4. The machine learning traffic prediction model

Table 5 and Figure 18 compare the performance of three machine learning models tested in the study. Table 5 shows that, while linear regression performed poorly with an r-square of 4% and an NRMSE of 0.25, the tree-based algorithms (decision tree and random forest) performed better with 22% and 58% r-square accuracy. Furthermore, the scatter graph between predicted and actual congestion (Figure 17) shows that the linear regression and decision tree models are not close to fitting the identity line. The random forest model, on the other hand, is more skewed diagonally, resembling the identity line. This graph confirms that the random forest algorithm is the best algorithm for predicting traffic congestion in this study case. Not only by r-square and NRMSE measurements but also visually in the scatter graph.

Table 5. R-square and NRMSE Scores for Each Machine Learning Algorithm

Measurement	Linear Regression (LR)	Decision Tree (DT)	Random Forest (RF)
R-square accuracy	4%	22%	**58%**
NRMSE value	0.26	0.23	**0.17**

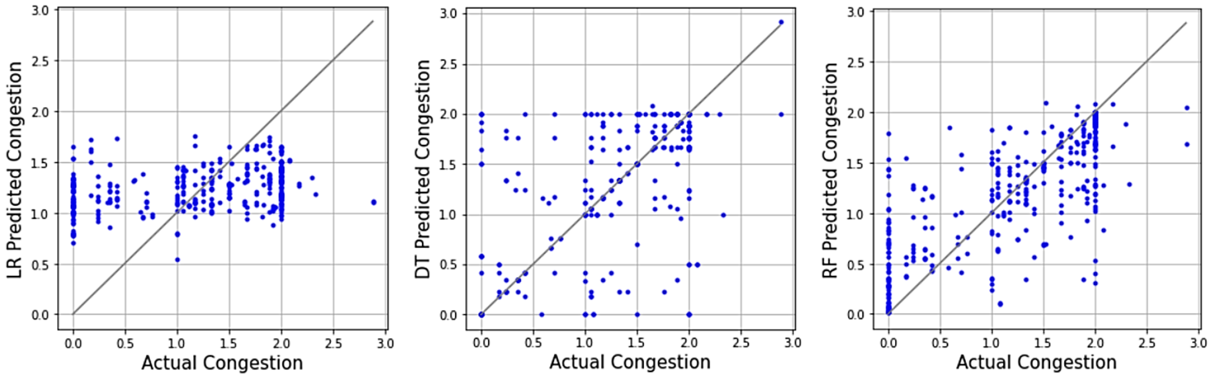


Figure 17. Scatter Graph for Each Machine Learning Algorithm

The feature importance analysis is a continuation of the machine learning modelling described in the previous paragraph. Because the random forest algorithm outperformed the other algorithms, the feature importance analysis was only applied to the random forest model. According to the analysis, the cul-de-sac ratio (CR), crossroad ratio (XR), and junction adjacency index play the most important roles in predicting traffic congestion in the model (Figure 18). In other words, the model's algorithm acknowledges that the presence of new cul-de-sacs and the disappearance of crossroads in the study area increases traffic congestion. Furthermore, the adjacent index, which measures the proximity of junctions, was also found to be an important feature. When this insight is combined with the results of the change detection maps, it is possible to conclude that lowering junctions adjacency (which occurs in location B) can increase traffic congestion. Overall, the feature importance analysis also shows that in this study, the street pattern is not a significant factor in predicting traffic congestion. Rather than a general street pattern, the fluctuation of traffic congestion is more sensitive to the change in street junction type and the proximity between those junctions.

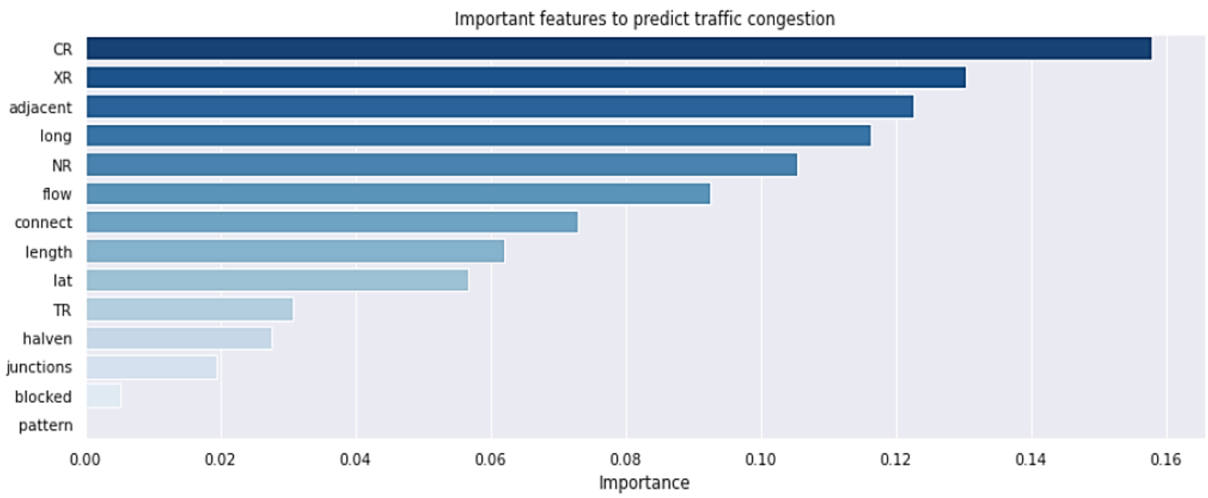


Figure 18. Feature Importance Visualization of The Random Forest Model

4.5. Performance of models based on the street pattern type

The final result that is produced in this study is the evaluation of model algorithms based on the street pattern types. From Table 6, it can be seen that for grid pattern type, the random forest performed best with 54% accuracy followed by linear regression and decision tree with 19% and 5% accuracy each. Furthermore, for the T-type pattern, it is calculated that random forests still manage to give the best model performance at 20% accuracy followed by linear regression and decision tree with 17% and 7% each. A notable result occurred in the 'transition type'. It is measured that for the 'transition type', the linear regression algorithm outperformed the random forest and decision tree quite significantly. The linear regression achieved the r-square accuracy of 45% while random forest and decision tree only managed to score 39% and 4% respectively.

