CHAIR OF URBAN STRUCTURE AND TRANSPORT PLANNING

TECHNICAL UNIVERSITY OF MUNICH

Master’s Thesis

Understanding Mode Choice using Values and Attitudes

A Structural Equation Modelling and Machine Learning Approach

Author:	Sana Jaafar
Matriculation:	03675619
Supervisor:	Dr.-Ing. David Durán Rodas
Advisor:	Dr.-Ing. Benjamin Büttner
Submission Date:	May 10, 2023

Hereby, I confirm that this document, corresponding to the report of the study project for the MSc. Environmental Engineering program, is my own work and I have documented all sources and material used.

Munich, May 10, 2023

Sana Jaafar

Sana Jaafar

Lehrstuhl für Siedlungsstruktur und Verkehrsplanung
Ingenieurfakultät Bau Geo Umwelt
Technische Universität München

Exposé- Master Thesis

Study program M.Sc. Environmental Engineering

Sana Jaafar
Matr.-Nr. 0367519

Processing period: six months 01.12.2022 - 31.05.2023

Title: Can personal values predict mode choice?

Background:

The transport sector is facing the challenge of majorly reducing CO_2 emissions in order to reach the ambitious goal of climate neutrality by 2050 [1]. This goal contradicts the fact the transport sector is the only sector that did not see any reduction in CO_2 emissions between 1990 and 2020 [2]. In fact, it was responsible for one-fifth of the global CO_2 emissions in 2018 [3]. Specifically, passenger road vehicles account for the highest percentage of overall transport emissions [3], making it the primary target for improvement. Currently, the supremacy of passenger car transport is seen across the EU, reaching a share higher than 80 percent in 2020 [4]. In order to understand this phenomenon, some researchers have explored the relationship between travel behavior and sociodemographics, and land use [14,15,16]. Other researchers have focused on how attitudes and perceptions affect travel behavior [17,18]. Nevertheless, there is still an incomplete understanding of the underlying cognitive process that results in a person's behavior, specifically travel behavior [5].

Abstract psychological concepts, such as values, attitudes, and perceptions, are essential to this process [6-7-8]. Values are beliefs and motivational constructs that refer to desirable and abstract goals. They are a subjective evaluation of abstract ideas that are tied to emotions [9]. As values lay at the core of an individual's belief system, they reflect the basic traits of adaptation, hence why they are considered to be the basis on which attitudes and behaviors are later constructed [6]. Accordingly, attitudes and behaviors are subjected to change or evolution during one's life, but values consist of more enduring beliefs that go beyond circumstances, objects, and issues [6]. Although various studies that have explored travel behavior have mentioned the need to include values as influencing factors, most travel behavior analyses are mainly focused on attitudes [10, 11, 12].

Homer and Kahle [13] proposed a value-attitude-behavior hierarchy model that has been successful in examining the impact of values on behavior [13]. This model implies that values influence individual's behavior through attitudes, and it exists within a cognitive hierarchy in which a causal sequence of values (more abstract cognitions) to attitudes (mid-range cognitions) to behaviors can be depicted [13].

Further exploration of the effect of values on travel mode choice may support the understanding of people's choices and thereby, the appropriate policies that may support the choice of more sustainable travel modes. Therefore, the aim of this thesis is to explore the relationship between values and travel mode choice. To reach this objective the main research questions are:

- Are values associated with travel mode choice?
- Can only values help predict travel mode choice or instead in conjunction with socio-demographic characteristics?
- What modeling techniques are best suited to associate values with travel mode choices?

Methodology:

To achieve the obove mentionned objectives, the following methodology is proposed: first a literature review on psychological theories used to understand human behavior in general is conducted, then I will focus on mode choice specifically. After that, personal values in psychology are researched, and surveys used to assess them are selected. Following the literature review results, a survey will be designed and data collection in the city of Munich will start. Once the data is collected, the relationship between values and travel mode choice is explored through correlation, modeling techniques like factor analysis, PCA and clustering. The expected outcome is to identify the best suited modeling techniques, which give as output an association between values and travel mode choices.

Supervision:

The candidate will present to his supervisor a draft of the structure for his master thesis and a work plan two weeks after this approval. Other supervision meetings will be planned with the candidate when necessary. The Chair of Urban Structure and Transport Planning supports the candidate with contact to relevant actors and or experts if needed. After two weeks after the submission of his thesis, the candidate must defend it by means of a presentation (20 minutes) and the following discussion. The results are the responsibility of the author. The Chair does not take responsibility for those results.

Dr.-Ing. Benjamin Büttner

Dr.-Ing. David Telmo Duran Rodas

References

[1]“Delivering the European green deal,” *European Commission - European Commission*. https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal/delivering-european-green-deal_en (accessed Nov. 15, 2022).

[2]Eurostat, “Greenhouse gas emissions falling in most source sectors,” *Eurostat*, Aug. 23, 2022. Accessed: Nov. 15, 2022. [Online]. Available: https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/ddn-20220823-1

[3]“Cars, planes, trains: where do CO2 emissions from transport come from?,” *Our World in Data*. https://ourworldindata.org/co2-emissions-from-transport (accessed Nov. 15, 2022).

[4]“EU-Transport,” *German Federal Statistical Office*. https://www.destatis.de/Europa/EN/Topic/Key-indicators/Transport.html;jsessionid=863F9027C5F5973B275A96F5FD48E72F.live742 (accessed Nov. 15, 2022).

[5]M. Paulssen, D. Temme, A. Vij, and J. L. Walker, “Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice,” *Transportation*, vol. 41, no. 4, pp. 873–888, Nov. 2013, doi: 10.1007/s11116-013-9504-3.

[6]I. Ajzen, “Values, Attitudes, and Behavior,” in *Methods, Theories, and Empirical Applications in the Social Sciences*, Wiesbaden: VS Verlag für Sozialwissenschaften, 2012, pp. 33–38. Accessed: Nov. 15, 2022. [Online]. Available: http://dx.doi.org/10.1007/978-3-531-18898-0_5

[7]E. J. Ihemezie, M. Nawrath, L. Strauß, L. C. Stringer, and M. Dallimer, “The influence of human values on attitudes and behaviours towards forest conservation,” *Journal of Environmental Management*, vol. 292, p. 112857, Aug. 2021, doi: 10.1016/j.jenvman.2021.112857.

[8]G. Cicia, M. Furno, and T. Del Giudice, “Do consumers’ values and attitudes affect food retailer choice? Evidence from a national survey on farmers’ market in Germany,” *Agricultural and Food Economics*, vol. 9, no. 1, Jan. 2021, doi: 10.1186/s40100-020-00172-2.

[9]S. Schwartz, “Toward Refining the Theory of Basic Human Values,” in *Methods, Theories, and Empirical Applications in the Social Sciences*, Wiesbaden: VS Verlag für Sozialwissenschaften, 2012, pp. 39–46. Accessed: Nov. 15, 2022. [Online]. Available: http://dx.doi.org/10.1007/978-3-531-18898-0_6

[10]M. Vredin Johansson, T. Heldt, and P. Johansson, “The effects of attitudes and personality traits on mode choice,” *Transportation Research Part A: Policy and Practice*, vol. 40, no. 6, pp. 507–525, Jul. 2006, doi: 10.1016/j.tra.2005.09.001.

[11]J. De Vos, D. Ettema, and F. Witlox, “Changing travel behaviour and attitudes following a residential relocation,” *Journal of Transport Geography*, vol. 73, pp. 131–147, Dec. 2018, doi: 10.1016/j.jtrangeo.2018.10.013.

[12]S. Y. He and J. Thøgersen, “The impact of attitudes and perceptions on travel mode choice and car ownership in a Chinese megacity: The case of Guangzhou,” *Research in Transportation Economics*, vol. 62, pp. 57–67, Jun. 2017, doi: 10.1016/j.retrec.2017.03.004.

[13]P. M. Homer and L. R. Kahle, “A structural equation test of the value-attitude-behavior hierarchy.,” *Journal of Personality and Social Psychology*, vol. 54, no. 4, pp. 638–646, Apr. 1988, doi: 10.1037/0022-3514.54.4.638.

Acknowledgments

First and foremost, I would like to thank my mentor and supervisor David for his incredible support, guidance and patience. You gave me the freedom and space to take this Master's thesis in the direction I felt is best, which has allowed me to learn and grow so much over the past few months. The best part of this work was discussing methods and ideas with you.

I would like to thank my friends, a special shout-out goes to Fatma, Fai, Natalie and my brother Mhamed, for listening to me when I talked about this thesis, sometimes complaining and sometimes sharing my enthusiasm. Your presence kept me sane and made going forward seem less daunting. And finally, I would like to thank my parents. Without the unconditional love and trust you constantly show me, I would not be here.

Abstract

Reliance on the private car remains a widespread problem. This behaviour results in the private car accounting for the largest share of total transport emissions. Policymakers have taken various measures to limit the use of private cars, but their effectiveness has not been as hoped. This raises the question whether the complexity of mobility patterns goes beyond the analysis of daily trips and requires new approaches. Psychological factors influencing mode choice are still under-researched. Therefore, exploring and understanding the relationship between psychological factors such as values and attitudes and transport mode choice would be a valuable gap to fill.

Based on a survey conducted in Munich between October and December 2019, 250 valid responses could be evaluated. The value-attitude-behaviour hierarchy was tested using structural equation modelling. Four machine learning algorithms were trained with different input data to determine whether predicting transport mode choice using values and attitudes can lead to accurate predictions. The fit indices for the structural model were acceptable, with a CFI of 0.86, a TLI of 0.83 and a RMSEA of 0.07. The results suggest that not all values are causally related to attitudes and mode choice. However, openness to change was found to be causally related to attitudes towards cycling, which in turn were related to private bike use. On the other hand, conservation was to was found to be causally related to attitudes towards car driving, which in turn were related to private car use.

Four machine learning algorithms were used to predict mode choice: K-Nearest Neighbours (KNN), Random Forest (RF), Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost). RF, SVM and XGBoost performed comparatively. In general, combining values with attitudes and socio-demographic variables made more accurate predictions. Shap values were used to determine the effect of the input feature on the output of the model. For private car use, age, income, attitude of "car driving is relaxing" and value conservation were the most important features that influenced the model. Regarding bike use, the attitudes "cycling is convenient" and "cycling is fun" the value conservation were the three features with the greatest influence on the model output. Finally, for public transport use, the category student in employment status, age between 18 and 29 and the value of self-transcendence had the greatest influence on the model result.

This thesis highlights the importance of values and attitudes in predicting mode choice, such as the value of conservation, openness to change and attitudes towards fun for cycling and relaxation for car driving. Based on these findings, it is recommended that cycling be promoted to people who hold the openness to change values and that it be promoted as a fun mode of transport rather than an environmentally friendly one. Future work could consider a larger and more diverse sample. A combination of structural equation modelling and machine learning is useful to understand the complex interactions between values, attitudes and mode choice.

Contents

1. Introduction

1.1. Motivation

Climate change is a global challenge that affects every aspect of our lives. From air and water pollution to rising global temperatures and diminishing natural resources, the impacts of climate change are a reality that must be prioritized [1]. One strategy to amend the effects of climate change is to reduce the amount of greenhouse gases released into the atmosphere [2, 3, 4]. The transportation sector is one of many sectors targeted for emission reduction [5, 6]. In fact, it was responsible for one-fifth of the global CO2 emissions in 2018 [3]. Specifically, passenger road vehicles account for the highest percentage of overall transport emissions [3].

Although the European Union have set ambitious goals to achieve climate neutrality by 2050 [7], the transport sector has seen little reduction in CO_2 emissions in the last decades and is considered a barrier to mitigation efforts in many countries [8]. According to recent estimates, 56% of adults in 28 European countries use a private car for daily trips, with an average car occupancy rate of 1.7 persons/car [9]. Interestingly, 82% of the survey respondents considered themselves to live in a well-served or relatively well-served public transport location [9].

1.1.1. The problem with private car use

Air pollution is the "largest single environmental health risk", according to the World Health Organization [10]. In 2016, 4.2 million premature deaths were attributed to exposure to ambient air pollution, 300,000 of which were children [11]. Road transport often accounts for a significant proportion of released nitrogen oxides (NO_x) in the air [12], which is associated with adverse health impact [13]. For example, in France, 64% and of NO_x was accounted for by road vehicles [14]. Although a threshold for the annual mean limit value for NO_2 is set by Directive 2008/50/EC on ambient air quality and cleaner air for Europe, researchers report that it is not held and frequently exceeded in several European Member States [10].

On top of emissions, commuters' time spent in congested results in wasted time and money [15]. In their Urban Mobility Report, Schrank et al. [16] estimated that Americans, on average, spent 99 hours a year in congestion, which turns into $88 billion in total and costs almost $1.380 per driver [16]. The 2019 Urban Mobility in the EU report stated that congestion costs the Member state economies €110 billion annually. To put this number into perspective, that is more than 1% of the European Union's gross domestic product (GDP) [17]. Cities have tried different approaches to try to solve this problem, namely using traffic management systems [15], lane expansion [18] and introducing congestion pricing [19].

1.1.2. Hard and soft measures to influencing car use behaviour

Policymakers have taken various measures to restrict the use of private cars to address these issues. These measures can be classified as hard or soft depending on their degree of coercion.

Hard measures are coercive, such as car-free zones [20], road pricing [21] and parking restrictions [22]. Hard measures can be effective in mitigating the severity of some of these problems. For example, London, Stockholm and Milan have had great success with road pricing, reducing congestion and air pollution [23]. However, they can be circumvented by motorists. J. N. Gonzalez et al.[22] have found interesting findings on parking restrictions in their study on the effectiveness of parking restrictions in Milan. The results show that high-income respondents are 75% more likely to use a private garage for parking and have a low probability of switching to more environmentally friendly modes of transport. Moreover, their model suggests that people who travel for work are less likely to give up their private car when driving in the zone affected by parking restrictions [22]. Furthermore, hard measures are not easy to implement because they are politically unenforceable or may meet public resistance. For this reason, soft measures are more attractive to policymakers[24].

Soft measures do not involve coercion but try to convince people to use alternative modes of transport by changing their perception and motivation [25]. Commonly used measures include marketing alternative transport modes such as public transport and cycling, and transport awareness campaigns [25]. Several studies have reported on soft measures' effectiveness in reducing private car use. For example, promoting bike-sharing schemes in Hangzhou, China, was associated with decreased car use [26]. Over the past decade, soft measures have been introduced in countries such as the UK [27], Australia [28], Japan [29] and Germany [30], all of which reported a decrease in car use and an increase in other modes of transport.

Despite the promising results, these studies have been criticised for using a quasi-experimental design with a single treatment group and a before-after test that does not take into account several risks to the internal validity of the causal conclusions [24]. Meta-analyses have been used to assess the validity of soft measures as they provide better methodological quality and are not narratively synthesized. A systematic review and meta-analysis by A. Semenescu et al.[31] examined 30 years of soft measures that were implemented in various countries, such as Japan, Germany, England and Sweden [31]. The results highlighted a 7% decrease in the modal split share of car use for 41 measures. However, significant variability in the effectiveness of the measures was found, which was explained by the psychological variable targeted by the measures. Measures targeting social, cultural and moral norms significantly reduced the share of car use in the modal split [31].

Although soft measures might have positive impacts, they lack theoretical grounding and standardised reporting, leading to inconclusive results about the degree of their effectiveness [31]. As knowledge about transport mode choice is still incomplete, alternative approaches are needed to take action to reduce car use. Exploring the psychological factors affecting travel mode choice is still an understudied area that can address travel behaviour and give new perspectives.

1.1.3. Psychology of transport mode choice

As mentioned above, it is essential to adequately explain and predict how individuals will respond to policies to increase the effectiveness of travel demand management strategies to reduce car usage. To understand the reasons behind travel mode choice, researchers have examined the relationship between travel behaviour, socio-demographics, and land use [32, 33, 34]. Others have focused on how attitudes and perceptions affect travel behaviour [35, 36]. However, there is still a lack of understanding of the underlying cognitive processes that drive a person's behaviour, particularly travel behaviour [37].

Over the last twenty years, psychological studies on travel mode choice have primarily been driven by the Theory of Planned Behaviour (T.P.B) [38]. T.P.B assumes that an individual's attitude towards a behaviour, subjective norms, and perceived behavioural control influence their intentions to engage in the behaviour [38]. Attitudes refer to a person's positive or negative evaluation of a particular behaviour, as they reflect that person's perception of it as "good" or "bad". Subjective norms, on the other hand, are the influence of social and cultural factors on a person's behaviour. They manifest as perceived social pressures to exhibit or not to exhibit a particular behaviour, with the expected response from significant others such as family and friends. Finally, perceived behavioural control refers to the belief that the behaviour can or cannot be performed. Perceived obstacles or facilitators play a role in the decision of whether or not to perform a behaviour. The TPB assumes that these three factors play a key role in forming the intention to perform a behaviour, and the higher the intention, the more likely the behaviour will be performed [38]. Using this theory, travel behaviour has been dominated by attitude models [39, 40, 41].

Although values are essential in decision-making [42], they are often not given much importance in travel behaviour research [37]. Values are defined as enduring beliefs and motivational constructs that refer to desirable goals; they are a subjective evaluation of abstract ideas tied to emotions [43]. Values at the core of an individual's belief system reflect the basic traits of adaptation. Hence, they are the basis on which attitudes and behaviours are later constructed [42]. Accordingly, attitudes and behaviours are subject to change or evolution during one's life, but values consist of enduring beliefs that go beyond circumstances, objects, and issues [42].

Values' centrality and transcendental nature made them attractive to many sciences related to human behaviour [44, 45]. For instance, Homer and Kahle [46] proposed a value-attitude-behaviour hierarchy model that has successfully examined the impact of values on behaviour [46]. The model implies that values influence an individual's behaviour through attitudes. They theorized that a cognitive hierarchy exists, in which a causal sequence of values (more abstract cognitions) to attitudes (mid-range cognitions) to behaviours can be depicted [46].

1.1.4. Statistical models and transport mode choice

Various statistical methods were used to understand and predict mode choice, including structural equation modelling [47, 48], logit models [49, 50, 51], integrated choice and latent variable [37], and machine learning [52].

Structural equation modelling (SEM) is a statistical technique that allows researchers to explore complex causal relationships between variables [53]. SEM can be used to test a theoretical model and estimate the strength and direction of causal relationships between variables and test the goodness of fit of the theoretical model [53].

Logit models are widely used in research on transport mode choice [54]. They are a regression model that predicts the probability of choosing a particular mode of transport based on the independent variables [54]. Logit models are simple and intuitive to interpret, and their coefficients can be used to estimate the effects of each variable on mode choice [49, 50, 51].

Integrated choice and latent variable (ICLV) models are an extension of logit models[54]. They can handle both observed and unobserved variables. ICLV models can capture the heterogeneity of decision-making processes between individuals and estimate latent variables that cannot be directly observed [54].

Machine learning (ML) algorithms are becoming increasingly popular in transport choice research [52]. ML algorithms, such as Random Forest and Support Vector Machines can handle complex data structures, capture non-linear relationships between variables and make accurate predictions [55]. These algorithms are more flexible than logit models and provide the ability to detect new patterns and relationships in the data that traditional statistical models may not capture [56].

In summary, each method has its own strengths and limitations and the choice of the appropriate method depends on the research question and the available data.

1.2. Research question

Moving away from private car use through hard and soft measures has not brought the success that policymakers had hoped for. Perhaps the complexity of mobility patterns goes beyond the analysis of daily trips and requires new approaches. Therefore, exploring and understanding the link between values and behaviour could be a valuable gap to fill. The main objective of this thesis is therefore to explore the impact of values and attitudes on transport mode choice. Is the conjunction with socio-demographic characteristics necessary or can values and attitudes explain transport mode choice. The research questions of this master thesis are formulated as follows:

- To what extent are values associated to mode choice?
- To what extent are attitudes assciated to mode choice?
- Can values alone or attitudes alone help to predict mode choice, or is this required in conjunction with socio-demographic characteristics?
- Which modelling techniques are best suited to relate values and attitudes to mode choice?

1.3. Thesis structure

After the introduction, chapter 2 presents the theoretical framework on values and attitudes, their association and their impact on transport choice. An introduction to the main statistical and machine learning models used to analyse and predict transport mode choice is also included. The chapter ends with the identified literature gap and the proposed conceptual framework.

chapter 3 presents the methodology used to answer the research questions. All steps are documented and described. Then, in chapter 4, the results of the different models developed and the evaluation metrics are presented. The results are then discussed in chapter 5, outlining the advantages and limitations of the two approaches. Finally, chapter 6 summarises the results, answers the research questions and provides recommendations based on the results as well as suggestions for future work.

2. Background and state of research

This literature review examines theories that conceptualise values and attitudes and their relationship to behaviour. First, the main theories of values are examined. Second, the role of attitudes in predicting and explaining behaviour is examined using the Theory of Reasoned Action (T.R.A) / Theory of Planned Behaviour (T.P.B). Third, the relationship between values, attitudes and behaviour is examined using the values-attitude-behaviour hierarchy. Fourth, the relationship between values, attitudes and mode choice is examined in the literature. Finally, the different approaches to modelling transport mode choice are reviewed, and the research gap is summarised.

2.1. Theoretical framework: Values and attitudes

2.1.1. The Concept of values

Rokeach Value Theory

Milton Rokeach [57] conceptualised human values in his book "The Nature of Human Values". He defined them as " an enduring conviction that a particular mode of behaviour or end-state of existence is personally or socially preferable to an opposite or reverse mode of behaviour or end-state of existence." He further distinguished between two categories of values: terminal values and instrumental values. The first focuses on desirable end states of existence, such as inner harmony. The second is more about behaviours such as self-control and honesty [57]. In Table 2.1, the 18 terminal values and instrumental values proposed by Rokeach are listed.

Table 2.1.: Terminal and instrumental values.

Terminal values	Instrumental values
True Friendship	Cheerfulness
Mature Love	Ambition
Self-Respect	Love
Happiness	Cleanliness
Inner Harmony	Self-Control
Equality	Capability
Freedom	Courage
Pleasure	Courage
Social Recognition	Honesty
Wisdom	Imagination
Salvation	Independence
Family Security	Intellect
National Security	Broad-Mindedness
A Sense of Accomplishment	Logic
A World of Beauty	Obedience
A World at Peace	Helpfulness
A Comfortable Life	Responsibility
An Exciting Life	Forgiveness

Rokeach Value Survey

Rokeach has developed a survey to assess the importance of different values within a person's belief system. Respondents are given the task of rating each value from the terminal and instrumental values in order of importance to themselves. In this way, each respondent is given two value systems of importance [58]. Despite the simplified and limited value options, Rokeach's [57] survey has been widely used as a tool to study the relationship between values and behaviour[58, 59].

Schwartz Theory of Basic Values

Based on Rokeach's work, several concepts for the construct of values emerged [60, 61]. Nevertheless, there was no consensus on the basic values, the content and structure of the relationships between these values, and agreed upon empirical methods to quantify them [43]. To fill this gap, Schwartz proposed the theory of basic human values [43]. The theory offers a conception of values that have six main characteristics [43]:

1. Values are beliefs that are fundamentally related to affect. Once activated, they are linked to feelings [43].

2. Values represent desired objectives that motivate individuals to take action. [43]

3. Values are not limited to specific actions and have a broader significance. [43]
4. Values shape the evaluation and selection of actions, people, policies and events. [43]
5. Values are ranked according to their level of importance, forming a system of priorities that define an individual. [43]
6. The relative importance of multiple values determines an individual's actions and attitudes, as they often have implications for more than one value. [43]

Table 2.2.: The ten basic values proposed by Schwartz [43].

Value	Defining goal
Self-Direction	independent thought and action-choosing, creating, exploring.
Stimulation	excitement, novelty, and challenge in life.
Hedonism	pleasure or sensuous gratification for oneself.
Achievement	personal success through demonstrating competence according to social standards.
Power	social status and prestige, control or dominance over people and resources.
Security	safety, harmony, and stability of society, of relationships, and of self.
Conformity	restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms.
Tradition	respect, commitment, and acceptance of the customs and ideas one's culture or religion provides.
Benevolence	preserving and enhancing the welfare of those with whom one is in frequent personal contact
Universalism	understanding, appreciation, tolerance, and protection for the welfare of all people and for nature

In Table 2.2 above, the 10 basic values are listed and defined.

The theory also provides a dynamic relational structure between the proposed values [43]. Fundamental to the value structure is that actions taken in pursuit of a particular value may have consequences that contradict some values but are consistent with others [43]. For example, the pursuit of power may conflict with the goals of universalism. Similarly, focusing on one's own success may discourage desired behaviour from improving the well-being of others. However, the pursuit of power and achievement are usually compatible [43].

Figure 2.1 shows the circular structure that represents the conflict and the harmonious relationship between the basic values. Since tradition and conformity pursue the same comprehensive goal, they are in a single wedge [43]. This representation of basic values illustrates the opposition between competing values [62]. One contrasting dimension is the values of "openness to change" and " conservation". In this case, there is a conflict between values that emphasise freedom of thought and action and readiness for change (self-direction, stimulation) and values that emphasise obedience and resistance to change (security, conformity, tradition) [62]. The second dimension is the contrast between "self-enhancement" and "self-transcendence". The conflict here is between values that emphasise the well-being and benefit of others (universalism, benevolence) and those that emphasise the pursuit of self-interest and relative success and control over others (power, achievement). Hedonism has the unique characteristic that it carries elements of both openness to change and self-enhancement [62].

The circular representation of values helps to view them as a motivational continuum [43]. Values that are close together in the circle have similar underlying motivations, while values that are far away have conflicting motivations [43].

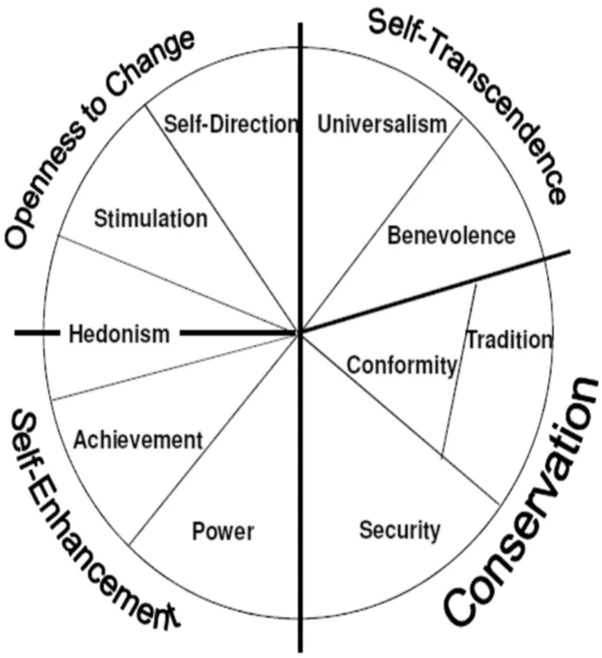

Figure 2.1.: Theoretical model of relations among ten motivational types of value [63]

Schwartz's Value Survey

The Schwartz's Value Survey (SVS) consists of a series of 57 Likert-type items that ask respondents to rate the importance of the above values. Example questions would be "A world of beauty, balance and harmony","A world in which people are responsible and dependable". Respondents would have the option to choose on a scale from "opposite to my values" to "supreme importance". The SVS measures the importance that individuals attach to different values but does not assess how people see themselves in relation to these values [64]. To combat these limitations, an alternative to the SVS is proposed. The Portrait Values Questionnaire (PVQ) is also based on Schwartz's theory, but instead focuses on the individual's self-concept and how they see themselves in relation to these values [65]. Respondents rate how much of the statement they see in themselves. An example question is " He believes that people should do what they are told. He thinks people should follow the rules at all times, even when no one is watching." The respondent has a range from "very much like me" to "not like me at all" [65].

Schwartz and colleagues [66] not only theorised but also empirically demonstrated the existence of the ten basic individual values [66]. This theory of values is perhaps the most widely accepted and disseminated. Its universal dimension and cross-cultural relevance have made it the standard theory [67].

In summary, the presented theories of values presented have proposed a list of values and their structure, and have presented their relationship to each other. However, this is insufficient to fully understand the impact of values on behaviour, especially on mode choice. The most influential research in this area has looked at attitudes and how they can explain and influence behaviour.

2.1.2. The Concept of attitudes

Ajzed [68] defined an attitude as "a disposition to respond favourably or unfavourably to an object, institution or event" [68]. They refer to a person's overall positive or negative feelings, beliefs, and behavioural intentions toward a particular object or situation [69]. Attitudes can be deduced from three different types of responses, all of which can be verbal or nonverbal [69]:

1. Cognitive response refers to an individual's knowledge and perceptions of the object of attitude.
2. Affective response refers to the emotional response towards the object of the attitude.
3. Conative response refers to the intention to behave in a certain manner towards the object of the attitude. [69]

Attitudes, which are less abstract than values, have been the focus of interest for many researchers, and it has been hypothesised that they actually influence behaviour. It is therefore important both to understand the relationship between attitudes and behaviour and to find a link between them and values.

Theory of Reasoned Action (T.R.A) and Theory of Planned Behavior (T.P.B)

Developed by Icek Ajzen and Martin Fishbein in the 1970s, The Theory of Reasoned Action (T.R.A) is a social psychological theory that depicts the relationship between people's attitudes, subjective norms, and behaviours [68]. Icek Ajzen and Martin Fishbein theorized that a person's intention to engage in a behaviour is the most important predictor of that behaviour [70]. Two factors determine the individual's intention: their attitude toward the behaviour and their perception of the normative influence to engage in the behaviour [70]. Subjective norms represent the influence of social and cultural factors on a person's behaviour. They manifest as perceived social pressures to exhibit or not to exhibit a particular behaviour, with the expected response from significant others such as family and friends [38].

Icek Ajzen built on the T.R.A and extended it to The Theory of Planned Behavior (T.P.B) [38]. The difference between the two theories is that the T.P.B introduces a new construct of perceived behavioural control [38], in other words, the belief that the individual has the actual capacity to perform the behaviour. Figure 2.2 presents the framework of the two theories and the link between the different factors.

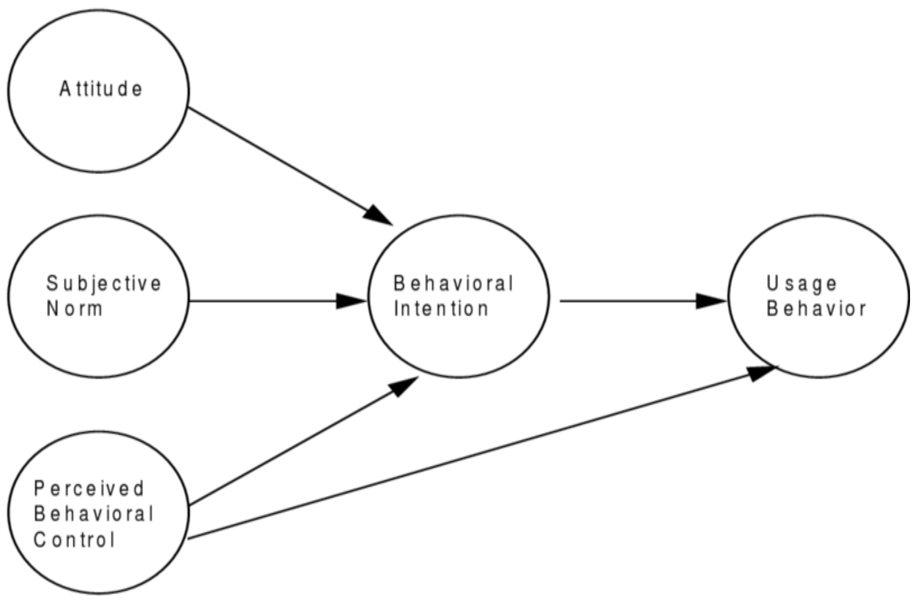

Figure 2.2.: Schematic Representation of T.R.A and its extension T.P.B [71]

To sum up, attitudes can be used to understand behaviour. In contrast to values, attitudes have a more tangible impact on behaviour. Some studies have addressed the lack of consideration for values in T.P.B. and proposed an extension to the theory [37]. However, causal relationships between attitudes and behaviour remain unexplored in both T.R.A and T.P.B and will be explored in the next section.

2.2. Connecting values, attitudes and behaviour

Homer and Kahle [46] noticed the lack of attention paid to values and tested their relationship to attitudes and behaviour [46]. Using a survey of 831 food shoppers in supermarkets and natural food stores, they hypothesised that values influence attitudes towards nutrition, which in turn influence shopping behaviour [46]. They were able to empirically demonstrate that values were more strongly associated with attitudes than with shopping behaviour. But at the same time, attitudes towards nutrition significantly influenced shopping behaviour for natural foods [46].

Based on these findings, they proposed a value-attitude-behaviour hierarchy as a framework to examine the effects of values on behaviour [46]. The hierarchy suggests that values are the most abstract and stable level of influence on behaviour, followed by attitudes and finally behaviour. According to the hierarchy, values are the underlying beliefs that guide people's attitudes and behaviours. Therefore, values are considered the basis for attitudes and behaviours, and changes in values can lead to changes in attitudes and behaviours over time [46].

In testing the value-attitude-behaviour hierarchy, T. L. Milfont et al. [72] its cross-cultural applicability was examined on data from Brazil, New Zealand and South Africa. This study was primarily concerned with the mediating role in the relationship between value and behaviour [72]. The main findings are that the hierarchy between values, attitudes and behaviour shows significant correlations between values, attitudes and behaviour in all three countries studied. However, the specific values and attitudes most strongly associated with behaviour varied somewhat across cultures. The authors conclude that the hierarchy provides a valuable framework for understanding the relationship between values, attitudes and behaviour in different cultures. However, it is important to consider cultural differences in the values and attitudes most important for behaviour in different contexts [72].

2.3. Factors impacting transport mode choice

2.3.1. Socio-demographic attributes

Understanding and predicting travel mode choice is extremely important to make appropriate recommendations to policymakers. J. Ko et al. [73] examined factors related to mode choice for commuters in Seoul and concluded that characteristics such as income, occupation and gender significantly influence mode choice [73]. Looking at household income, A. Papagiannakis et al. [74] analysed the change in transport mode choice in Greek households after the economic crisis. Their results showed that the lowest income groups experienced a more significant decrease in private car use [74].

Regarding gender and age, L. Böcker et al. [75] in their case study in Rotterdam, the Netherlands, showed the difference between elderly men and women, highlighting in particular that elderly women rely more on walking, cycling and public transport. In contrast, elderly men are frequent car users [75].

2.3.2. Attitudes

Recently, several studies have examined the role of attitudes, perceived behavioural control and subjective norm in determining their influence on transport mode choice. Donald et al. [40] used the T.P.B framework on data from 827 participants who travelled to work either by car or public transport. They found that car use was determined by intention and habit. In contrast, public transport use was only influenced by intention [40].

Similarly, E. Heinen et al. [41] collected data from over 700 adults in the Netherlands through an online survey to investigate the role of attitudes towards the choice of transport for commuting to work. The result of the study shows that attitudes towards distance and effort are an important factor in cycling behaviour. Moreover, these attitudes were more important for longer trips, while attitudes towards safety were more important for shorter trips [41].

Vredin Johansson et al. [76] investigated the effect of environmental preferences, safety, comfort, convenience and flexibility on an individual's choice of transport mode. They used a sample of Swedish commuters and found that individuals with positive flexibility and comfort attitudes towards public transportation are more likely to use sustainable transport modes such as walking, cycling and public transport. Moreover, personality traits like openness to experience and neuroticism were also linked to mode choice but did not have as strong of an effect as attitudes [76].