This result demonstrates that, while the random forest algorithm performs best when all street patterns are combined, it does not consistently outperform other algorithms when they are separated. This also means that for future traffic model studies, before deciding on the best algorithm for the project or research, a comparison of algorithms is still needed to be performed. One cannot simply generalize their algorithm selection for traffic modelling tasks to the random forest.

Table 6. R-square and NRMSE Score Comparison for Different Street Pattern Type Model

Pattern	Parameter	Linear Regression	Decision Tree	Random Forest
Grid	R-square accuracy	19%	5%	**54%**
	NRMSE	0.24	0.26	**0.18**
T - Type	R-square accuracy	17%	7%	**20%**
	NRMSE	0.22	0.25	**0.22**
Transition	R-square accuracy	**45%**	4%	39%
	NRMSE	**0.27**	0.35	0.28

5. DISCUSSION

This chapter discusses five components. First, it discusses the implication of traffic modification. Second, it discusses the connection between street indices, street pattern types, and traffic congestion. Following that, it reviews the machine learning-based traffic model. And after that, it attempts to yield valuable insights for the superblock's next expansion stage. Finally, it investigates the study's limitations.

5.1. The implication of traffic modification

The study results show two distinct implications of traffic modification: a) a change in traffic congestion pattern and b) a change in street morphology (observed through the change detection mapping). The comparison of traffic maps reveals that there is an increase in the magnitude of congestion and the emergence of congestion hotspots following the superblock (Figure 10). This implies that the traffic modification alters the traffic pattern in the study area, particularly around intervention areas such as locations A and B. However, it is worth noting that the shift in traffic patterns may occur outside of the intervention area as well. This is since traffic congestion has a propagative character, which means that traffic incidents at one location point can unexpectedly spread to a larger area (Nguyen, Liu, & Chen, 2016). This phenomenon also occurs in this study. For example, it can be seen that overall traffic has increased significantly in the northeast area (Figure 10, t2 map). Although this area is not necessarily located near intervention locations, its road line is a continuation of location A's one-way road intervention. As a result, its increased traffic may be caused by the propagation effect from location A. This phenomenon shows that the implication of traffic modification on traffic patterns turns out to have a faint range effect and requires a more sensitive-comprehensive observational approach.

Traffic modification has an impact on street morphology as well. Figure 14 shows that the street indices value is fluctuating (which reflects the dynamic change in street morphology). If observed thoroughly, the blocking intervention (location B) is detected to increase the TR index value while decreasing the XR value. This indicates that this type of intervention has been detected to promote the conversion of a crossroad (TR) to a T-junction (TR). This type of intervention may result in a less efficient traffic flow and is therefore unsuitable for freight transport (Han et al., 2020). This finding deduced that a traffic modification has a different implication on street morphology. Thus, for future reference, a decision to deploy a certain traffic modification has to be followed by an understanding of the consequences that might happen if a certain street morphology is changed.

5.2. Relationship between the selected street indices, street pattern types, and traffic congestion

The study is based on the idea that traffic changes can cause incidental changes in street patterns, resulting in an unprecedented distribution of traffic flow and the emergence of new congestion hotspots. Many previous studies are aligned with this claim. For instance, it is explained that street patterns are assembled by street indices (Wang & Debbage, 2021a) and that many street indices have a significant correlation with traffic congestion (Choi & Ewing, 2021). Based on these premises, it is reasonable to assume that changes in street index values can alter street patterns and influence traffic congestion trends.

Having said that, these claims are not prominently reflected in the study results. Firstly, although the correlation tendency of street indices and traffic congestion is mostly aligned (the positive and negative correlation), none of them has a high correlation score that is deemed as significant (Section 4.2; Figure 11). Secondly, although street indices can be used to classify street networks into street pattern types, in this study, the street patterns that existed before the traffic modification (t1) and the street patterns that emerged after the traffic modification (t2) do not show striking change. Presumably, this is because the current stage of the superblock is still focusing on changing traffic orientation rather than changing the actual morphological structure of the streets. Ergo, there is not much street pattern conversion detected, visually and quantitatively.

Although there is not much street pattern conversion detected, new congestion hotspots have emerged and the overall congestion has increased. There are two possible explanations for this. First, perhaps traffic congestion that occurs post-traffic modification cannot be simply explained through the change of street patterns and should be examined in conjunction with other components (e.g. the change of land use zoning). Second, there is always a possibility that the selected indices are not properly prepared or not the right set for the task, thus it was not reflecting the patterns changes properly.

In summation, the relationship between the selected street indices, street pattern types, and traffic congestion in this study is noticeable but not significant. Keeping in mind all the limitations mentioned, their relationship does not have a strong quantitative correlation but does have a good spatial correlation (demonstrated by the fact that the increasing congestion is mostly spotted in the location where traffic modifications are implemented). These intricate relationships support the speculation that plain modelling such as linear modelling will not work well in the study case, therefore machine learning methods such as random forests and decision trees which are allegedly able to provide good predictions on non-linear data are deemed appropriate.

5.3. Review of the machine learning traffic model

The traffic model developed in the study is tested with three different machine learning algorithms, linear regression, decision tree, and random forest. From these three algorithms, random forest excels as the best approach with an accuracy of 58%, followed by a decision tree at 22% and linear regression at 4%. There are some assumptions on why this condition occurs. First, since nominal data such as latitude and longitude are included in the modelling, an algorithm that can only process ratio data such as linear regression is expected to perform poorly. In addition, tree-based algorithms such as decision tree and random forest are designed to handle nominal and ratio data subsequently (Satya Sree et al., 2021), hence, in this case, they performed better with the random forest edging decision tree due to its ensemble approach.

Other than the algorithm characteristic, this modelling result may also be caused by the dataset condition. For example, when the model is tested with the dataset that is already separated by street pattern type, the random forest algorithm is observed to not have consistently decent performance. As shown in Table 6, while random forest has 54% predictive accuracy in grid type datasets, it only has 20% and 39% in T-type and transition type datasets. This condition is most likely caused by the number of data entries that each street pattern type has. Figure 16 shows that the grid type is the most dominant street pattern in the study area. This means that the grid pattern has more data entries than the other pattern types, allowing the random forest algorithm to see more variation in training data, test more data entries into the trees, and thus have a higher probability of producing better accuracy.

Another aspect of the traffic model worth discussing is the feature importance analysis. Although the street pattern is expected to be the most important feature for traffic congestion prediction, in this study, the random forest model suggests that it is not. Perhaps, it is because pattern type is an aggregate of many street indices. Meaning classifying street networks into a street pattern type could significantly reduce the variability within those street network data. This condition can make it difficult for the machine learning algorithm to detect a noticeable tendency that leads to accurate traffic congestion prediction. To summarize, while the machine learning model in this study does not perform exquisitely well, it did produce additional insights into the relationship between traffic modification and traffic congestion that cannot be drawn through generalized traffic modelling such as O-D matrix, four-step travel model, or macroscopic modelling (e.g. give an insight about the importance of street junction conversion).

5.4. Valuable insights for the next expansion stage of the superblock

The study results provide some valuable insights that can be incorporated into the planning process for the next stage of the superblock. First, The Pearson correlation test can be used as an indication of what to expect when a certain street intervention is applied. Looking at Figure 11, for example, it can be deduced that the presence of T-junctions (TR) is negatively correlated with a good flow of traffic (as measured by the 'flow' index). Furthermore, the flow index is negatively related to traffic congestion. This means that when a crossroad is converted into a T-junction, there is a chance that traffic flow will decrease and traffic congestion will increase. In another case, the adjacent index is found to be highly correlated with the flow index. This correlation suggests that when street junctions are designed to be closer to each other, good traffic flow can be expected.

Other than the correlation test, the feature importance analysis from the traffic model can also be used as a consideration when designing the next stage of the superblock. As can be seen in Figure 18, cul-de-sac (CR), crossroads (XR), and the alteration of adjacency between junctions (Adjacent) are the most influential features to the dynamic of congestion in Barcelona. Hence, before expanding the implementation area of the superblock, is better to be aware of junction conversion that may happen. This is because, based on the findings of this study, it can be argued that one should a) avoid the conversion of any junctions into a cul-de-sac, b) avoid road blocking that made a big gap between junctions, and c) maintain the existence of crossroads.

5.5. Limitations

Even though the study met all of its research objectives, the findings and conclusions drawn from it should be interpreted with caution. First, while the Barcelona traffic sensor produced decent traffic data on a high time scale, the installation locations are rather dispersed. The density of sensors in the area is high in the district's centre, but there is a gap on the west side (Annex 2). This implies that the study's congestion mapping has a blind spot and may miss significant traffic activity that exists in that area but cannot be examined due to a sensor's lack. However, because the hollow spots only exist in a small portion of the study area and traffic modifications do not occur there, this spatial shortcoming does not disrupt nor cause anomalies in the study result. Second, the provided traffic dataset from the municipality of Barcelona has already been pre-processed and categorically labelled. This means that, while congestion mapping is still possible, the map's symbology does not represent a standard measurement of traffic volume (i.e V/C ratio). This also means that the dataset cannot be used for further investigation, such as traffic capacity analysis. However, because this is a comparison study, this limitation is deemed acceptable as long as the recorded traffic data has a consistent range of value.

Additionally, even though the study has carefully selected and calculated suitable street indices for the task, there is a possibility that other kinds of indices can capture a more significant relation between street patterns and traffic congestion. This particular limitation, however, is a subject of endless possibility, especially because the measurement and the calculation of street form is an everchanging subject that will evolve along with the development of GIS tools. Lastly, in comparison to the other urban street network studies, the geographic scope of this study is small. Therefore, the variations of street indices and street patterns collected are also less diverse when compared to other studies that collect network data across the country and continent. Having said that, taking into consideration the context of traffic modification. To the best of the writer's knowledge, the superblock is by far the most eligible study case due to its well-published masterplan and publicly available street data which is often not the case for most cities.

6. CONCLUSION AND RECOMMENDATION

6.1. Conclusions

In this study, using Barcelona's superblock policy as a case study, a machine learning modelling approach is proposed to better understand traffic congestion that emerges post the implementation of traffic modification. Following the set objectives, first, a traffic mapping comparison between before superblock and after superblock is conducted. It was then followed by the calculation of some selected street indices and a data exploration between these indices and traffic data. Next, a machine learning-based traffic modelling is performed by comparing linear regression, decision tree, and random forest algorithms, which were then followed by algorithms evaluation based on street pattern types.

Through the traffic congestion mapping, new congestion hotspots are spotted to emerge and the overall traffic congestion in the study area was noticed to be increased. After that, in data exploration, several analyses concluded that in this study case (given the mentioned limitations), there is no strong collinearity between the selected street indices and traffic congestion, but spatially, there is a significant correlation between them. On top of that, the study also did not find a striking change in street patterns post the implementation of the superblock. Furthermore, from the traffic modelling, it is concluded that for this study case, random forest algorithms outperformed other algorithms and yielded a feature importance analysis which suggests that street junction alteration is more influential to traffic conditions than street patterns alteration. Lastly, from the evaluation of the algorithms based on street pattern types, it is concluded that although in this study case random forest outperforms other algorithms, it does not have consistent accuracy in all types of the street pattern.

Overall, this study provides a novelty on how the machine learning approach can be used when a relationship between two variables is too intricate to investigate using traditional statistical methods. It also demonstrates that a predictive transport model can still be constructed from non-collinear data using an appropriate machine learning algorithm. And at last, it broadens our understanding of the effects of traffic modification on street pattern and traffic congestion.

6.2. Recommendations for further studies

Looking forward, as more urban data such as traffic records are generated at a higher velocity and volume, there is a potential for better insight into our cities. Here are some potential future research directions:

1. In the future, after the superblock has been expanded and running for a longer time, additional studies covering a larger study area can be proposed to include more variations of urban districts. If this proposed study is carried out when more areas have already been converted and more traffic changes have occurred, better results and insights can be expected.

2. This study's dataset can be used for further social research, such as an overlay of traffic congestion hotspots and demographic maps. This type of study can assist in identifying and delineating areas of high congestion - high economic vulnerability, which in most cases should be prioritized in infrastructure provision.

3. It will be interesting to construct a traffic model that takes into account temporal variables. This means that variables like the day of the week, time of day, and month of the year can be combined. This approach will provide a better understanding of the impact of traffic modifications and their relationship to time-related travel behaviour.