L. Li and Y. Zhang [77] investigated the factors that determine the intention to use carsharing in China. Using a survey of 1165 participants, they applied an extended theory of planned behaviour. The attitudes used in the survey are convenience, economy, comfort, flexibility and environmental benefits. The author's results showed that intention to use carsharing is directly influenced by attitude, subjective norm and perceived behavioural control, rather than environmental concern [77].

W. Chen et al. [78] extended T.P.B., considered public transport's impact in a mixed pro-environment travel model. They concluded that attitude has the largest effect on intention to use environmentally friendly transport, while the quality of public transport has a mediating effect on intention. They also found that the perceived quality of public transport, social and economic characteristics play a role in travel mode choice [78].

Devika, R. et al. [48] applied the T.P.B to investigate the effect of attitudes, subjective norm, perceived behavioural control and intention on mode choice. Analyzing the results of a questionnaire completed by 400 commuters, they identified attitude as the most significant factor affecting the intention to use public transit. They used the following attitudes towards public transit: enjoyment, satisfaction, and easiness. As for attitudes towards private vehicle, they used: need, safety, convenience and pestige[48].

W. Li and M. Kamargianni [79] adopted an ICLV model to assess the influence of three attitudes towards motorized and non-motorized mode choices. The used attitudes are: "willingness to be a green traveler", "satisfaction with cycling environment", and "advocacy of car-sharing services". The results show a more positive attitude towards “willing to be a green traveler” can increase the likelihood of opting for bike sharing and walking to work. Interestingly, the positive correlation between the choice of bike sharing with “willingness to be a green traveler” significantly outweigh “satisfaction with cycling environment” [79].

In trying to understand the implications of morality in T.P.B., X. Hu et al. [80] extended the theory by introducing environmental concern and moral obligation. Their results showed that attitudes perceived behavioural control, environmental concern and moral obligation correlate with young people’s intention to travel low-carbon [80].

The causal relationship between attitudes and behaviour remains unexplored in T.P.B and is explored in the next section.

2.3.3. Values

Although the body of work on values and mode choice is not extensive, some studies have examined this relationship. Linking values with attitudes using the value-attitude-behaviour theory, Paulssen et al. [37] adopted the IVCL model to explore the potential causal relationship between values, attitudes and travel mode choice. Using data from 519 German commuters, who either used private car or public transport, the results showed that power, hedonism and security values affect flexibility, comfort, convenience, and ownership attitudes. This impacts, in return, the mode choice [37].

Similarly, R. Arroyo. et al [81] used the value-attitude-behaviour theory to examine the relationship between values, attitudes and different modes of transport. Data from 404 respondents were used to build structural equation models that estimate transport mode choice. The results showed that stimulation positively influenced attitudes towards cycling, while security had a negative influence. Power was found to be negatively associated with walking. The results of the model also show that a positive attitude towards a particular mode of transport influences the use of that mode [81].

On predicting the use of a new railway line in Sweden, A. Nordlund and K. Westin. [82] conducted a questionnaire that was randomly sent to residents living along the new railway line. Their results showed that values have an impact on environmental awareness. They also found a direct correlation between holding the openness to change values vs conservation values in relation to intention to use the railway line [82].

Hunecke [83] performed a hierarchical regression to analyse how practical an attitude-based targeting approach could be in predicting the ecological impact of mobility behaviour compared to using socio-demographic data. Adding psychological factors, i.e. values and attitudes, increases the explained variance of the model by 7% for public transport users and 25% for bicycle users. The results also showed a negative association between openness-to-change and public transport rejecters [82].

J. Garcıa et al. [84] examined the relationship between values, attitudes and travel intentions concerning cycling and walking and compared them with the actual behaviour. They found that openness to change and self-transcendence values are correlated with cycling and walking [84].

R. Wal et al.[85] conducted a study among staff and students at a British university to investigate the relationship between personal values and car use. The results showed that a reduction in car use intention was associated with individuals with higher levels of altruism [85].

2.4. Statistical models and travel mode choice

2.4.1. Structural Equation Modelling

Structural Equation Modelling (SEM) is a flexible statistical modelling technique that can handle multiple types of variables. SEM is particularly suited when exploring the indirect effect of specific variables and examining the causality relationship between them. This technique has been used to model travel behaviour since 1980 [86]. Usually, when dealing with SEM, two types of variables are specified:

1. Latent variables: non-directly observable concepts
2. Observed variables: directly observable variables

SEM is a unique method to verify complex phenomena because it can formalize relations between cause-and-effect variables [87]. This is done through measurement (also called confirmatory factor analysis) and structural models. The first is used to relate observed to latent variables, and the latter to relate latent to latent variables [87].

Confirmatory factor analysis (CFA)

The first step of a CFA is to present the hypothesis to be tested in the form of a diagram and equation, then to determine the model statistically, and finally to evaluate the statistical assumptions used to support the model [88]. The measurement model depicts how latent variables are measured through observed indicators [87]. The measurement model can be compared to a simple linear regression, where the main predictor is latent. A simple regression equation is defined as [53]:

$$y = b_0 + b_1 x + \epsilon \tag{2.1}$$

where b_0 is the intercept, b_1 is the coefficient, ϵ is the error and x is an observed predictor. The same logic is applied to the measurement model for one item [53]:

$$y_1 = \tau_1 + \lambda_1 \eta + \epsilon_1 \tag{2.2}$$

where τ is the intercept of the item, λ is the loading of the item and ϵ_1 is the residual for the item.
The measurement model for one latent variable can be written as follows [53]:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} + \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} (\eta_1) + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix} \tag{2.3}$$

Each term of the measurement model is defined as [53]:

- τ: the item's intercepts
- λ: the loadings, which represent the correlation of the item with the factor
- η: the latent predictor of the items, i.e. the factor
- ϵ: the residuals of the measurement model

Path diagram

A path diagram facilitates the understanding of the model to be tested. Figure 2.3 shows the standard symbols used in a path diagram and their meanings [53]:

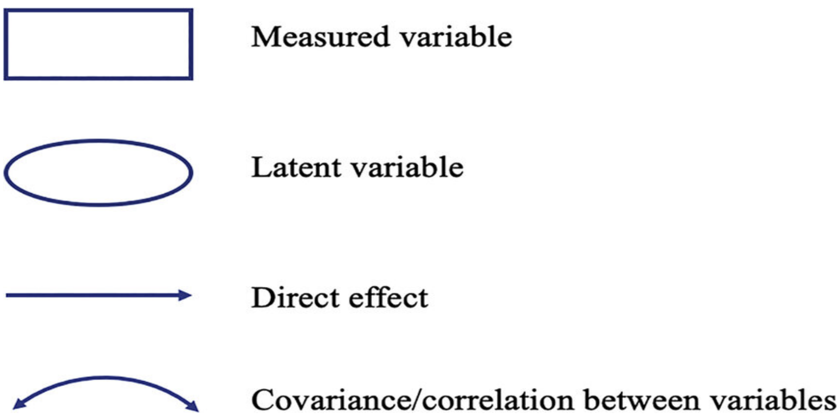

Figure 2.3.: Standard symbols used in a path diagram and their meanings [89]

Structural model model

The causal relationships between the latent variables are defined in the structural model [87]. Now Figure 2.4 shows an example of a SEM with two latent variables and the relationship between them:

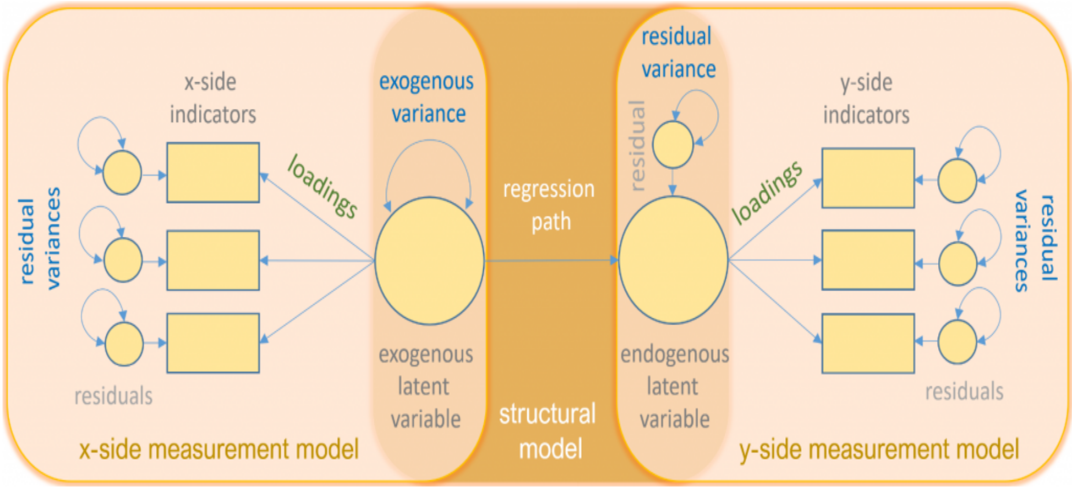

Figure 2.4.: Example of a path diagram [53]

In a structural model, the latent variables represented by a circle can be exogenous or endogenous. Other variables have no influence on the exogenous latent variables. They are independent variables that have a causal influence on the observed variables in the model. On the other hand, endogenous latent variables are influenced by other variables in the model. They are dependent variables caused by the relationships between the other variables in the model, including other latent variables and observed variables [53].

The main difference between exogenous and endogenous latent variables is that exogenous variables have a causal influence on the model but are not influenced by the other variables in the model. In contrast, endogenous variables are influenced by the other variables in the model. The measured variable, represented by a square, is the observed variable found in the data set. It can also be referred to as the manifest variable. The loading parameter represents the strength of the relationship between a latent variable and an observed variable. The residual variance is the unexplained variance of an observed variable not explained by the relationships specified in the model [53].

Fit Indices of SEM Model

To assess the goodness-of-fit of a structural equation model, the following metrics need to be checked:

- Chi-Square Test is a test that compares the observed data with the expected data based on the model. A non-significant chi-square value ($p > 0.05$) indicates that the model fits the data well. The chi-square value is sensitive to sample size, so a large sample size may result in a significant chi-square value even if the model fits well [53].
- Comparative Fit Index (CFI) is a fit index that compares the fit of the model with the fit of a null model where all variables are uncorrelated. It ranges from 0 to 1, with values

closer to 1 indicating a better fit. A CFI of 0.95 or higher is considered to indicate a good fit [53].

- The Tucker-Lewis Index (TLI) compares the fit of the hypothesized model to that of a null model.It ranges from 0 to 1, with values closer to 1 indicating a better fit. A TLI of 0.95 or higher are preferred [53]
- Root Mean Square Error of Approximation (RMSEA) measures the discrepancy between the model and the data, with lower values indicating a better fit. A value of 0.06 or lower is generally considered to indicate a good fit [53].

2.4.2. Logit Models

Traditionally, the logit model has been used in travel behaviour research. It is a mathematical framework used in econometrics that draws upon theories from mathematical psychology [54]; namely, Thurstone's law [90] of comparative judgement.

Thurstone [90] assumed that an individual's perception of an alternative depends on unknown parameters and alternative-specific characteristics so that an individual n perceives an alternative j as:

$$U_{jn} = f(\theta_{jn}) + \varepsilon_{jn} \tag{2.4}$$

where the function $f(\theta_{jn})$ depends on unknown parameters θ_{jn} and characteristics specific to alternative j. The error term ε_{jn} is normally distributed, independent, and varies across individuals and alternatives.
When an individual n is set to choose between two alternatives j and j', they will prefer the alternative j over j' prime if $U_{jn} - U_{j'n} > 0$ and the probability of the alternative j is equal to:

$$P(j \mid \theta_{jn}, \sigma_{j,j'}) = \Phi\left(\frac{f(\theta_{jn}) - f(\theta_{j'n})}{\sigma_{j,j'}}\right) \tag{2.5}$$

where $(f(\theta_{jn}) - f(\theta_{j',n})$ is the mean and $\sigma_{j,j'}$ is the standard deviation. This form has been generalized to stochastic utility maximization over multiple alternatives using the choice axiom proposed by Luce [91] and was defined as random utility maximization by Marschak [92].

Luce's [91] choice axiom postulates that when an individual n chooses an alternative j in a set of alternatives C_n, the ratio of choice probabilities for two alternatives j and j' is constant for every choice set C_n:

$$\frac{P(j \mid C_n)}{P(j' \mid C_n)} = \frac{P_{j,j'}(j)}{P_{j,j'}(j')} \tag{2.6}$$

This axiom is called Independence from Irrelevant Alternatives (IIA); it suggests that the probability of selecting a choice from a given set of alternatives is independent of all other alternatives, and when it holds, a positive strict utility V_{jn} is associated with each alternative in this manner:

$$P(j \mid C_n) = \frac{V_{jn}}{\sum_{j' \in C_n} V_{j'n}} \tag{2.7}$$

McFadden [93] introduced the application of this model in travel behaviour analysis by proposing a parametric exponential function for the strict utility:

$$V_{jn} = exp\left(X_{jn}^{\intercal}\beta\right) \tag{2.8}$$

where X_{jn} are the attributes of j of a specific individual n and β are to be estimated parameters. This form is called the multinomial logit (MNL) model and was used to predict the mode choice of work commuting trips and the generation, destination and mode of shopping trips in urban areas [94].

The MNL provides a clear mathematical framework and makes parameter estimation easy, so it has been extensively used to model discrete travel behaviour choices [49, 50, 51]. The MNL model is, however, challenged for its IIA assumption as it lies at the core of every rational choice theory. Many studies have demonstrated that individuals making travel mode choices often disobey this assumption, which leads to biased predictions and inconsistent parameter estimation [95]. To overcome these limitations, researchers have proposed extensions to the MNL model, namely multinomial probit models [96, 97], nested logit models [49, 98, 99] and mixed logits models [100, 37, 101].

However, due to these models' high dimensionality and complexity, parameter estimation is complicated and requires sophisticated techniques that can be computationally intensive [102]. In summary, MNL models and its extensions offer a sound statistical framework to predict and analyze travel behaviour. However, they rely on assumptions that can lead to erroneous model estimation if not fulfilled.

Integrated choice and latent variable models

Several variables, observed ones such as socio-demographics, the built environment and travel cost, can influence travel mode choice. Recently, studies have attributed latent behaviour variables to account for unobserved effects. These effects are called latent variables, such as values and attitudes. Latent variables can be indirectly measured through psychometric indicators. The integrated choice and latent variable (ICLV) model proposed by Ben-Akiva et al. [103] can capture these effects by extending the random utility modelling presented in subsection 2.4.2 [54]. The ICLV model has two components: a choice model and a latent variable model. Figure 2.5 showcases the structure of an ICLV model.

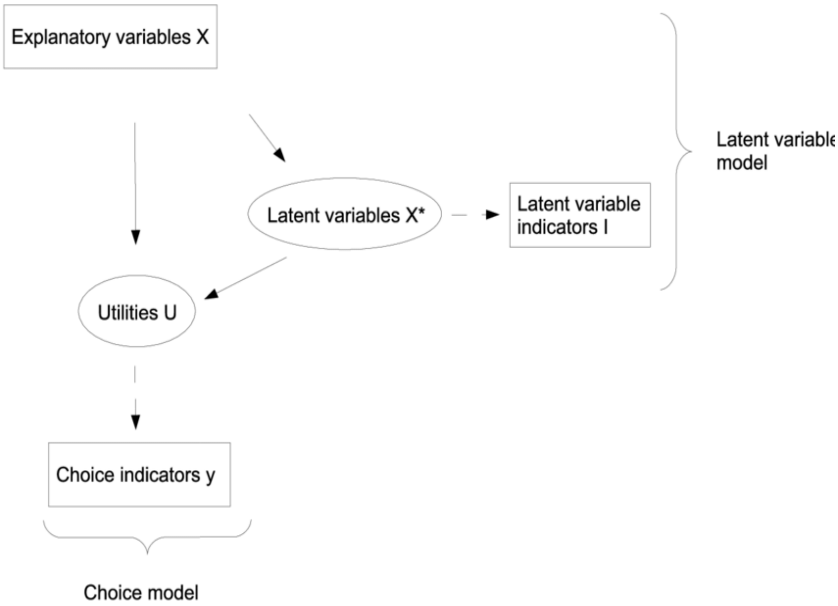

Figure 2.5.: Integrated choice and latent variable model [104]

The utility function of the alternative j is as follows [54]:

$$U_{jn} = V_{jn}\left(X_{jn}, \phi_n; \beta\right) \tag{2.9}$$

where the systematic utility V_{jn} is dependent on the observed attributes X_{jn}, the latent variables ϕ_n and parameters β.

A structural and measurement components are needed to model the latent variables. Suppose that there are K latent variables indexed from k= 1,....,K and L measureemnt indicators indexed frpm l=1,....,L. The structural component is as follows [54]:

$$\phi_{kn} = W_{kn}\left(Z_{kn}, \gamma_k\right) + \eta_{kn} \tag{2.10}$$

where W_{kn} consists of the function of the individual-specific characteristics Z_{kn} and parameters γ_k, and η_{kn} is a random disturbance.

The measurement component is as follows:

$$m_{ln} = \delta_l + \phi_n^{\intercal}\lambda_l + \xi_{ln} \tag{2.11}$$

where δ_l is the constant, λ_l is the vector of loadings and ξ_{ln} is an error term.

The joint probability of the observed choice and the indicators are given by [54]:

$$P(y, m \mid X, \beta, \gamma, \delta, \lambda) = \int_\phi P(y \mid X, \phi, \beta) P(m \mid \delta, \phi, \lambda) f(\phi \mid Z, \gamma) d\phi \tag{2.12}$$

Integrated Choice and Latent Variable (ICLV) models have explored the complex relationship between latent variables and transport mode choice. They can provide valuable insights for transport research by capturing the latent factors that determine transport mode choice[76, 37, 79].