LIST OF REFERENCES

Abdulkareem, Ahmed, Adulaimi, Ahmed, Pradhan, Biswajeet, Chakraborty, Subrata, & Alamri, Abdullah (2021). Traffic Noise Modelling Using Land Use Regression Model Based on Machine Learning, Statistical Regression and GIS. *Energies*, *14*(16), 5095.

Ajuntament de Barcelona (2014). Urban Mobility Plan of Barcelona,. Barcelona. Retrieved from http://www.bcnecologia.net.

Ajuntament de Barcelona (2020). Towards Superblock Barcelona,. Barcelona. Retrieved from www.barcelona.cat/premsa.

Alpaydın, Ethem (2010). *Introduction to Machine Learning, The Wikipedia Guide. Natural Language Engineering* (*2nd ed.*, Vol. 19). Cambridge, MA: The MIT Press.

Alqudah, Nour, & Yaseen, Qussai (2020). Machine Learning for Traffic Analysis: A Review. *Procedia Computer Science*, *170*, 911–916.

Alwosheel, Ahmad, van Cranenburgh, Sander, & Chorus, Caspar G. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, *28*, 167–182.

Aman Gupta (2021). Fuzzy C-Means Clustering (FCM) Algorithm,. Retrieved from https://medium.com/geekculture/fuzzy-c-means-clustering-fcm-algorithm-in-machine-learning.

Ambühl, Lukas, Loder, Allister, Becker, Henrik, Menendez, Monica, & Axhausen, Kay W (2018). Evaluating London's congestion charge An approach using the macroscopic fundamental diagram. In *7th Transport Research Arena*. Vienna: Institute for Transport Planning and Systems.

Aronsson, Linus, & Bengtsson, Aron (2019). *Machine learning applied to traffic fore-casting*. University of Gothenburg, Sweden.

Badstuber, Nicole (2019). London congestion charge: what worked, what didn't, what next,. Retrieved from https://theconversation.com/london-congestion-charge-what-worked-what-didnt-what-next-92478.

Banerjee, Sudipto (2005). On Geodetic Distance Computations in Spatial Modeling. *Biometrics*, *61*(2), 617–625.

Bathaee, Yavar (2018). The Artifical Intelligence Black Box and The Failure of Intent and Causation. *Harvard Journal of Law & Technology*, *31*(2).

BBC News (2018). Minicabs including Uber to pay London congestion charge,. Retrieved from https://www.bbc.com/news/uk-england-london-46618762.

Bhatia, Parteek (2019). Introduction to Data Mining. *Data Mining and Data Warehousing*, *5*(2), 17–27.

Boeing, Geoff (2019). Urban spatial order: street network orientation, configuration, and entropy. *Applied Network Science*, *4*(1), 1–19.

Boeing, Geoff (2020). *Street Network Models and Indicators for Every Urban Area in the World* (*1st ed.*). Los Angeles: Department of Urban Planning and Spatial Analysis, University of Southern California.

Bratsas, Charalampos, Koupidis, Kleanthis, Salanova, Josep Maria, Giannakopoulos, Konstantinos, Kaloudis, Aristeidis, & Aifadopoulou, Georgia (2020). A comparison of machine learning methods for the prediction of traffic speed in Urban places. *Sustainability*, *12*(1).

Breiman, Leo (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Chan, S. H. Y., Donner, R. V., & Lämmer, S. (2011a). Urban road networks — spatial networks with universal geometric features? *The European Physical Journal B 2011 84:4*, *84*(4), 563–577.

Chan, S. H.Y., Donner, R. V., & Lämmer, S. (2011b). Urban road networks - Spatial networks with universal geometric features?: A case study on Germany's largest cities. *European Physical Journal B*, *84*(4), 563–577.

Chauhan, Nagesh Singh (2022). Decision Tree Algorithm, Explained,. Retrieved from https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html.

Choi, Dong ah, & Ewing, Reid (2021). Effect of street network design on traffic congestion and traffic safety. *Journal of Transport Geography*, *96*(November), 103200.

Christodoulou, Aris, & Christidis, Panayotis (2020). Measuring congestion in European cities. Luxembourg.

Climate & Clean Air Coalition (2015). Air pollution measures for Asia and the Pacific,. Retrieved from https://www.ccacoalition.org/en/content/air-pollution-measures-asia-and-pacific.

Cox, Nicholas J (2009). Speaking Stata: Paired, parallel, or profile plots for changes, correlations, and other

comparisons. *The Stata Journal*, *9*(4), 621–639.
Cozens, Paul, Hillier, David, Cozens Ã, Paul, & Hillier Ãã, David (2008). The Shape of Things to Come: New Urbanism, the Grid and the Cul-De-Sac. *International Planning Studies*, *13*(1), 51–73.
CSE India (2017). Congestion on Delhi roads has worsened,. Retrieved from https://www.cseindia.org/congestion-on-delhi-roads-has-worsened--6994.
Cubukcu, K. Mert (2015). Examining the Street Patterns in Izmir in the 19th Century: A network based spatial analysis. *Procedia - Social and Behavioral Sciences*, *202*, 436–441.
Cutler, Adele, Cutler, D Richard, & Stevens, John R (2012). Random Forests. In Cha Zhang & Yunqian Ma (Eds.), *Ensemble Machine Learning: Methods and Applications* (pp. 157–176). Utah: Springer.
Deng, Liu, Tai-Yi, & Yun, Luo (2019). Detecting Urban Polycentric Structure from POI Data. *ISPRS International Journal of Geo-Information*, *8*, 283.
Duchêne, Damien (2019). *Superblocks A Study of Public Life in Barcelona*. KTH Royal Institute of Technology, Stockholm.
ESRI (2018). Network Analysis,. Retrieved from https://support.esri.com/en/other-resources/gis-dictionary/term/8c565a36-9137-42ec-a600-6ce4f294e509.
ESRI (2020). How To: Detect and quantify temporal changes using the Raster Calculator tool,. Retrieved from https://support.esri.com/en/technical-article/000001209.
Evans, Richard H (1996). An Analysis of Criterion Variable Reliability in Conjoint Analysis. *Perceptual and Motor Skills*, *82*(3), 988–990.
Forina, M., Armanino, C., & Raggio, V. (2002). Clustering with dendrograms on interpretation variables. *Analytica Chimica Acta*, *454*(1), 13–19.
Friendly, Michael, & Denis, Daniel (2005). The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, *41*(2), 103–130.
Gaber, Yasser, Ahmed, Nasser, & Ali Farrag, Farrag (2021). Change detection for map updating using very high resolution satellite images. *JES. Journal of Engineering Sciences*, *0*(0), 424–445.
Gavrilyuk, Maxim, Vorob'Yova, Tat'yana, & Shalagina, Elena (2020). Effects of road blocking on traffic flows in Moscow. *Transportation Research Procedia*, *50*, 1–11.
Gössling, Stefan (2013). Urban transport transitions: Copenhagen, city of cyclists. *Journal of Transport Geography*, *33*, 196–206.
Han, Baorui, Sun, Dazhi, Yu, Xiaomei, Song, Wanlu, & Ding, Lisha (2020). Classification of urban street networks based on tree-like network features. *Sustainability*, *12*(2).
Hays, Tanna, Olds, Preston, & Spence, John (2016). Happiness and Traffic: An Analysis of Long Term Effects. Denver.
Heidari Soureshjani, Milad (2016). Assessing Physical-spatial Impacts of Superblocks and Inner Urban Historical Fabrics Adjacency, Case study analysis of Great Mosque historical context of ShahreKord. *International Journal of Advances in Agricultural & Environmental Engg*, *3*(2).
Helm, J. Matthew, Swiergosz, Andrew M, Haeberle, Heather S, Karnuta, Jaret M, Schaffer, Jonathan L, Krebs, Viktor E, & Ramkumar, Prem N (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Musculoskeletal Medicine*, *13*(1), 69–76.
Hung, Ming Chuan, & Yang, Don Lin (2001). An efficient fuzzy C-means clustering algorithm. *Proceedings - IEEE International Conference on Data Mining, ICDM*, (September), 225–232.
Jenkins-Smith, Hank, Ripberger, Joseph, Copeland, Gary, Nowlin, Matthew, Hughes, Tyler, Fister, Aaron, & Wehde, Wesley (2017). *Quantitative Research Methods for Political Science, Public Policy and Public Administration* (*3rd ed.*). Oklahoma: University of Oklahoma Library.
Jiang, Yinan, Kang, Rui, Li, Daqing, Guo, Shengmin, & Havlin, Shlomo (2017). Spatio-temporal propagation of traffic jams in urban traffic networks. *Urban Geography*, *3*(4), 56–85.
Kaparias, Ioannis, Liu, Peng, Tsakarestos, Antonios, Eden, Niv, Schmitz, Pierre, Hoadley, Suzanne, & Hauptmann, Susanna (2020). Predictive road safety impact assessment of traffic management policies and measures. *Case Studies on Transport Policy*, *8*(2), 508–516.
Keim, Daniel A., Hao, Ming C., Dayal, Umeshwar, Janetzko, Halldor, & Bak, Peter (2010). Generalized scatter plots. *Information Visualization*, *9*(4), 301–311.
Knoop, Victor L. (2017). *Introduction to Traffic Flow Theory: An introduction with exercises* (*1st ed.*). Delft University of Technology.
Laetz, Thomas J. (1990). Predictions and perceptions: Defining the traffic congestion problem. *Technological Forecasting and Social Change*, *38*(3), 287–292.

Lamíquiz, Patxi J., & López-Domínguez, Jorge (2015). Effects of built environment on walking at the neighbourhood scale. A new role for street networks by modelling their configurational accessibility? *Transportation Research Part A: Policy and Practice*, *74*, 148–163.
Lau, Johann (2020). Google Maps 101: How AI helps predict traffic and determine routes,. Retrieved from https://blog.google/products/maps/google-maps-101-how-ai-helps-predict-traffic-and-determine-routes/.
Lee, Yong Ju, & Min, Okgee (2018). Long Short-Term Memory Recurrent Neural Network for Urban Traffic Prediction: A Case Study of Seoul. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, *2018-November*, 1279–1284.
Li, Xiao, Wang, Yu, Basu, Sumanta, Kumbier, Karl, & Yu, Bin (2019). A debiased MDI feature importance measure for random forests. *Advances in Neural Information Processing Systems*, *32*(NeurIPS).
Li, Yingcheng, Xiong, Weiting, & Wang, Xingping (2019). Does polycentric and compact development alleviate urban traffic congestion? A case study of 98 Chinese cities. *Cities*, *88*, 100–111.
Liu, Yongli, Zhang, Xiaoyang, Chen, Jingli, & Chao, Hao (2019). A Validity Index for Fuzzy Clustering Based on Bipartite Modularity. *Journal of Electrical and Computer Engineering*, *2019*.
Louf, Rémi, & Barthelemy, Marc (2014). A typology of street patterns. *Journal of The Royal Society Interface*, *11*(101).
Lowry, John H., & Lowry, Michael B. (2014). Comparing spatial metrics that quantify urban form. *Computers, Environment and Urban Systems*, *44*, 59–67.
Lundberg, Ian, Johnson, Rebecca, & Stewart, Brandon M. (2021). What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory: *https://doi.org/10.1177/00031224211004187*, *86*(3), 532–565.
Lynch, H. J., Crosbie, K., Fagan, W. F., & Naveen, R. (2010). Spatial patterns of tour ship traffic in the Antarctic Peninsula region. *Antarctic Science*, *22*(2), 123–130.
Malik, Sumbal, El Sayed, Hesham, Khan, Manzoor Ahmed, & Khan, Muhammad Jalal (2021). Road Accident Severity Prediction - A Comparative Analysis of Machine Learning Algorithms. *2021 IEEE Global Conference on Artificial Intelligence and Internet of Things, GCAIoT 2021*, 69–74.
Martínez Sotoca, José, & Pla, Filiberto (2010). Supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recognition*, *43*(6), 2068–2081.
Mesa, Héber, & Restrepo, Guillermo (2008). On dendrograms and topologies. *Match*, *60*(2), 371–384.
Ministry of Transport New Zealand (2019). Defining Congestion. Auckland.
Nagy, Attila M., & Simon, Vilmos (2021). Improving traffic prediction using congestion propagation patterns in smart cities. *Advanced Engineering Informatics*, *50*, 101343.
Nayak, Janmenjoy, Naik, Bighnaraj, & Behera, H. S. (2015). Fuzzy C-means (FCM) clustering algorithm: A decade review from 2000 to 2014. *Smart Innovation, Systems and Technologies*, *32*, 133–149.
Nguyen, Hoang, Liu, Wei, & Chen, Fang (2016). Discovering Congestion Propagation Patterns in Spatio-Temporal Traffic Data. *IEEE Transactions on Big Data*, *3*(2), 169–180.
Nielsen, Frank (2016). Hierarchical Clustering. In *Introduction to HPC with MPI for Data Science* (pp. 195–211). Zurich: Springer.
Noorlander, Martijn (2021). *Predicting traffic flow with machine learning*. University of Twente.
Patorniti, Nicholas P, Stevens, Nicholas, & Salmon, Paul Matthew (2020). Catalysing Urban Redevelopment to Optimise the Existing Use and Future Delivery of Energy and Water Services. In *Proceedings of The 2nd Sustainable Road Freight Conference*. Sippy Downs.
Peng, Hao, Du, Bowen, Liu, Mingsheng, Liu, Mingzhe, Ji, Shumei, Wang, Senzhang, … He, Lifang (2021). Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning. *Information Sciences*, *578*, 401–416.
Peponis, J, Allen, D, Haynie, S Dawn, Scoppa, Martin, & Zhang, Z (2007). Measuring the configuration of street networks. *Proceedings of the Sixth International Space Syntax Symposium*, *2*, 1–16.
Pramanik, Md Aktaruzzaman, Rahman, Md Mahbubur, Anam, A. S.M.Iftekhar, Ali, Amin Ahsan, Amin, M. Ashraful, & Rahman, A. K.M.Mahbubur (2021). Modeling Traffic Congestion in Developing Countries Using Google Maps Data. *Advances in Intelligent Systems and Computing*, *1363 AISC*(October), 513–531.
Prasad Pai (2021). Hierarchical clustering explained,. Retrieved from https://towardsdatascience.com/hierarchical-clustering-explained-e59b13846da8.
Riccardo (2012). Change Detection in GIS,. Retrieved from https://www.gislounge.com/change-