2.4.3. Machine learning to predict mode choice

To overcome the limitations of logit models, machine learning (ML) has been of interest in transport research for years [56]. ML refers to methods or algorithms that enable computers to mechanise data-driven model programming and create models that use methodically recognise patterns in statistically significant data [52]. There are different types of ML algorithms; the most used ones are [52]:

- Supervised learning: algorithms generate a function by using input data to predict output data.
- Unsupervised learning: no output data is provided; ML models the input data looking for clustering.
- Reinforcement learning: ML algorithms learn when interacting in an environment. The learning process is done by receiving feedback on the accuracy of the response. [52]

Mode choice modelling can be viewed as a classification problem [56], often a use case for supervised learning [105]. The ML algorithm must learn a function that can classify a vector into one of several classes by looking at the input and output examples of the function [105]. Several studies have reported the effective usage of machine learning classifiers in predicting travel mode choice.

For example,L. Cheng et al. [55] used Random Forest (RF), Support Vector Machine (SVM) and MNL to predict walking and cycling in Nanjing, China. They used travel data from 2013 and used household attributes, socio-demographic data, travel time and trip purpose as variables. Their results show the better performance of RF with an overall accuracy of 85% [55].

X. Zhao et al. [56] also use different classifiers to predict travel mode choice based on socio-economic and demographic attributes, parking cost, waiting time, access to rideshare and public transport. The best-performing model was RF with an accuracy of 85.6% [56].

In his review of machine learning algorithms used to model mode choice [52], J. D. Pineda-Jaramill concluded that RF is the best algorithm for modelling mode choice [52]. XGBoost is also a powerful classifier that has been used to model mode choice. F. Wang and C. L. Ross. [106] tested the algorithm on survey data from the Delaware Valley, Pennsylvania. They used travel time and socio-demographic attributes as variables to pedict car, walking, cycling or transit use. XGBoost proved robust with an average testing error of 15.5% [106].

Machine learning classifiers

Popular classifiers used to predict travel mode choice are briefly presented in this section. There will be no detail on how they function, as it is out of the scope of this thesis.

K-Nearest Neighbors(KNN) is a non-parametric classification algorithm that assigns a class label to a new data point based on the class labels of its k-nearest neighbours in the training data [107].

Random Forest (RF) works by constructing many decision trees at the time of training and outputting the class that corresponds to the classes' mode or the individual trees' mean prediction [107].

Support Vector Machines (SVM) work by finding the hyperplane that maximally separates the different classes. SVM classifiers have many parameters that can be tuned to optimise performance, including the kernel function, the regularisation parameter and the penalty parameter [107].

Extreme Gradient Boosting (XGBoost) is a tree-based ensemble method built on iteratively growing low-depth decision trees. XGBoost combines several weak predictive models (often decision trees) to create a strong model that makes very accurate predictions [106].

Performence evaluation

For classification problems, different metrics can be used, including:

- Accuracy is the proportion of correct predictions out of the total number of predictions [108].
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.13}$$
- Precision is the proportion of true positives (correct predictions) over the total number of positive predictions [108].
$$Precision = \frac{TP}{TP + FP} \tag{2.14}$$
- Recall is the proportion of true positives (correct predictions) over the total number of actual positives [108].
$$Recall = \frac{TP}{TP + FN} \tag{2.15}$$
- F1 score is the harmonic mean of precision and recall. It provides a balanced measure of the model's accuracy [108].
$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{2.16}$$

- The Receiver Operating Characteristic (ROC) curve is a graphical representation of the performance of a binary classifier. It shows the trade-off between the true and false positive rates at different thresholds [108].

2.5. Literature gap

In the context of transport choice, several studies have shown that attitudes towards transport modes such as driving, cycling, walking and public transport can influence transport choices. However, the role of personal values in transport mode choice has not yet been studied in depth. These findings can be seen in Table 2.3. Thus, there is a gap in the literature on the relationship between mode choice and values. In addition, machine learning has gained attention in predicting mode choice and has proven to be a useful tool [109]. Still, no literature has been found that uses values or attitudes as explanatory variables.

Ref	Latent Variables	Key Findings	Model and Evaluation
[47]	Attitudes towards service preference and public transport: comfort, efficiency, safety, punctuality and economy	Except for service preference safety, all other attitudes significantly influenced the travel mode choice	SEM NFI= 0.829 CFI= 0.863 RMSEA = 0.088
[48]	Attitudes towards public transit: enjoyment, satisfaction, easiness. Attitudes towards private vehicle: need, safety, convenience, prestige	Attitudes towards private vehicles were found to have a more significant influence on intention to use public transit than attitudes towards public transit-favouring attitude.	SEM NFI= 0.988 CFI= 0.998 RMSEA = 0.007
[81]	Attitudes towards cycling, Attitudes towards walking, Achievement, Stimulation,Security, Universalism, power	stimulation positively influenced attitudes towards cycling, while security had a negative influence. Power negatively associated with walking	SEM CFI= 0.867 TLI= 0.855 RMSEA = 0.043
[110]	Attitudes towards public transport: convenience, personal safety, modal comfort, service environment, and waiting feelings	Convenience and service environment have a bigger influence on the choice of public transport, while modal comfort has a minor influence on it.	SEM-DCM goodness of fit = 0.446

[76]	Attitudes towards public transport: environmental preferences, safety, comfort, convenience and flexibility	Individuals with positive flexibility and comfort attitudes towards public transport are more likely to use sustainable transport modes such as walking, cycling and public transport	Regression R^2_{Adj}=0.122
[83]	Values: Opennes-to-change, conservation, self-transcendence, self-enhancement, attitudes towards car: autonomy, privacy, excitement, attitudes towards bicycle: autonomy, excitement	Negative association between openness-to-change and public transport. Adding attitudes helps make better predictions. rejecters	Regression McFadden Pseudo-R^2= 0.21
[76]	Attitudes towards public transport: environmental preferences, safety, comfort, convenience and flexibility	Individuals with positive flexibility and comfort attitudes towards public transport are more likely to use sustainable transport modes such as walking, cycling and public transport	Regression R^2_{Adj}=0.122
[37]	Values: power, hedonism, security, attitudes towards transport modes: flexibility, convenience/-comfort, ownership	power, hedonism and security values affect flexibility, comfort, convenience, and ownership attitudes. This impacts, in return, the mode choice	ICLV -

Table 2.3.: Selected research from the literature review

3. Methodology

Figure 3.1 describes the methodology used in this study, explaining each step in the section. The methodology consists of three main parts: first, the data is pre-processed, then the first approach is described, which tests the causality relationship between values, attitudes and mode choice using structural equation modelling. Then the second approach is described, which involves predicting mode choice using different machine learning models.

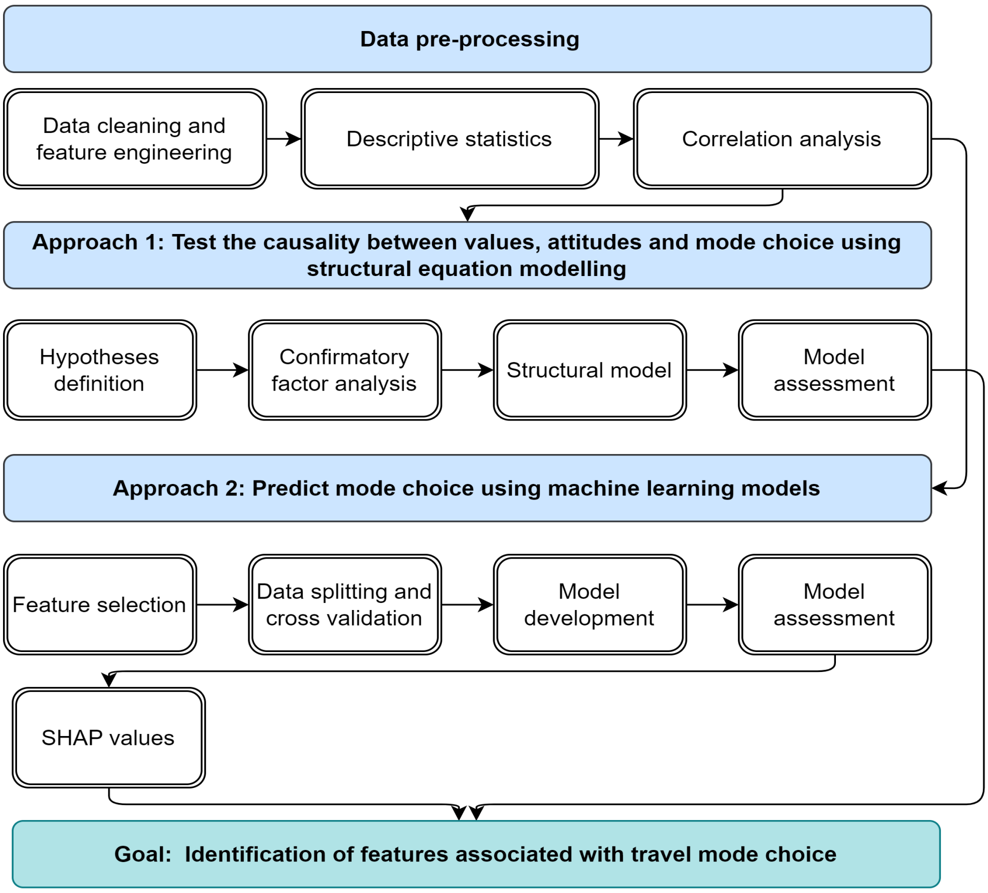

Figure 3.1.: Methodology followed in this thesis

3.1. Data pre-processing

3.1.1. The survey

Before examining the data, the survey used in this paper will be briefly described. The dataset comes from a survey conducted in Munich between October and December 2019. The survey was part of Michael Stoeckle's work for his Master's thesis at the Technical University of Munich [111]. The survey was conducted online and on-site and 451 responses were collected. The survey consists of five main parts:

Awareness and use of bike sharing schemes

In this part, participants were asked if they knew about four bike-sharing systems currently operating in Munich: MVG Rad, DB Call a Bike, Donkey Republic and Jump Bike. Next, they are asked about the frequency of using bike-sharing systems, and in case they do not use them, they are given a list of possible reasons to choose from.

Frequency of usage of different means of transport and trip purpose

In this part, different modes of transport were given, e.g. walking, cycling, public transport and own car. The participant could choose seven options: daily or almost daily, one to three days a week, one to three days a month, less than once a month, do not use, used in the past and never used. The purpose of the trip was asked. For each mode choice, there were seven options: to work, to school/university, for shopping, for private errands, for leisure activities, to get home and for other purposes.

Attitudes towards modes of transport

This part asked about attitudes towards the different modes. Attitudes towards each mode of transport were related to convenience, relaxation, fun, health, safety, environmental friendliness and accessibility of use. Participants could choose from these options: I fully agree, I partly agree, I neither agree nor disagree, I partly disagree and I fully disagree

Psychographic questions

The 21-item version of the Portrait Values Questionnaire was used [112] to collect data on values. Each item corresponds to a basic value. The questions and their corresponding items can be found in A. The participant has the following options: Very much like me, Like me, Somewhat like me, A little like me, Not like me, Not like me at all and Don't know.

Socio-demographics

In this part, the socio-demographic characteristics of the participants are asked: age, gender, level of education, occupation, income, number of people in the household, whether they have a migration background and finally, where they live.

3.1.2. Data pre-processing

Data cleaning: In this step, the missing values in the data are identified, and the corresponding row is deleted. Rows containing answers with "I do not know" in the attitudes and values parts were also removed.

Feature engineering: A user definition was established for each mode of transport to evaluate the participants' choice of transport mode. In this thesis, a user is defined as someone who uses a particular mode of transport daily, almost daily, one to three times a week. Following this logic, each participant was given three additional attributes: Public transport user, a bike user and private car user. These are the target features. A user has a 1 as code and a non-user has a zero as code.

In the socio-demographic part, many categories in the question about occupation and education led to outliers. They were removed from the data set. The features age and income were agregated into four categories instead of the eight and ten categories provided by the survey.

The next step was to code the responses from the attitudes and values parts. A scale of 1 to 5 was used for attitudes, from 1 for "I fully disagree" to 5 for "I fully agree". For the PQV questions, the scale ranged from 1 to 6, starting with 1 for "Not like at all" to 6 for "Very much like me". Finally, each value was calculated according to the values given by Schwartz by taking the mean of the corresponding items (see Appendix A).

Descriptive statistics: In this step, summary statistics such as mean, mode and standard deviation are calculated for continuous variables in the data set. This provides information about the distribution of the data and identifies potential problems.

Correlation analysis: Pearson correlation coefficients between values, attitudes and mode choice were calculated. The aim is to identify statistically significant relationships between the variables.

3.2. Approach 1: Testing the causality between values, attitudes and mode choice using structural equation modelling

3.2.1. Hypotheses

Before using SEM, causal relationships between variables need to be hypothesized. The aim is to test the hypothesis and decide whether to confirm or reject it. The two-step approach is the most commonly used method in developing SEM. The first step is to develop an acceptable measurement model through confirmatory factor analysis (CFA) and then test the hypothesis using a structural model.

The hypotheses to be tested are:

- H1. Personal values are related to attitudes.

  -H1.a Personal value Conservation is associated positively with attitudes towards driving and negatively towards public transport and cycling

  -H1.b Personal value Openness to change is associated negatively with attitudes towards driving and positively towards public transport and cycling

- H2. Attitudes are related to transport mode choice.

  -H2.a Attitudes towards driving are positively associated with driving and negatively associated with public transport use and cycling

  -H2.b Attitudes towards public transport are positively associated with public transport use and cycling and negatively associated with driving

  -H2.c Attitudes towards cycling are positively associated with cycling and negatively associated with driving

3.2.2. Confirmatory factor analysis

Before using SEM, causal relationships between variables need to be hypothesized. The aim is to test the hypothesis and decide whether to confirm or reject it. The two-step approach is the most commonly used method in developing SEM. The first step is to develop an acceptable measurement model through confirmatory factor analysis (CFA) and then test the hypothesis using a structural model [53].

In CFA, a model is specified in which the observed variables are assumed to be indicators of unobserved latent factors. First, correlation matrices are created between the indicators in question. Indicators that do not show strong correlations are removed. Second, the latent variables are contracted. In this case, the latent variables are values, attitudes towards private cars, attitudes towards cycling and attitudes towards public transport. Each variable has a set of measured values, i.e. manifest variables. The measurement model is built by writing each latent variable as a function of the measurements [53]. For example, attitude towards driving can be written as follows:

$$Att\_Driving =\sim ATT\_CarDriving\_convenient + ATT\_CarDriving\_safe + ATT\_CarDriving\_fun \tag{3.1}$$

where $Att\_Driving$ is the latent variable and $ATT\_CarDriving\_convenient$, $ATT\_CarDriving\_safe$, $ATT\_CarDriving\_fun$ are its corresponding measurements from the data set.
The software used is R with the Lavaan package. The method "cfa" from the package is used to build the model. The factor loadings for each measurement are then checked for relevance. The measurement model is then evaluated using the Comparative Fit Index, the Tucker-Lewis Index and the Root Mean Square Error of Approximation.

Table 3.1 and Table 3.2 introduce the different measurements and summarise the different steps taken to construct the latent variables.

Variables		Coding of Input Value
Latent Variable	Observed Variable	
Attitudes towards driving	Car driving is convenient Car driving is safe Car driving is environmentally friendly Car driving is relaxing Car driving is fun	1 → I fully disagree 2 → I partially disagree 3 → I neither agree nor disagree 4 → I partially agree 5 → I fully agree
Attitudes towards cycling	Cycling is convenient Cycling is safe Cycling is environmentally friendly Cycling is relaxing Cycling is fun	
Attitudes towards public transport	Public transport is convenient Public transport is safe Public transport is environmentally friendly Public transport is relaxing Public transport is fun	

Table 3.1.: Attitudes and their measurements and coding

Variables		Coding of Input Value	
Latent Variable	Observed Variable	Step 1	Step 2
Conservation	Conformity Tradition Security	Each item is given a code 1→ Not like me at all 2 → Not like me 3 → A little like me 4 → Somewhat like me 5 → Like me 6 → Very much like me	mean(item7,item16) mean(item9,item20) mean(item5,item14)
Self-enhancement	Power Achievement		mean(item2,item17) mean(item9,item20)
Opennes to change	Hedonism Stimulation Self-direction		mean(item10,item21) mean(item9,item20) mean(item1,item11)
Self-transcendence	Universalism Benevolennce		mean(item3,item8) mean(item12,item18)

Table 3.2.: Values and their measurements and coding

3.2.3. Structural model

Once the CFA model is validated, a structural model can be tested. The structural model specifies the hypothesized relationships between latent and/or observed variables. A structural model is usually presented graphically in SEM as a path diagram. The path diagram shows the hypothesized causal relationships between the latent factors and the observed variables. In this case, the arrows represent the causal relationships between values and attitudes and between attitudes and transport choices. The path diagram is shown in Figure 3.2 below:

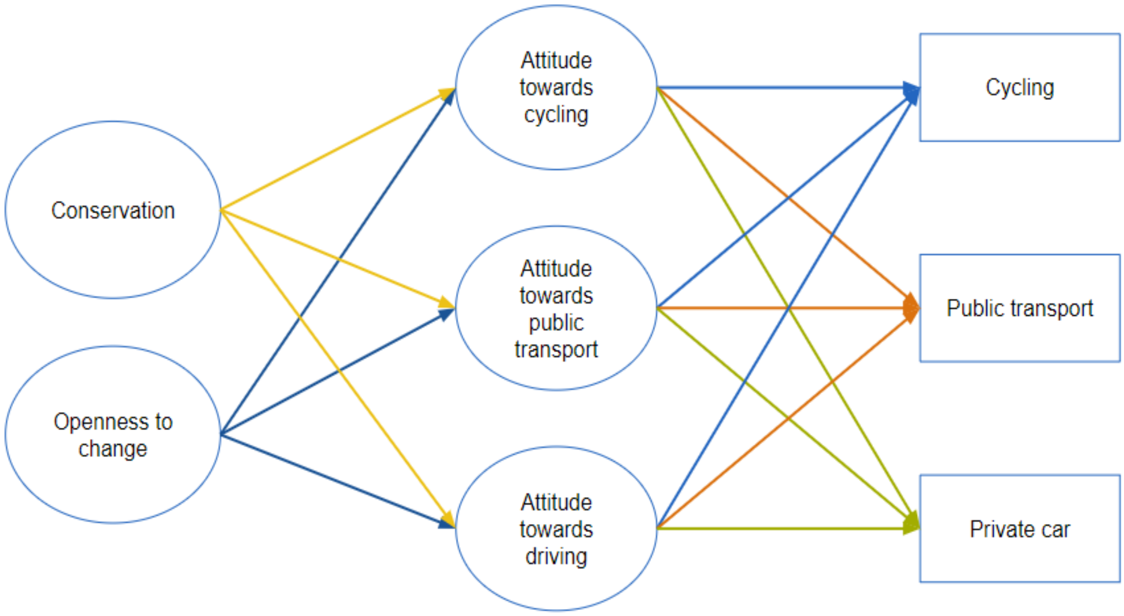

Figure 3.2.: Path diagram

3.2.4. Model estimation and results

The fit of the structural model is assessed using the Comparative Fit Index, the Tucker-Lewis Index and the Root Mean Square Error of Approximation. The proposed hypothesis is then either confirmed or rejected.

3.3. Approach 2: Predicting mode choice using machine learning models

3.3.1. Feature selection

Since the data is already cleaned from previous steps, the next step is to select the features that will be used to train the model. Table 3.3 summarizes the steps to create each value. Socio-demographics attributes are categorical data such as gender, age and income. They were encoded using the get_dummies() method from the Pandas library. The method is used to convert categorical variables into dummy variables. It creates a new DataFrame with binary columns for each category in the original data. For example, the categorical

Variables		Coding of Input Value		
Output variable	Intermediate variable	Step 1	Step 2	Step 3
Conservation	Conformity Tradition Security	Each item is given a code	C=mean(item7,item16) T=mean(item9,item20) S=mean(item5,item14)	mean(C,T,S)
Self-enhancement	Power Achievement	1→ Not like me at all 2 → Not like me 3 → A little like me	P=mean(item2,item17) A= mean(item9,item20)	mean(P,A)
Opennes to change	Hedonism Stimulation Self-direction	4 → Somewhat like me 5 → Like me 6 → Very much like me	H=mean(item10,item21) St=mean(item9,item20) Sd= mean(item1,item11)	mean(H,St,Sd)
Self-transcendence	Universalism Benevolennce		U=mean(item3,item8) B=mean(item12,item18)	mean(U,B)

Table 3.3.: Values and their measurements and coding

column "Gender" with the values 'Male' and 'Female' would be converted into two new columns: "Gender_Male" and "Gender_Female". The values in these columns would be 1 or 0, indicating whether the original column contained this category for a particular row. Table 3.4 displays the categorical data used and the corresponding categories.

3.3.2. Data splitting and cross-validation

After cleaning and feature selection, the data is split into train and test sets. The train data set fits the model, while the test data set evaluates the performance of the model. Cross-validation is used to assess the model's performance across multiple data folding [108]. Cross-validation involves splitting the data into multiple subsets and training the model on each subset while testing it on the remaining data. This allows a more accurate assessment of the model's performance on new data [108]. Additionally, oversampling techniques are recommended to balance the training set when dealing with an unbalanced data set [113]. In the data set, the class distributions are highly imbalanced (see chapter 4). The data was split into 75% train and 25% test sets. The library used is Scikit-learn and the method is train_test_split. The synthetic minority oversampling technique (SMOTE) was applied to the train set to address the class imbalance.

3.3.3. Model development

From the literature review, four commonly used machine learning classifiers were identified and used for this thesis, which are K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM) and eXtreme Gradient Boosting (XGBoost). The four models are

Categorical data	Categories
Gender	Male Female
Age	18-29 years 30-39 years 40-49 years above 50 years
Employment status	Full-time employed Part-time employed Student (university)
Education	Bachelor´s degree Master´s degree Technical school degree
Household income	0€-900€ 900€-2000€ 2000€-4000€ Above 4000€

Table 3.4.: Categorical data

developed with four different input data:

- Socio-demographics
- Attitudes
- Values
- Socio-demographics, attitudes and values

In addition, hyperparameter tuning is used to optimise the performance of the model by adjusting the parameters of the algorithm. The GridSreachCV method was used with a cross-validation of 5.

3.3.4. Model assessment metrics

Finally, the performance of the model is evaluated using various metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). These metrics measure the performance of the model and can be used to compare different models.

3.3.5. SHAP values

SHAP (SHapley Additive exPlanations) values are a technique for explaining the output of machine learning models. They provide a way to calculate the importance of each feature to a particular prediction by calculating each feature's contribution to the prediction [95]. The SHAP values make the interpretability of machine learning algorithms possible by helping to understand which features are most important for a particular prediction and how each feature contributes to the overall prediction of the model [95]. By understanding the importance of each feature, it is possible to determine which factors most influence mode choice. The shap library in Python was used. The TreeExplainer and shap_values and summary_plot methods were used to determine the importance of the features.

4. Results

4.1. Descriptive statistics

Table 4.1 shows the distribution of the dataset by socio-demographic variables. The dataset consists of 250 respondents, of whom 60% identify as male and 40% as female. The majority of the respondents are aged between 18 and 29 years (56%), followed by 30 to 39 years (24%), 40 to 49 years (10%) and above 50 years (10%). In terms of employment status, 46% of the respondents were employed full-time, 42% were employed part-time and 12% were students. Regarding education, 49% of the respondents had a Master's degree, followed by 32% with a Bachelor's degree and 19% with a technical school degree. Finally, looking at household income, the largest group of respondents reported a monthly income between 2000€-4000€ (32%), followed by above 4000€ (28%), 0€-900€ (22%) and 900€-2000€ (18%).

		Dataset distribution	
		Count	%
Gender	Male	148	60
	Female	102	40
Age	18-29 years	141	56
	30-39 years	59	24
	40-49 years	26	10
	above 50 years	24	10
Employment status	Full-time employed	114	46
	Part-time employed	106	42
	Student (university)	30	12
Education	Bachelor´s degree	81	32
	Master´s degree	123	49
	Technical school degree	46	19
Household income	0€-900€	55	22
	900€-2000€	46	18
	2000€-4000€	79	32
	Above 4000€	70	28

Table 4.1.: Distribution of the dataset

Table 4.2 presents the descriptive statistics of the 10 basic values: conformity, tradition, benevolence, universalism, self-direction, stimulation, hedonism, achievement, power and security. The mean values ranged from 3.10 for power to 5.03 for benevolence.

The standard deviation values ranged from 0.68 for universalism to 1.21 for conformity. In addition, the median value for almost all values is 4. The values of conservation (conformity, tradition, security) have on average lower values than the values of self-transcendence (benevolence and universalism).

	mean	std	25%	50%	75%
Conformity	3.50	1.21	2.50	3.50	4.50
Tradition	3.52	1.03	3.00	3.50	4.38
Benevolence	5.03	0.72	4.50	5.00	5.50
Universalism	5.01	0.68	4.67	5.00	5.33
SelfDirection	4.64	0.80	4.00	4.50	5.00
Stimulation	3.80	1.08	3.00	4.00	4.50
Hedonism	4.19	1.02	3.50	4.00	5.00
Achievement	3.82	1.10	3.00	4.00	4.50
Power	3.10	0.97	2.50	3.00	4.00
Security	3.64	1.11	3.00	3.50	4.50

Table 4.2.: Descriptive statistics of values

Table 4.3 shows the descriptive statistics of the different attitudes towards public transport, cycling and driving. The results show that public transport is perceived as the most convenient and safest, with mean scores of 4.49 and 4.32, respectively. Cycling is perceived as the healthiest and most environmentally friendly mode of transport, with mean scores of 4.82 and 4.90, respectively. Driving a car is perceived to be the least healthy and environmentally friendly, with mean scores of 1.82 and 1.48, respectively. Cycling is also perceived to be the most convenient mode of transport, with a mean score of 4.73, while driving a car is perceived to be the least relaxing mode of transport, with a mean score of 2.62.

	mean	std	25%	50%	75%
ATT_PT_convenient	4.49	0.79	4.00	5.00	5.00
ATT_PT_relaxing	3.03	1.19	2.00	3.00	4.00
ATT_PT_fun	2.62	1.01	2.00	2.00	3.00
ATT_PT_healthy	2.85	1.06	2.00	3.00	3.00
ATT_PT_safe	4.32	0.83	4.00	4.00	5.00
ATT_PT_EnvFriendly	4.41	0.70	4.00	5.00	5.00
ATT_Cycling_convenient	4.73	0.54	5.00	5.00	5.00
ATT_Cycling_relaxing	4.19	0.88	4.00	4.00	5.00
ATT_Cycling_fun	4.49	0.74	4.00	5.00	5.00
ATT_Cycling_healthy	4.82	0.45	5.00	5.00	5.00
ATT_Cycling_safe	3.20	1.08	2.00	3.00	4.00
ATT_Cycling_EnvFriendly	4.90	0.39	5.00	5.00	5.00
ATT_CarDriving_convenient	3.83	1.19	3.00	4.00	5.00
ATT_CarDriving_relaxing	2.62	1.20	2.00	2.00	4.00
ATT_CarDriving_fun	3.21	1.21	2.00	3.00	4.00
ATT_CarDriving_healthy	1.82	0.82	1.00	2.00	2.00
ATT_CarDriving_safe	3.27	1.06	3.00	3.00	4.00
ATT_CarDriving_EnvFriendly	1.48	0.71	1.00	1.00	2.00

Table 4.3.: Descriptive statistics of attitudes

4.1.1. Effects of socio-demographics on mode choice

Table 4.4 presents the dataset regarding the number of public transport users and non-users in each group. Statistical significance is indicated by p-values for the chi-square. The results show significant differences in public transport use between age groups, occupational groups, education levels and household income. There were no significant gender differences in the use of public transport.

		Public transport		p-value for the chi-square
		User	Non user	
Gender	Male	104	44	-
	Female	77	25	
Age	18-29 years	120	21	***
	30-39 years	37	22	
	40-49 years	12	14	
	above 50 years	12	12	
Occupation	Full-time employed	68	46	***
	Part-time employed	17	13	
	Student (university)	96	10	
Education	Bachelor´s degree	67	14	***
	Master´s degree	78	45	
	Technical school degree	36	10	
Household income	0€-900€	49	6	***
	2000€-4000€	55	24	
	900€-2000€	40	6	
	Above 4000€	37	33	

$^{-}$p>0.05, *p<0.05, **p<0.01, ***p<0.001

Table 4.4.: Descriptive analysis of public transport user and non-user

Table 4.4 presents the dataset in terms of the number of users and non-users of private car each group, and statistical significance is indicated by p-values for the chi-square. The results indicate that there are significant differences in private car use between age groups, occupational groups and household income. There were no significant gender and education differences in the use of private cars.

		Private car		p-value for the chi-square
		User	Non user	
Gender	Male	35	113	-
	Female	20	82	
Age	18-29 years	23	118	***
	30-39 years	11	48	
	40-49 years	14	12	
	above 50 years	7	17	
Occupation	Full-time employed	34	80	**
	Part-time employed	7	23	
	Student (university)	14	92	
Education	Bachelor´s degree	12	69	-
	Master´s degree	31	92	
	Technical school degree	12	34	
Household income	0€-900€	10	45	***
	2000€-4000€	16	63	
	900€-2000€	3	43	
	Above 4000€	26	44	

$^{-}$p>0.05, *p<0.05, **p<0.01, ***p<0.001

Table 4.5.: Descriptive analysis of private car users and non-users

Table 4.6 presents the dataset regarding the number of users and non-users of private bikes in each group. Statistical significance is indicated with p-values for the chi-square. The results indicate no statistical significant differences in the use of private bicycles between the different occupational groups and household incomes except for age.

		Private bike		p-value for the chi-square
		User	Non user	
Gender	Male	117	31	-
	Female	76	26	
Age	18-29 years	100	41	*
	30-39 years	51	8	
	40-49 years	21	5	
	above 50 years	21	3	
Occupation	Full-time employed	90	24	-
	Part-time employed	27	3	
	Student (university)	76	30	
Education	Bachelor´s degree	60	21	-
	Master´s degree	111	26	
	Technical school degree	35	12	
Household income	0€-900€	37	18	-
	2000€-4000€	60	19	
	900€-2000€	36	10	
	Above 4000€	60	10	

$^{-}$p>0.05, *p<0.05, **p<0.01, ***p<0.001

Table 4.6.: Descriptive analysis of private bike users and non-users

4.1.2. Correlation between values, attitudes and mode choice

Table 4.7 presents the correlation results between attitudes towards the three modes of transportation (public transport, cycling, and car driving) and five basic values: conformity, tradition, benevolence, universalism, and self-direction. The statically significant correlations are coloured in green.

First, there is a significant positive correlation between attitudes towards public transport and the moral values of universalism and safety. Second, attitudes towards cycling show a negative correlation with convenience and a positive correlation with universalism and stimulation. Third, attitudes towards driving are strongly correlated with tradition and power. Negative correlations are found between all attitudes towards driving and universalism.

	Conformity	Tradition	Benevolence	Universalism	SelfDirection
ATT_PT_convenient	-0.0	-0.02	0.11	0.22***	-0.03
ATT_PT_relaxing	0.06	0.08	0.07	0.1	0.02
ATT_PT_fun	0.01	-0.0	0.07	0.07	-0.02
ATT_PT_healthy	0.1	0.06	0.04	0.04	-0.11
ATT_PT_safe	-0.03	-0.11	0.07	0.15*	0.07
ATT_PT_EnvFriendly	0.07	-0.1	0.02	0.08	-0.09
ATT_PT_accessible	0.04	0.06	0.06	0.16*	0.08
ATT_Cycling_convenient	-0.09	-0.07	0.01	0.18**	0.02
ATT_Cycling_relaxing	-0.14*	-0.01	0.0	0.12	0.03
ATT_Cycling_fun	-0.06	-0.03	0.08	0.14*	0.03
ATT_Cycling_healthy	-0.02	-0.01	0.08	0.19**	0.01
ATT_Cycling_safe	-0.11	0.03	-0.02	0.01	0.08
ATT_Cycling_EnvFriendly	0.02	-0.08	0.06	0.02	-0.06
ATT_CarDriving_convenient	0.1	0.27***	0.02	-0.13*	0.03
ATT_CarDriving_relaxing	0.07	0.22***	0.05	-0.13*	0.05
ATT_CarDriving_fun	0.08	0.13*	0.12	-0.18**	0.02
ATT_CarDriving_healthy	0.16*	0.2**	-0.03	-0.23***	-0.05
ATT_CarDriving_safe	0.09	0.11	-0.05	-0.16**	-0.01
ATT_CarDriving_EnvFriendly	0.09	0.22***	-0.05	-0.28***	0.0

*p<0.05, **p<0.01, ***p<0.001

Table 4.7.: Correlation between attitudes and values

Table 4.8 shows the correlations between the attitudes and the rest of the ten values: stimulation, hedonism, achievement, power, and security. The statically significant correlations are coloured in green.

The results show that attitudes towards cycling are generally positively correlated with stimulation, hedonism and achievement, while negatively correlated with power and security. Specifically, attitudes towards cycling positively correlate with convenience, relaxation, fun and health, while negatively correlated with security. Attitudes towards driving are positively correlated with power, security, and convenience while negatively correlated with relaxation. Attitudes towards driving are positively correlated with hedonism, achievement, power and security.

	Stimulation	Hedonism	Achievement	Power	Security
ATT_PT_convenient	-0.04	-0.01	0.04	-0.05	-0.01
ATT_PT_relaxing	0.02	-0.03	0.02	0.01	0.06
ATT_PT_fun	0.08	0.01	0.12	-0.01	0.0
ATT_PT_healthy	0.09	0.01	-0.04	-0.03	0.13*
ATT_PT_safe	0.08	0.07	-0.2**	-0.11	-0.09
ATT_PT_EnvFriendly	-0.01	0.04	-0.0	-0.02	-0.13*
ATT_Cycling_convenient	0.18**	0.07	-0.12	-0.14*	-0.17**
ATT_Cycling_relaxing	0.2**	0.02	0.0	-0.07	-0.09
ATT_Cycling_fun	0.24***	0.15*	0.06	-0.07	-0.11
ATT_Cycling_healthy	0.18**	0.18**	0.02	-0.06	-0.02
ATT_Cycling_safe	0.08	-0.11	0.05	-0.02	-0.05
ATT_Cycling_EnvFriendly	0.06	0.17**	0.12*	0.02	-0.07
ATT_Cycling_accessible	0.06	0.08	0.09	0.07	0.02
ATT_CarDriving_convenient	-0.07	0.07	0.06	0.14*	0.16*
ATT_CarDriving_relaxing	0.0	0.13*	0.15*	0.24***	0.18**
ATT_CarDriving_fun	0.09	0.27***	0.18**	0.26***	0.13*
ATT_CarDriving_healthy	-0.08	0.08	0.21***	0.26***	0.21***
ATT_CarDriving_safe	-0.02	0.04	0.06	0.23***	0.06
ATT_CarDriving_EnvFriendly	-0.04	0.03	0.13*	0.14*	0.11

*p<0.05, **p<0.01, ***p<0.001

Table 4.8.: Correlation between attitudes and values

Table 4.9 shows the correlation between the three transport modes and the corresponding attitudes towards them. The statically significant correlations are coloured in green.

There are negative correlations between private car use and all attitudes towards public transport except safety. The only statically significant correlation between public transport use and attitudes is for convenience and relaxation on public transport. There is also a significant positive correlation between convenience of cycling and use of private bike and a negative correlation with use of private care. Private car use correlates positively with convenience, relaxation, fun and health when driving.