detection-in-gis/.
Ruktanonchai, Nick Warren, Ruktanonchai, Corrine Warren, Floyd, Jessica Rhona, & Tatem, Andrew J. (2018). Using Google Location History data to quantify fine-scale human mobility. *International Journal of Health Geographics 2018 17:1*, *17*(1), 1–13.
Sangare, Mamoudou, Gupta, Sharut, Bouzefrane, Samia, Banerjee, Soumya, & Muhlethaler, Paul (2021). Exploring the forecasting approach for road accidents: Analytical measures with hybrid machine learning. *Expert Systems with Applications*, *167*, 113855.
Satya Sree, K. P.N.V., Karthik, Jayavarapu, Niharika, Ch, Srinivas, P. V.V.S., Ravinder, N., & Prasad, Chitturi (2021). Optimized Conversion of Categorical and Numerical Features in Machine Learning Models. *Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2021*, 294–299.
Seber, George A.F., & Lee, Alan J. (2003). *Linear Regression Analysis* (*2nd ed.*). Hoboke, New Jersey: John Wiley & Sons, Inc.
Sedgwick, Philip (2012). Pearson's correlation coefficient. *BMJ Clinical Research*, *345*(7864).
Sharifi, Ayyoob (2019). Resilient urban forms: A review of literature on streets and street networks. *Building and Environment*, *147*, 171–187.
Shi, Jia, Lee, Ching Hung, Guo, Xuesong, & Zhu, Zhengwei (2020). Constructing an integrated stakeholder-based participatory policy evaluation model for urban traffic restriction. *Technological Forecasting and Social Change*, *151*, 119839.
Słomska-Przech, Katarzyna, Panecki, Tomasz, & Pokojski, Wojciech (2021). Heat maps: Perfect maps for quick reading? comparing usability of heat maps with different levels of generalization. *ISPRS International Journal of Geo-Information*, *10*(8).
Song, Yan, & Knaap, Gerrit Jan (2007). Measuring Urban Form: Is Portland Winning the War on Sprawl? *https://doi.org/10.1080/01944360408976371*, *70*(2), 210–225.
Soriguera, F., & Robusté, F. (2011). Estimation of traffic stream space mean speed from time aggregations of double loop detector data. *Transportation Research Part C: Emerging Technologies*, *19*(1), 115–129.
Supriana, Ferhat Januar Rediat, Siregar, Martha Leni, Tangkudung, Ellen Sophie Wulan, & Kusuma, Andyka (2020). Evaluation of Odd-Even Vehicle Registration Number Regulation Before and After Expansion of the Rule in Jakarta. In *2nd International Symposium on Transportation Studies in Developing Countries* (Vol. 193) (pp. 151–156). Depok: Atlantis Press.
Surhone, L M, Timpledon, M T, & Marseken, S F (2010). *NCSS Statistical Software*. Riga: VDM Publishing.
SWARCO (2018). What is the transport mode shift?,. Retrieved from https://www.swarco.com/stories/what-transport-mode-shift.
The Jakarta Post (2019). Odd-even traffic policy trial is a success,. Retrieved from https://www.thejakartapost.com/news/2016/08/26/odd-even-traffic-policy-trial-is-a-success-police.html.
Tiwari, Geetam (2003). Transport and land-use policies in Delhi 1. *Bulletin of the World Health Organization*, *81*, 444–450.
Topirceanu, Alexandru, Iovanovici, Alexandru, Cosariu, Cristian, Udrescu, Mihai, Prodan, Lucian, & Vladutiu, Mircea (2016). Social cities: Redistribution of traffic flow in cities using a social network approach. *Advances in Intelligent Systems and Computing*, *356*, 39–49.
Tran, Hieu (2019). *Survey of Machine Learning and Data Mining Techniques used in Multimedia System*. *The Library of University of Texas*. The University of Texas, Dallas.
Tsekeris, Theodore, & Geroliminis, Nikolas (2013). City size, network structure and traffic congestion. *Journal of Urban Economics*, *76*(1), 1–14.
Wang, Mingshu, & Debbage, Neil (2021a). Urban morphology and traffic congestion: Longitudinal evidence from US cities. *Computers, Environment and Urban Systems*, *89*, 101676.
Wang, Mingshu, & Debbage, Neil (2021b). Urban morphology and traffic congestion: Longitudinal evidence from US cities. *Computers, Environment and Urban Systems*, *89*, 101676.
Werner, Rolf, Todorov Valev, Dimitar, & Lyubenov Danov, Dimitar (2009). The Pearson's correlation-a measure for the linear relationships between time series? *Fundamental Space Research*, (September).
Wilbur, Antoinette (2006). Traffic Detector Handbook (*3rd ed.*). Virginia: Turner-Fairbank Highway Research Center.
Wolhuter, Keith (2015). *Geometric Design of Roads Handbook*. *CRC Press* (*1st ed.*). Boca Raton: Talyor & Francis Groups.