	Private bike user	Private car user	Public transport user
ATT_PT_convenient	-0.06	-0.17**	0.26***
ATT_PT_relaxing	0.04	-0.09	0.17**
ATT_PT_fun	0.07	-0.02	0.09
ATT_PT_healthy	-0.03	-0.01	0.08
ATT_PT_safe	0.15*	0.02	-0.03
ATT_PT_EnvFriendly	0.06	-0.04	0.12
ATT_PT_accessible	0.05	-0.11	0.05
ATT_Cycling_convenient	0.42***	-0.17**	-0.01
ATT_Cycling_relaxing	0.25***	-0.05	-0.06
ATT_Cycling_fun	0.35***	-0.08	-0.04
ATT_Cycling_healthy	0.2**	-0.06	0.01
ATT_Cycling_safe	0.02	0.02	-0.12
ATT_Cycling_EnvFriendly	0.18**	-0.01	0.03
ATT_CarDriving_convenient	-0.12	0.22***	0.05
ATT_CarDriving_relaxing	-0.17**	0.19**	0.03
ATT_CarDriving_fun	-0.16*	0.15*	0.05
ATT_CarDriving_healthy	-0.31***	0.14*	0.06
ATT_CarDriving_safe	-0.09	0.12	0.01
ATT_CarDriving_EnvFriendly	-0.2**	0.19**	-0.04

*p<0.05, **p<0.01, ***p<0.001

Table 4.9.: Correlation between attitudes and the use of the modes of transport

Table 4.10 presents the correlation results between values and the use of the three transport modes. The statically significant correlations are coloured in green.

The values conformity, tradition and security are negatively correlated with private bike use, while the values stimulation and universalism are positively correlated with private bike use. Private car use is positively correlated with tradition and power, while it is negatively correlated with security and universalism. Public transport use is not significantly correlated with any of the measured values.

	Private bike user	Private car user	Public transport user
Conformity	-0.14*	0.06	0.02
Tradition	-0.15*	0.13*	0.12
Benevolence	-0.07	0.01	0.01
Universalism	0.2**	-0.1	0.0
SelfDirectionr	-0.01	0.0	-0.06
Stimulation	0.22***	-0.11	-0.02
Hedonism	-0.03	-0.04	0.1
Achievement	-0.04	0.02	0.03
Power	-0.13*	0.03	0.05
Security	-0.19**	0.05	-0.05

*p<0.05, **p<0.01, ***p<0.001

Table 4.10.: Correlation between values and the use of the modes of transport

4.2. Approach 1: Structural equation modelling

4.2.1. Confirmatory factor analysis

Before starting a CFA, correlation matrices are calculated between the measurements of the latent variables to be constructed. This ensures that the measurements are indicators of the same concept. A correlation of 0.30 or more is considered a medium effect size in psychology and the social sciences [53].

Table 4.11 shows the correlation between the ten basic values. The correlations between values with the same motivations are coloured with the same colour. The correlation coefficients are all higher than 0.3, which makes the indicators suitable for a CFA.

	Conformity	Tradition	Benevolence	Universalism	SelfDirection	Stimulation	Hedonism	Achievement	Power	Security
Conformity	1.0***									
Tradition	0.38***	1.0***								
Benevolence	0.11	0.13*	1.0***							
Universalism	-0.1	-0.05	0.43***	1.0***						
SelfDirection	-0.13*	-0.04	0.2**	0.17**	1.0***					
Stimulation	-0.19**	0.01	0.1	0.23***	0.33***	1.0***				
Hedonism	0.05	0.07	0.19**	0.11	0.32***	0.49***	1.0***			
Achievement	0.11	0.03	0.07	-0.1	0.14*	0.13*	0.23***	1.0***		
Power	0.27***	0.11	-0.06	-0.29***	0.12*	0.1	0.24***	0.49***	1.0***	
Security	0.5***	0.36***	0.18**	-0.08	0.02	-0.18**	0.01	0.15*	0.21**	1.0***

*p<0.05, **p<0.01, ***p<0.001

Table 4.11.: Correlation matrix between the 10 basic values

Table 4.12 shows the correlation between the different attitudes towards public transport. The attitudes toward safety and environmental friendliness do not correlate highly with the other attitudes; therefore, they are excluded from the CFA.

Attitudes	Convenience	Relaxation	Funness	Health	Safety	Env. friendliness
Convenience	1.0***					
Relaxation	0.48***	1.0***				
Funness	0.44***	0.66***	1.0***			
Health	0.22***	0.39***	0.52***	1.0***		
Safety	0.15*	0.24***	0.22***	0.24***	1.0***	
Env. friendliness	0.29***	0.19**	0.16*	0.2**	0.37***	1.0***

*p<0.05, **p<0.01, ***p<0.001

Table 4.12.: Correlation between attitudes towards public transport

Table 4.13 the table shows the correlation between the different attitudes toward driving. The attitudes toward safety and environmental friendliness do not correlate highly with the other attitudes; therefore, they are excluded from the CFA.

Attitudes	Convenience	Relaxation	Funness	Health	Safety	Env. friendliness
Convenience	1.0***					
Relaxation	0.62***	1.0***				
Funness	0.39***	0.68***	1.0***			
Health	0.31***	0.42***	0.42***	1.0***		
Safety	0.26***	0.41***	0.29***	0.33***	1.0***	
Env. friendliness	0.24***	0.33***	0.29***	0.49***	0.36***	1.0***

*p<0.05, **p<0.01, ***p<0.001

Table 4.13.: Correlation between attitudes towards car driving

Table 4.14 shows the correlation between the different attitudes towards cycling. Safety was removed in this case because of low correlation coefficients.

Attitudes	Convenience	Relaxation	Funness	Health	Safety	Env. friendliness
Convenience	1.0***					
Relaxation	0.36***	1.0***				
Funness	0.37***	0.61***	1.0***			
Health	0.29***	0.28***	0.39***	1.0***		
Safety	0.08	0.35***	0.21**	0.05	1.0***	
Env. friendliness	0.31***	0.16*	0.29***	0.54***	-0.1	1.0***

*p<0.05, **p<0.01, ***p<0.001

Table 4.14.: Correlation between attitudes towards cycling

4.2.2. Results of the CFA model

The first CFA model did not converge when using the four values. I decided to remove the two values self-transcendence and self-enhancement. This left two opposing values: of conservation and openness to change. The results of the second CFA model indicate that the latent variables of "Attitude car driving", "Attitude public transport", "Attitude cycling", "Conservation", and "Openness to change" are well-defined by their respective manifest variables. This is indicated by the standardized loadings (Std.all), which are higher than 0.4 [53]. The estimate values represent the regression coefficients for the factor loadings, while the Std.Err indicates the standard error of the estimates, and z and p indicate the statistical significance of each estimate.

The estimates in the first column are the result of the marker method. In this method, the factor loading of one of the marker variables is fixed at 1. Then the factor loadings of the remaining marker variables and the factor loadings of the remaining observed variables are estimated with respect to the fixed marker variable. This method allows the determination of the relative scale of the latent variables with respect to the fixed marker variable. Standardized estimates (Std.all) are usually preferred because they provide a standard measure for comparing the strength of the relationship between each observed variable and the latent construct. Standardized factor loadings range from -1 to 1, with higher absolute values indicating a stronger relationship between the observed variable and the latent construct [53].

For "Attitude car driving," all manifest variables are statistically significant. The highest factor loading is for the manifest variable "Driving is relaxing" with a standardised loading of 0.93, meaning that for one standard deviation in "Attitude car driving"," the variable "Car driving is relaxing" increases by 0.93 standard deviation units. "Driving is fun" and "Driving is convenient" also have high loadings, 0.73 and 0.64, respectively. "Driving is relaxing" has the lowest residual variance (13%) compared to all other manifest variables. indicating that the model explains this observed variable the best. "Car driving is healthy" has the lowest factor loading (0.48) and the highest residual variance, indicating limited variability and not a good indicator of the latent variable "Attitude car driving".

As for "Attitude public transport", the manifest variables are also statistically significant. The highest factor loading has the manifest variable "Public transport is fun" (0.85), followed by "Public transportation is relaxing" (0.78), "Public transport is healthy" (0.55), and "Public transport is convenient" (0.54). The manifest variables "Public transport is convenient" and "Public transport is healthy" have the highest covariances with 71% and 67%, respectively.

Regarding "Attitude cycling", the manifest variables are statistically significant. The highest factor loading belongs to the manifest variable "Cycling is fun" (0.83), followed by "Public transport is relaxing" (0.71), "Public transport is healthy" (0.47), and "Public transport is convenient" (0.49). "Public transport is convenient" and "Public transport is healthy" have the highest covariances, 78% and 76%, respectively.

Both values, "Conservation" and "Openness to change", have statistically significant manifest variables. "Conservation" is strongly associated with the manifest variables "Conformity" (0.69) and "Security" (0.71). "Tradition" has the lowest loading (0.53) and the highest residual variance (72%). "Openness to change" is most strongly associated with stimulation (0.80), and least associated with "Self-direction" (0.42). "Self-direction" has a residual variance of 82% and "Stimulation" of 35%.

The CFI value of 0.918 and the TLI value of 0.899 indicate a reasonable fit. The root mean square error of approximation (RMSEA) measures model fit given model complexity and sample size. The RMSEA value of 0.056 indicates a good fit of the model.

	Estimate	Std. Err.	z	p	Std.all
		Factor Loadings			
Attitude car driving					
ATT.CarDriving.convenient	1.00^{+}				0.64
ATT.CarDriving.fun	1.15	0.12	9.77	***	0.73
ATT.CarDriving.relaxing	1.45	0.14	10.25	***	0.93
ATT.CarDriving.healthy	0.51	0.07	6.80	***	0.48
Attitude public transport					
ATT.PT.convenient	1.00^{+}				0.54
ATT.PT.relaxing	2.16	0.27	8.03	***	0.78
ATT.PT.healthy	1.35	0.21	6.59	***	0.55
ATT.PT.fun	2.01	0.25	8.14	***	0.85
Attitude cycling					
ATT.Cycling.convenient	1.00^{+}				0.49
ATT.Cycling.relaxing	2.36	0.35	6.73	***	0.71
ATT.Cycling.fun	2.33	0.34	6.88	***	0.83
ATT.Cycling.healthy	0.79	0.15	5.41	***	0.47
Conservation					

Tradition	1.00+				0.53
Conformity	1.55	0.26	6.08	***	0.69
Security	1.47	0.24	6.04	***	0.71
Openness to change					
SelfDirection	1.00+				0.42
Stimulation	2.56	0.53	4.86	***	0.80
Hedonism	1.86	0.36	5.23	***	0.62
Residual Variances					
ATT.CarDriving.convenient	0.83	0.08	9.88	***	0.59
ATT.CarDriving.fun	0.68	0.08	8.71	***	0.47
ATT.CarDriving.relaxing	0.19	0.08	2.51	*	0.13
ATT.CarDriving.healthy	0.52	0.05	10.73	***	0.77
ATT.PT.convenient	0.44	0.04	10.27	***	0.71
ATT.PT.relaxing	0.55	0.08	7.12	***	0.40
ATT.PT.healthy	0.78	0.08	10.22	***	0.70
ATT.PT.fun	0.27	0.06	4.88	***	0.27
ATT.Cycling.convenient	0.22	0.02	10.31	***	0.76
ATT.Cycling.relaxing	0.38	0.05	7.92	***	0.49
ATT.Cycling.fun	0.17	0.03	4.83	***	0.31
ATT.Cycling.healthy	0.15	0.01	10.39	***	0.78
Tradition	0.76	0.08	9.36	***	0.72
Conformity	0.75	0.12	6.24	***	0.52
Security	0.60	0.10	5.77	***	0.49
SelfDirection	0.53	0.05	10.17	***	0.82
Stimulation	0.40	0.13	3.16	**	0.35
Hedonism	0.64	0.09	7.37	***	0.62
Fit Indices					
CFI	0.92				
TLI	0.90				
RMSEA	0.06				
χ^2(df)	223.61(125)			0.000	

+Fixed parameter
*p<0.05, **p<0.01, ***p<0.001

Table 4.15.: Summary of the measurement model

4.2.3. Results of the structural model

The results of the model shown in Table 4.16 aim to explain the relationship between attitudes toward different modes of transport, the underlying factors that shape those attitudes, i.e. values and the potential influence of those attitudes on mode choice.

The model provides estimates of the relationships between the latent and observed variables, as well as the relationships between the latent variables themselves.

	Estimate	Std. Err.	z	p
	Regression Slopes			
Attitude car driving				
Conservation	0.41	0.13	3.15	0.002
Openness	0.20	0.19	1.05	0.293
Attitude cycling				
Conservation	−0.07	0.05	−1.60	0.109
Openness	0.29	0.09	3.12	0.002
Attitude public transport				
Conservation	0.04	0.07	0.56	0.578
Openness	0.13	0.11	1.12	0.265
Private bike user				
attitude.bike	0.70	0.12	5.64	0.000
attitude.car	−0.07	0.03	−2.14	0.032
attitude.pt	−0.12	0.06	−2.03	0.042
Private car user				
attitude.bike	−0.10	0.10	−0.93	0.352
attitude.car	0.11	0.04	3.09	0.002
attitude.pt	−0.03	0.06	−0.51	0.609
Public transport user				
attitude.bike	−0.18	0.11	−1.63	0.104
attitude.car	0.03	0.04	0.73	0.468
attitude.pt	0.22	0.07	3.04	0.002
	Residual Covariances			
Private.bike.user w/Private.car.user	−0.00	0.01	−0.13	0.897
Private.bike.user w/pt.user	−0.02	0.01	−1.84	0.066
Private.car.user w/pt.user	−0.04	0.01	−3.54	0.000
	Latent Variances			
attitude.car	0.55	0.10	5.33	0.000
attitude.pt	0.19	0.04	4.35	0.000
attitude.bike	0.07	0.02	3.94	0.000
Conservation	0.29	0.08	3.69	0.000
Openness	0.10	0.04	2.70	0.007
	Latent Covariances			
Conservation w/Openness	−0.03	0.02	−2.09	0.036
	Fit Indices			
χ^2(df)	358.20(173)			0.000
CFI	0.86			
TLI	0.83			

RMSEA	0.07

Table 4.16.: Summary of the structural model

The results of the structural model show that conservation is only significantly positively associated with attitudes towards driving, with an estimate of 0.41 and a p-value of 0.002. Conservation is not statistically significantly associated with attitudes towards cycling and public transport. As for openness to change, the regression results show no statistically significant relationship with attitudes towards driving and using public transport. Openness to change is only significantly positively related to attitudes towards cycling, with an estimate of 0.29 and a p-value of 0.002.

Not all attitudes were significantly related to the use of each mode of transport. With the exception of cycling, where the three attitudes were significantly associated with its use, car driving and public transport use were only significantly and positively associated with attitudes towards car driving and public transport, respectively. For private bike users, attitudes towards cycling showed the strongest positive association with an estimate of 0.7. Both attitudes towards car and public transport are negatively related to cycling, with estimates of -0.07 and -0.12, respectively.

The residual covariances show a weak and non-significant relationship between the use of a private bike and the use of a private car. The same applies to the relationship between the use of a private bike and public transport. However, there is a significant negative relationship between the use of a private car and the use of public transport. Additionally, the latent variances show that attitude towards driving has the highest variance (0.55), followed by attitude towards public transport (0.19) and attitude towards cycling (0.07). In terms of values, conservation shows a latent variance of 0.29 and openness to change of 0.10. Finally, the fit indices for the model are acceptable, with a CFI of 0.86, a TLI of 0.83 and an RMSEA of 0.07.

4.3. Approach 2: Machine learning

4.3.1. Results of the models predicting private car use

Table 4.17 presents the results of the selected machine learning algorithms for predicting private car usage using socio-demographic data and attitudes as input data. The models' performances are evaluated based on their F1 score, accuracy, precision and recall. The KNN model has a low F1 score of 0.52 and an accuracy of 0.74, indicating many false positives and false negatives. The Random Forest (RF) model performs better with an F1 score of 0.65 and an accuracy of 0.76, but its precision and recall are low. The Support Vector Machine (SVM) model performs similarly to the RF model with an F1 score of 0.59 and an accuracy of 0.79. Finally, the XGBoost model performs best with an F1 score of 0.61 and an accuracy of 0.81.

In all models, performance deteriorates when attitudes are replaced as inputs. A range of 0.50 to 0.53 is observed for the F1 score. With an F1 score of 0.50 and an accuracy of 0.66, the Random Forest (RF) model performs worst in this case.

Predicting private car usage using								
Models	Socio-demographics				Attitudes			
	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall
KNN	0.52	0.74	0.28	0.15	0.53	0.66	0.25	0.25
RF	0.65	0.76	0.50	0.38	0.50	0.63	0.22	0.30
SVM	0.59	0.79	0.5	0.41	0.52	0.63	0.25	0.25
XGBoost	0.61	0.81	0.60	0.23	0.51	0.65	0.23	0.30

Table 4.17.: Car use prediction: model performances on test set

Table 4.18 presents the results of the selected machine learning algorithms for predicting private car use using values and socio-demographic data, attitudes and values as input data. Using only values as input data, the KNN model performed best with an F1 score of 0.58 and an accuracy of 0.68. The Random Forest (RF) model and XGBoost performed almost identically better, with the same F1 score of 0.65, precision of 0.26, and recall of 0.46. The Support Vector Machine (SVM) model performed the least well, with an F1 score of 0.40 and an accuracy of 0.79. Finally, the XGBoost model performs the best with an F1 score of 0.53 and an accuracy of 0.61.

For the sociodemographics, attitudes and values feature set, KNN and RF have the highest F1 score and accuracy of 0.57 and 0.73, respectively, leading to better performance when compared to the values feature set. However, the precision and recall values for both KNN and RF are still relatively low. SVM and XGBoost also perform better with this feature set but with lower performance than KNN and RF.

Predicting private car use using								
Models	Values				Socio-demographics+Attitudes+Values			
	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall
KNN	0.58	0.68	0.31	0.46	0.57	0.73	0.33	0.30
RF	0.53	0.61	0.26	0.46	0.57	0.73	0.33	0.30
SVM	0.40	0.46	0.16	0.38	0.49	0.61	0.21	0.30
XGBoost	0.53	0.53	0.26	0.46	0.56	0.71	0.30	0.30

Table 4.18.: Car use prediction: model performances on test set

Figure 4.1shows the plot of the ROC curves for all models using the different data sets. All four data sets do not yield good ROC curves.