Wolz, Laura, Abdalla, Filipe B., Alonso, David, Blake, Chris, Bull, Philip, Chang, Tzu Ching, … Shaw, Richard (2014). Foreground subtraction in intensity mapping with the SKA. *Proceedings of Science*, *7*(5), 14–28.

Wu, Chao Yun, Hu, Mao Bin, Jiang, Rui, & Hao, Qing Yi (2021). Effects of road network structure on the performance of urban traffic systems. *Physica A: Statistical Mechanics and its Applications*, *563*, 125361.

Xiong, Haoyi, Vahedian, Amin, Zhou, Xun, Li, Yanhua, & Luo, Jun (2018). Predicting traffic congestion propagation patterns: A propagation graph approach. *IWCTS 2018 - Proceedings of the 11th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, *18*, 60–69.

Xu, Lu, He, Zongyi, Zhang, Shaoqing, & Guo, Yan (2009). The comparative study of three methods of remote sensing image change detection. *2009 17th International Conference on Geoinformatics, Geoinformatics 2009*, 1595–1598.

Yiu, Tony (2019). Understanding Random Forest. How the Algorithm Works and Why It Is So Effective,. Retrieved from https://towardsdatascience.com/understanding-random-forest-58381e0602d2.

Yu Feng (2021). Fuzzy C-Means Clustering with Python,. Retrieved from https://towardsdatascience.com/fuzzy-c-means-clustering-with-python-f4908c714081.

Zambrano-Martinez, Jorge Luis, Calafate, Carlos T., Soler, David, Lemus-Zúñiga, Lenin Guillermo, Cano, Juan Carlos, Manzoni, Pietro, & Gayraud, Thierry (2019). A centralized route-management solution for autonomous vehicles in urban areas. *Electronics*, *8*(7), 1–21.

Zhang, Jun, Zhang, Xinxin, Yang, Yanni, & Zhou, Bing (2020). Study on the influence of one-way street optimization design on traffic operation system. *Measurement and Control (United Kingdom)*, *53*(7–8), 1107–1115.

Zhao, Guoming, Quan, Donghui, Zhang, Xia, Dahlan, Ahmad, Harahap, E, Darmawan, D, & Badruzzaman, F H (2020). Simulation of Traffic T-Junction at Cibiru-Cileunyi Lane Using SimEvents MATLAB. *Journal of Physics: Conference Series*, *1613*(1), 012074.

Zhao, Pengjun, & Hu, Haoyu (2019). Geographical patterns of traffic congestion in growing megacities: Big data analytics from Beijing. *Cities*, *92*, 164–174.

Zhao, Ying, & Karypis, George (2005). Hierarchical clustering algorithms for document datasets. *Data Mining and Knowledge Discovery*, *10*(2), 141–168.

APPENDIX

Annex 1: Electrical instrument detail of the inductive loop traffic sensor

The electrical instrument of the inductive loop sensor consists of four parts: wire inductive loop detectors (embedded in road pavement), a lead-in wire from the detector to a pull box, a lead-in cable connecting the pull box to the controller, and the controller cabinet which contain counter and data transmitter (Figure 21).

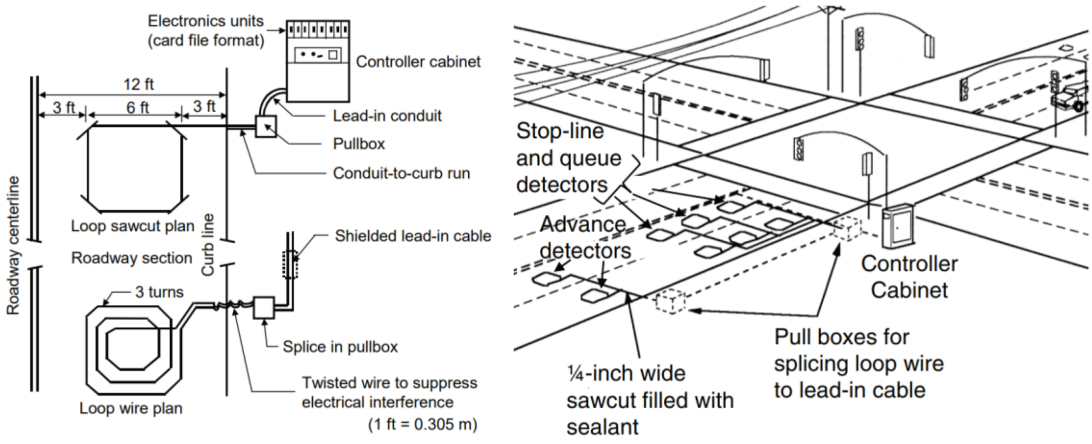


Figure 19. Detail Schema of Inductive-Loop Unit (Wilbur, 2006)