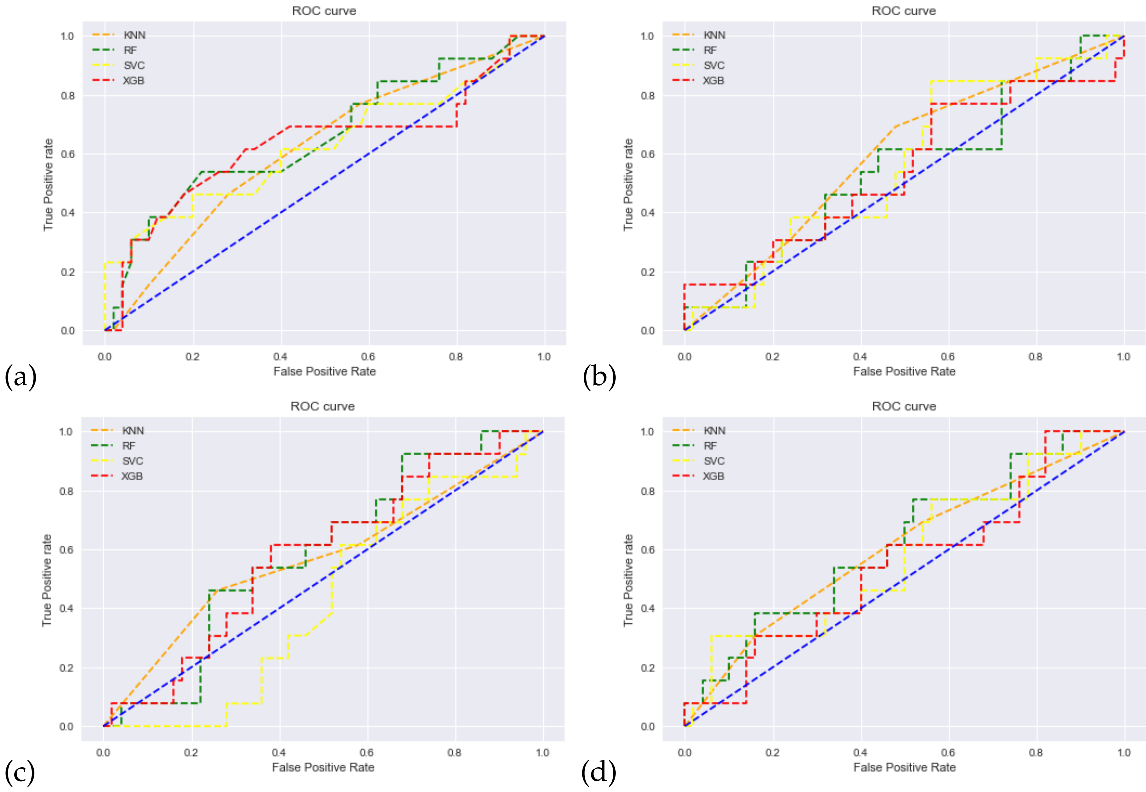

Figure 4.1.: (a) ROC curve using socio-demographic features (b) ROC curve using attitude features (c) ROC curve using value features (d) ROC curve using socio-demographic, attitude and value features

SHAP values

SHAP was used to interpret the model output using all the tree features. As depicted in Figure 4.2, the features age 40-49 years, net income 4000 and above, the attitude car driving is relaxing, and conservation are the top four features that influence the output of the RF model on the training set.

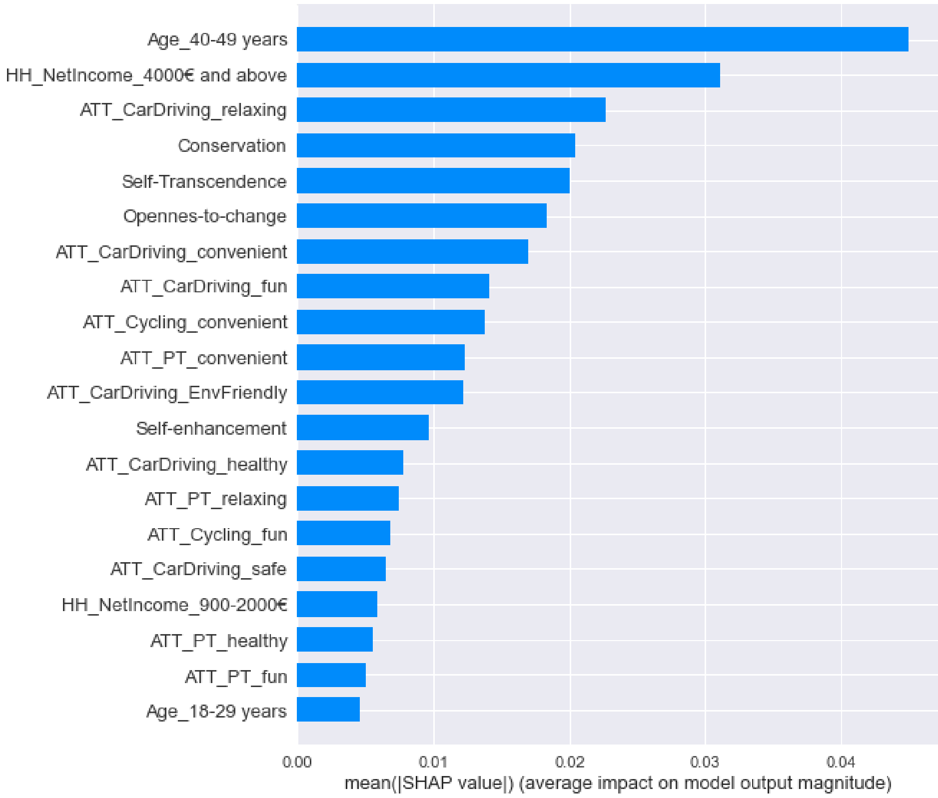

Figure 4.2.: SHAP Feature importance RF (training set)

4.3.2. Results of the models predicting private private bike use

Table 4.19 shows the performance of machine learning algorithms in predicting private bike usage using socio-demographic data and attitudes as input variables. When predicting bike usage using the socio-demographic features, the KNN model performs best with an F1 value of 0.49 and an accuracy of 0.61. However, the precision and recall scores for KNN are relatively low. When predicting bike usage using attitudes, the RF model performs the best with an F1 score of 0.72 and an accuracy of 0.81. This model's precision and recall scores are also high.

Predicting private bike use using								
Models	Socio-demographics				Attitudes			
	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall
KNN	0.49	0.61	0.77	0.71	0.61	0.69	0.84	0.75
RF	0.43	0.65	0.75	0.81	0.72	0.81	0.87	0.87
SVM	0.43	0.77	0.77	1.0	0.73	0.81	0.89	0.85
XGBoost	0.40	0.58	0.73	0.73	0.74	0.81	0.91	0.83

Table 4.19.: Private bike use prediction: model performances on test set

Table 4.22 shows the performance of machine learning algorithms in predicting private bike usage using values, then socio-demographics, attitudes and values as input data. The KNN model performs the worst of all the models, with an F1 score of 0.57 and an accuracy of 0.57. The SVM model performs the best, with an F1 score of 0.72 and an accuracy of 0.79. The XGBoost and RF models also perform well, with an F1 score between 0.70 and 0.72 and an accuracy between 0.79 and 0.81. Overall, the results suggest that including attitudes and values together with socio-demographic information as input data can improve the performance of the models in predicting private bike usage.

Predicting private bike use using								
Models	Values				Socio-demographics+Attitudes+Values			
	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall
KNN	0.47	0.61	0.76	0.73	0.57	0.57	0.81	0.79
RF	0.44	0.61	0.75	0.75	0.70	0.79	0.87	0.85
SVM	0.59	0.65	0.86	0.65	0.72	0.79	0.89	0.83
XGBoost	0.49	0.69	0.77	0.85	0.72	0.79	0.89	0.83

Table 4.20.: Private bike use prediction: model performances on test set

Figure 4.3 shows the plot of the ROC curves for all models using the different features. Using attitudes or combining socio-demographics, attitudes, and values results in the best ROC curves.

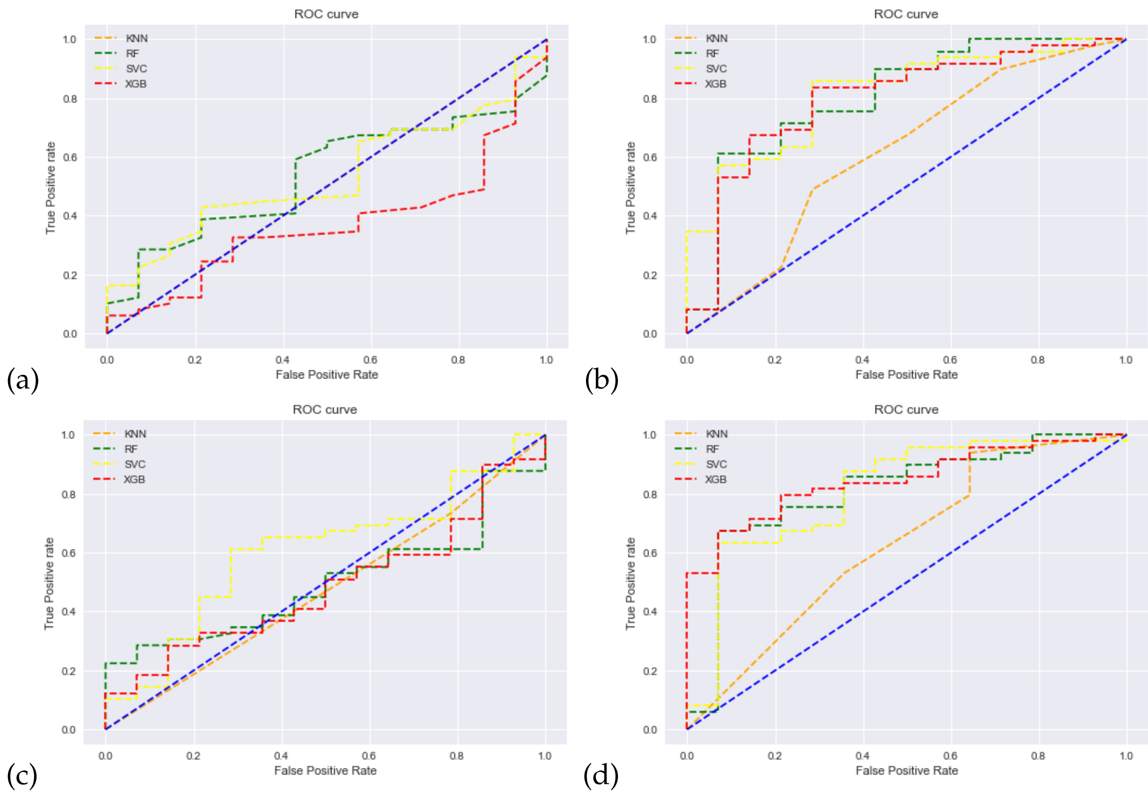

Figure 4.3.: (a) ROC curve using socio-demographic features (b) ROC curve using attitude features (c) ROC curve using value features (d) ROC curve using socio-demographic, attitude and value features

SHAP values

SHAP was used to interpret the model output using all the tree features. As depicted in Figure 4.4, the features three features that influence the output of the RF model on the training set.

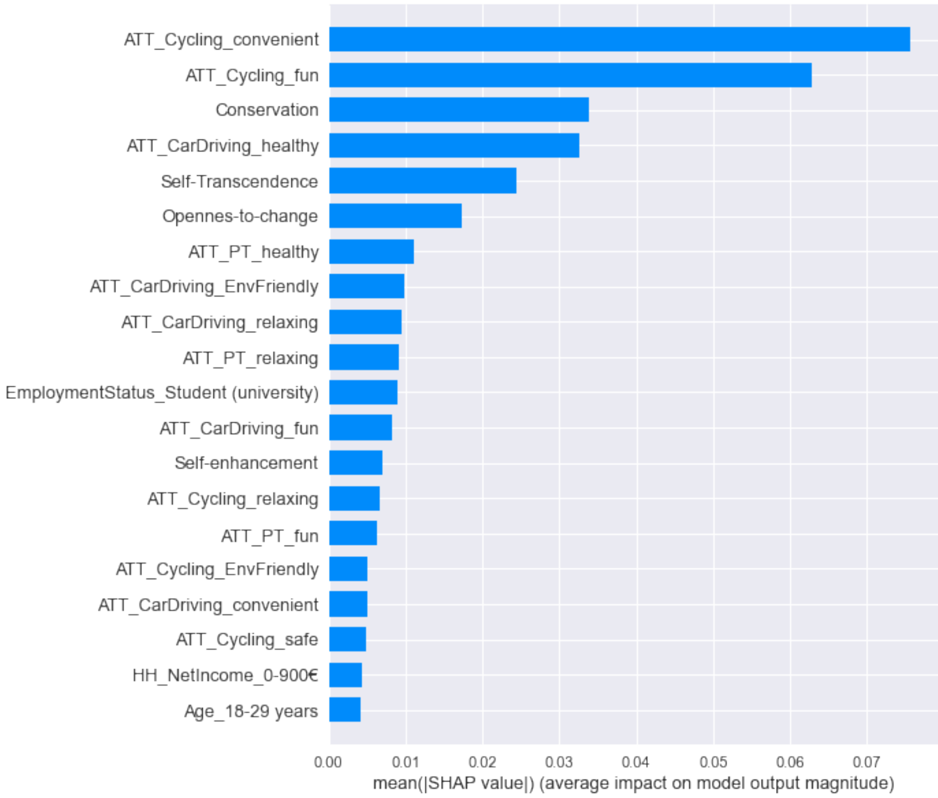

Figure 4.4.: SHAP Feature importance RF (training set)

4.3.3. Results of the models predicting public transport use

This table shows the results of machine learning algorithms predicting public transport use. The models were trained with either socio-demographic data or attitudinal data. The KNN and RF models achieved similar F1 scores of 0.65 when using socio-demographic data, while the SVM model had a lower F1 score of 0.41. The KNN model had a significantly lower F1 score of 0.45 when using attitude data, indicating poor performance. The RF model also had a low F1 score of 0.41, the SVM and XGBoost models performed better, with F1 scores of 0.45 and 0.56, respectively. Overall, it seems that attitude data does not perform as well as socio-demographic data in predicting public transport use.

Predicting Public transport usage with								
Models	Socio-demographics				Attitudes			
	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall
KNN	0.65	0.71	0.78	0.81	0.45	0.46	0.72	0.36
RF	0.65	0.68	0.81	0.70	0.41	0.49	0.64	0.61
SVM	0.41	0.69	0.69	1.0	0.45	0.45	0.66	0.59
XGBoost	0.59	0.63	0.76	0.68	0.56	0.61	0.73	0.70

Table 4.21.: Public transport use prediction: model performances on test set

Predicting public transport usage using								
Models	Values				Socio-demographics+Attitudes+Values			
	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall
KNN	0.52	0.60	0.71	0.72	0.58	0.61	0.76	0.76
RF	0.50	0.63	0.70	0.81	0.60	0.65	0.77	0.70
SVM	0.41	0.55	0.66	0.75	0.59	0.66	0.75	0.77
XGBoost	0.62	0.71	0.76	0.86	0.58	0.66	0.74	0.79

Table 4.22.: Public transport use prediction: model performances on test set

Table 4.22 shows the results of models trained using either value data or a combination of socio-demographic data, attitudes and values data. Overall, the models with the combined input data performed better than the models trained with only value data. Specifically, the SVM model had a lower F1 score and accuracy than the other models using socio-demographic, attitudes, and values data, while XGBoost had the highest F1 score and accuracy. The KNN model had higher precision and recall when using socio-demographic data, attitudes and values data than when using values data alone. The inclusion of socio-demographic, attitudes, and values data as input features can lead to improvement of the performance of machine learning models for predicting public transport use.

Figure 4.5 shows the plot of the ROC curves for all models using the different features. Using socio-demographics or combining socio-demographics, attitudes, and values results in the best ROC curves.

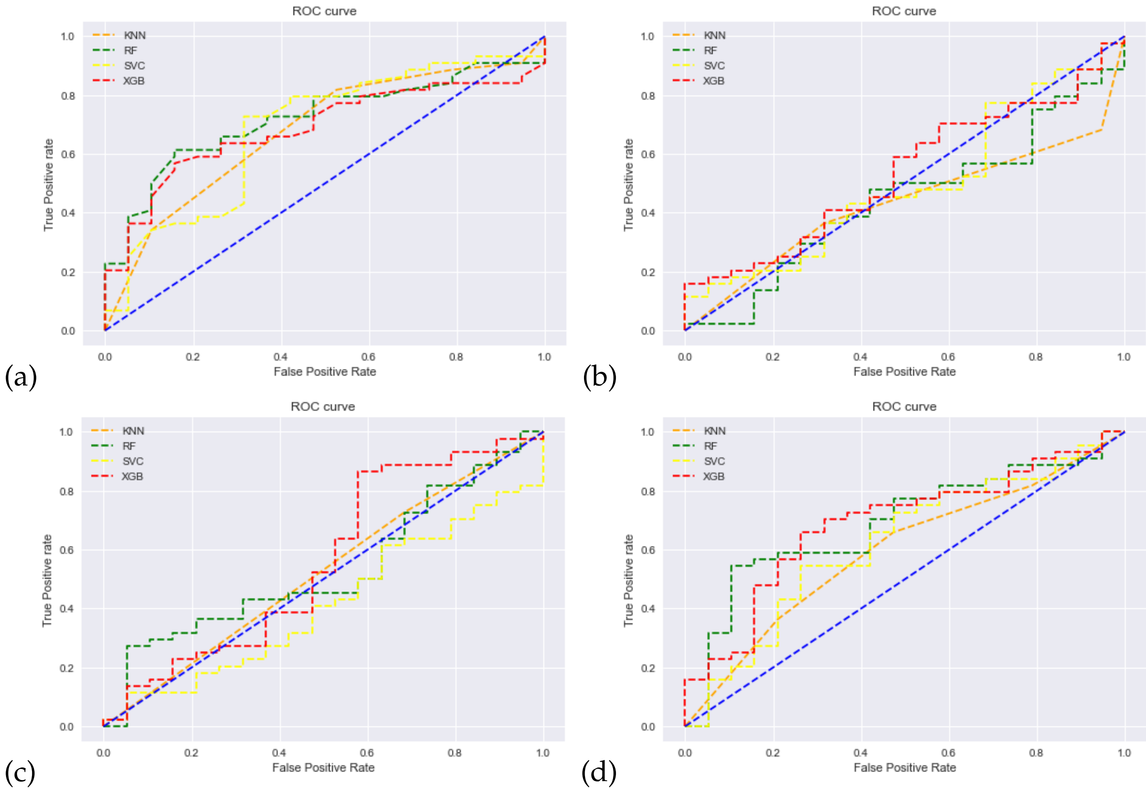

Figure 4.5.: (a) ROC curve using socio-demographic features (b) ROC curve using attitude features (c) ROC curve using value features (d) ROC curve using socio-demographic, attitude and value features

SHAP values

SHAP was used to interpret the model output using all the tree features. As depicted in Figure 4.6, the features that influence the output of the RF model on the training set are being a student, age between 18 and 29 years and self-transcendence.