Annex 2: Sensor locations in Eixample District, Barcelona

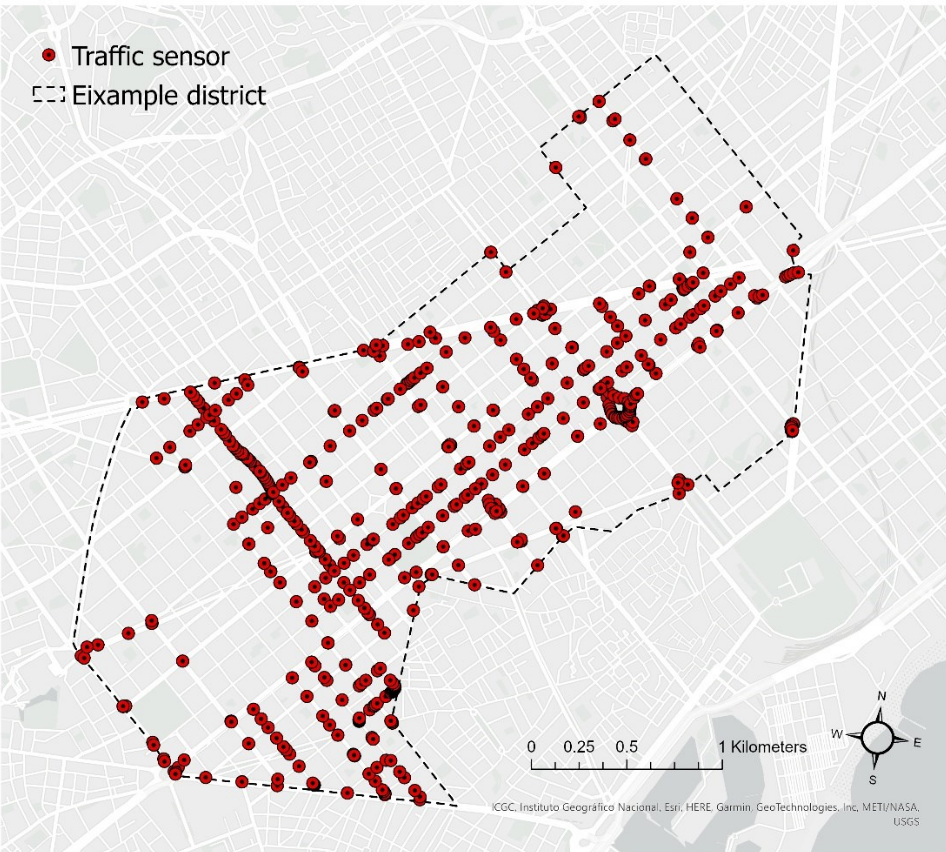


Figure 20. Map of Traffic Sensor Locations in Eixample District

Annex 3: Further explanation of the decision tree

To give a clearer picture, if there is a task to predict whether a Flamingo belongs to a group of mammals or non-mammals, the decision tree algorithm can make the first splitting task in the root node based on the body temperature feature (cold or warm). Once it got processed, the data will be classified based on the 'gives birth' feature and proceeded into the group of mammals or non-mammals. From this example, it can be elaborated that the first split of body temperature is the root node, the classification based on gives birth feature is the decision node, and the final prediction of mammals or non-mammas is the terminal node/leaf (Figure 21).

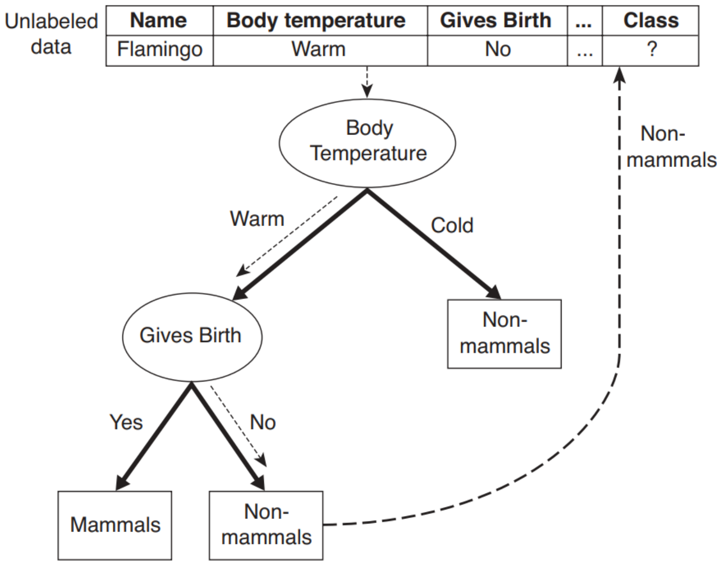


Figure 21. Example of Decision Tree Prediction Process (Bhatia, 2019)

Annex 4: Fuzzy c-means clustering visualization

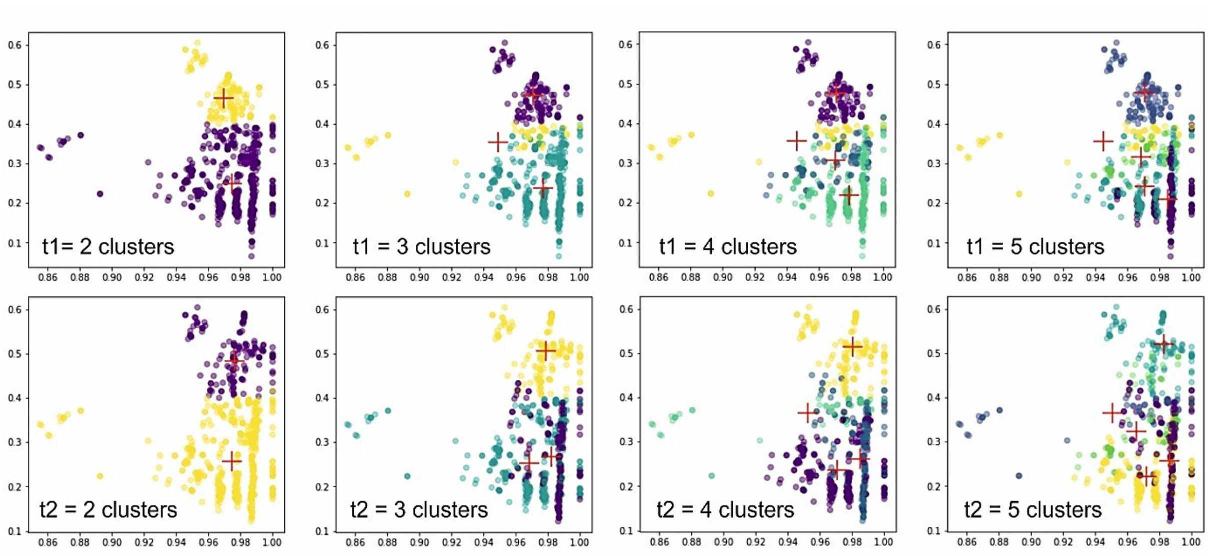


Figure 22. Fuzzy C-Means Clustering Visualization for t1 and t2 Datasets