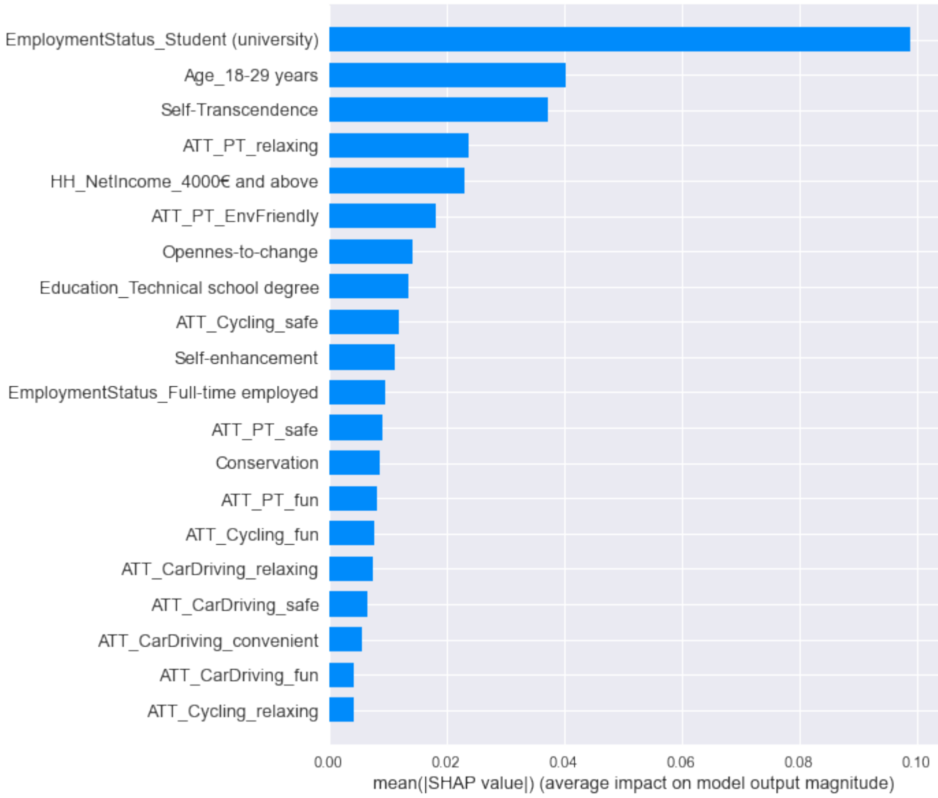

Figure 4.6.: SHAP Feature importance RF (training set)

5. Discussion

5.1. Approach 1: Structural equation modelling

5.1.1. Values and attitudes

The results of the CFA table indicate that the measurement model fits the data well. All factor loadings are statistically significant and have standardized coefficients greater than 0.40, indicating that they contribute to the construct they represent. This makes the interpretation of the relationships depicted by the structural model relevant.

The regression slope estimates suggest that the value of conservation influences attitudes toward driving. Specifically, higher levels of conservation are associated with more positive attitudes toward driving. On the other hand, there is a significant positive relationship between openness to change and attitudes toward cycling. This aligns with the work of R. Arroyo. et al. [81] and J. Garcıa et al. [84] and suggests that people who are more open to new experiences are more likely to have positive attitudes towards cycling. However, people with conservative values are more likely to have positive attitudes toward driving.

Attitudes towards public transport have no statistical significance with both conservation and openness to change. These results have already been indicated by the lack of a statistically significant correlation between the different measures of attitudes towards public transport and the scores representing conservation and openness to change. This non-significant relationship was also present in the work of R. Arroyo. et al. [81]. This can be explained by the attitudes selected for the survey that did not capture a good spectrum of perceptions of public transport, as indicated by the high residual variance of manifest variables such as "public transport is healthy" and "public transport is convenient". In the case of public transport, attitudes towards health, relaxation and fun are not as meaningful as for cycling and driving.

The latent variance estimates show that the conservation value and the attitude toward driving have the highest variance, suggesting that they best explain the variation in attitudes in the model. The Attitude towards biking factor has the lowest variance, indicating that it explains the least variation in attitudes in the model. Once again, this suggests that the survey questions did not capture significant attitudes towards public transport. Finally, the latent covariance estimates show a negative relationship between conservation and openness to change factors. This suggests that people with one value and not the other do not exhibit the same tendencies.

5.1.2. Attitudes and mode choic

The estimates of the regression slopes for private bike users, private car users, and public transport show that attitudes toward these different modes of transportation are significantly related. For example, positive attitudes toward cycling are associated with more negative attitudes toward driving and using public transportation. Conversely, more positive attitudes toward driving are associated with more negative attitudes toward cycling and public transportation. These results suggest that people who prefer a particular mode of transport are less likely to have positive attitudes toward other modes of transport.

Not all attitudes were significantly related to the use of each mode of transport. With the exception of cycling, where the three attitudes were significantly associated with its use, driving and using public transport were only significantly and positively associated with attitudes towards driving and public transport respectively. In addition, there is a significant negative association between attitudes towards driving and using a private bicycle and a significant positive association between attitudes towards driving and using a private car. This indicates that people with a positive attitude towards driving are more likely to use a car than a bicycle.

Interestingly, the residual covariances show a weak and non-significant relationship between using a private bike and using a private car. But there is a significant negative relationship between the use of a private bike and the use of public transport and a significant negative relationship between the use of a private car and public transport. This suggests that people who use a private bicycle or a private car are less likely to use public transport.

The latent variances show that attitudes towards driving have the highest variance, followed by attitudes towards public transport and cycling. This suggests that people have stronger attitudes towards driving than other transport modes.

5.1.3. Hypothesese assessement

Based on the results of the structural model, the hypotheses can now be confirmed or rejected. Hypothesis H1.: "Personal values are related to attitudes" is confirmed. The sub-hypothesis H1.a and H.1b are partially confirmed. Personal value conservation has a positive causal relationship with attitudes towards driving but not to attitudes towards public transport and cycling. The personal value of openness to change has no causal relationship with attitudes towards driving or attitudes towards public transport, but there is a positive causal relationship with attitudes towards cycling Hypothesis H2.: "Attitudes are related to transport mode choice" is confirmed. The sub-hypothesis H2.a is partially confirmed. Attitudes towards driving have a positive causal relationship with driving, but not to using public transport and cycling. Sub-hypothesis H2.b is rejected. Attitudes towards public transport have no statistically significant influence on the use of any of the three modes of transport. Sub-hypothesis H2.c is partially confirmed. Attitudes towards cycling have a positive causal relationship with cycling, but not with driving. Thus, the original conceptual framework presented in chapter 3 is not confirmed and needs to be modified. The alternative representation of the relationship between values, attitudes and mode choice is shown in the following figure:

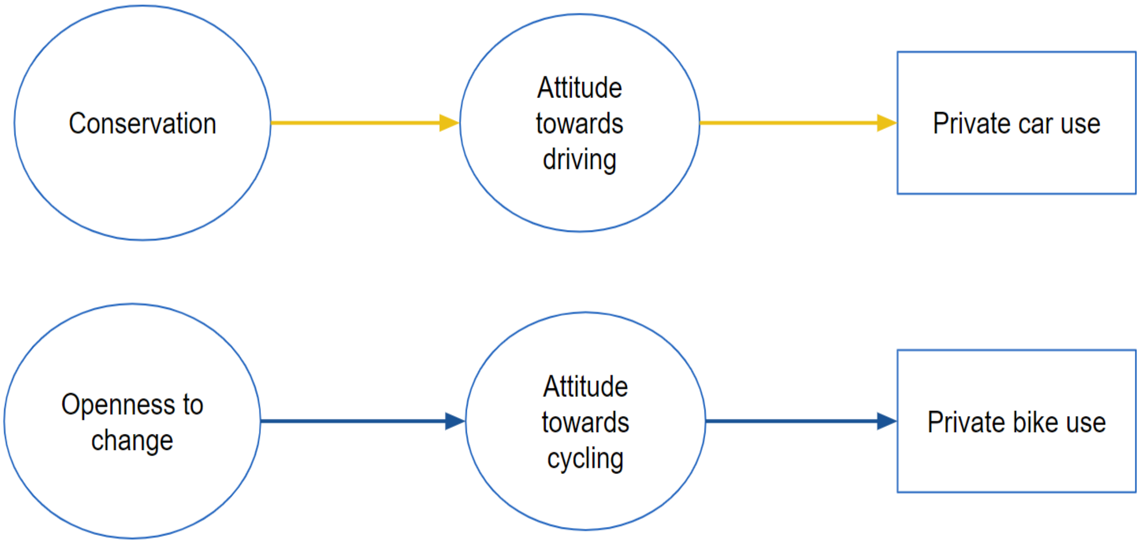

Figure 5.1.: New structure

5.2. Approach 2: Machine learning

For private car use, accuracy and precision scores were generally low for all models, which indicates that the models made many false positive predictions. Overall, the results suggest that the input data used in these models may not be sufficient to predict private car use accurately. This can be explained by the dataset's small number of private car users. A problem that was not solved by the oversampling technique. The ROC curve shows that the socio-demographic features led to the best performance of all models. The SHAP plot shows that age between 40 and 49 years and net income of 4000€and above are the two features that influence the model predictions most.

Regarding private bike use, the results suggest that the SVM, RF and XGBoost perform at a similar level when using either attitudes or a combination of socio-demographic data and values. The F1 scores for the three models ranged from 0.70 to 0.74 and the accuracy values were also high, ranging from 0.87 to 0.91, indicating that the models detect true positives. Combining socio-demographic data and values leads to better predicting private bike users. The SHAP plot shows that attitudes, convenience and fun, and value conservation are the features that influence the model predictions most.

In terms of public transport use, the results show that the KNN and RF models performed best when socio-demographic characteristics were used. However, the performance of the models performance was weaker when using attitudes features. Combining values, socio-demographics, and attitudes improved model performance compared to using values features alone. The XGBoost model performed best overall, achieving an F1 score of 0.62 and an accuracy of 0.71. However, the models still had relatively low accuracy and F1 scores, suggesting that there may be other factors influencing public transport use that are not captured by these features. The SHAP plot shows that being a student and age between 19 and 29 years and self-transcendence are features that influence the model predictions most.

5.3. Advantages and limitations of both approaches

The first approach, structural equation modelling, allows the testing of complex relationships between latent variables and observed variables. It provided estimates and residual variances that helped to understand the relationships between the different variables. In addition, its confirmatory purposes made it possible to assess the goodness of attitudes obtained from the data set. For this, this approach is very useful for constructing abstract variables, examining the relationship between them, and then relating them to the observed behaviour. All this can be interpreted and understood with the indices and fit of the model.

However, many measurements are needed to construct a well-fitting model, i.e. for each latent variable, which is not possible in reality because people do not fill out long surveys. So to get meaningful results, the survey should be thought out beforehand. For example, attitudes towards public transport did not yield statistical significance because it did not represent the overall contract.

The machine learning approach has the great advantage of processing complex datasets with multiple variables. Since it is a data-driven approach, it does not require assumptions about the relationship between variables. It can also deal with outliers. This approach provided reasonable predictions even though the dataset was not large. The use of shap values made it possible to see which variables influenced the models' predictions and to draw conclusions based on these results. However, machine learning algorithm results are not accurate on small datasets because the dataset did not contain enough variation to train properly. This makes exploring psychological features with machine learning difficult, as thousands of input data are needed to capture the necessary variation. Furthermore, machine learning cannot test a causality relationship.

Both approaches have their own advantages and disadvantages. The use of either method should depend on the objective. Testing for causal relationships based on an established theory is very possible with SEM, whereas with large amounts of data and the goal of only making predictions, a machine learning approach makes more sense.

6. Conclusion

6.1. General conclusion and answering the research questions

In this thesis, two approaches were taken to explore the realtionship of values and attitudes with mode choice. The first approach is structural equation modelling to test the value-attitude-behaviour theory by Homer and Kahle [46]. The second approach is machine learning to predict travel mode choice using four different sets of input data. The dataset used for this thesis is a survey completed by Michael Stoeckle for his Master's thesis at the Technical University of Munich.

The first approach consists of a structural equation model that is used to test and confirm or reject the value-attitude-behaviour theory. Schwartz's 10 basic values, attitudes towards driving, cycling and public transport were the variables used to understand transport choice. The main findings are that not all the values tested were found to be causally related to attitudes, and the same was true for the relationship between attitudes and mode choice.

The model's results empirically proved the causal relationship between the value of conservation and attitudes towards driving, and thus the causal relationship between them and private car use. The results also proved the causal relationship between openness to change and attitudes towards cycling, and thus the causal relationship between them and cycling. The weak and non-significant relationship between the use of a private bike and the use of a private car and public transport suggests that these modes may be used for different purposes. Finally, the fit indices for the structural model are acceptable, suggesting that the model provides a reasonable explanation for the data.

Four machine learning algorithms were used to predict mode choice. The algorithms are K-Nearest Neighbours (KNN), Random Forest (RF), Support Vector Machine (SVM) and eXtreme Gradient Boosting (XGBoost). Each algorithm was used with four different sets of features. The first feature is socio-demographics, the second is attitudes, the third is values and finally, a combination of these three features. All four algorithms were tasked with predicting car use, bike use and public transport use. The performance of the models in predicting each travel mode choice was assessed. The results suggest that the accuracy and precision scores for predicting private car use were generally low. However, the performance of the models in predicting private bike use was better, with the SVM, RF, and XGBoost models performing at a similar level. The KNN and RF models performed best for public transport use when using socio-demographic features, but the performance was weaker when using attitudes. Combining socio-demographic data, attitudes, and values improved model performance.

SHAP values were used to assess the importance of the features when the combination of socio-demographics, values and attitudes was used. Age, income, "Car driving is relaxing" attitude and conservation value were the most important features influencing the model. Regarding bike use, the attitudes "cycling is convenient" and "cycling is fun" and the value conservation were the three features with the most significant influence on the model output. Finally, the use of public transport was influenced by the category of student in employment status, age between 18 and 29 and the value of self-transcendence.

Based on these findings, the research questions can be answered. First, values are associated with mode choice. They directly affect attitudes, which in turn affect travel mode choice. This makes values have a mediator effect on mode choice. Used alone as input variables in the context of machine learning did not deliver the best performance. Combining values with attitudes and socio-demographic variables leads to better predictions. The machine learning algorithms Random Forest (RF), Support Vector Machine (SVM) and XGBoost have performed the best across the input features and could relate values to mode choice. Structural equation modelling is a useful technique that captures the complex interactions between factors that influence mode choice, such as values and attitudes and provides insights into how they interact. Although machine learning cannot provide such insights, results have shown that psychological factors can be added to input data to predict mode choice. With new librairies like shap,the results of machine learning algorithms can be better interpreted and important features extracted.

6.2. Future work and recommendations

Based on the conclusions of this thesis, there are several starting points for future work and recommendations. First, the sample could be larger and the characteristics of the respondents could be more diverse. In this way, the subtleties of values and attitudes can be better mapped. A combination of structural equation modelling and machine learning should be considered to better understand the relationships between personal values, attitudes and transport choices. This leads to a better understanding of the complex interactions between these factors and helps identify key variables that influence mode choice. It is worth noting that the attitudes used in this thesis were not all appropriate to capture the complexity of each mode of transport. Attitudes that better describe perception can only lead to a better understanding of behaviour.

The results of this thesis can help to make some recommendations that can reduce private car use: Cycling can be promoted, especially among people who are open to change. These people are likely interested in trying something new and seeking stimulation, so they are more likely to cycle. The results of the causal relationship between attitudes towards cycling and cycling can be used to promote cycling as fun rather than focusing on the environmental friendliness of cycling. Developing interventions that target specific attitudes towards driving might turn out to be effective, given the causal relationship between attitudes towards driving. Policymakers can use this information to develop targeted interventions that address the perceived fun that comes from driving a car and targets the relaxing part of driving.

A. Appendix

Basic Value	PVQ Items as Numbered and Labeled in the survey
universalism	• Item 3: He thinks it is important that every person in the world should be treated equally. He believes everyone should have equal opportunities in life. • Item 8: It is important to him to listen to people who are different from him. Even when he disagrees with them, he still wants to understand them. • Item 19: He strongly believes that people should care for nature. Looking after the environment is important to him.
benevolence	• Item 12: It is very important to him to help the people around him. He wants to care for their well-being. • Item 18:It is important to him to be loyal to his friends. He wants to devote himself to people close to him.
tradition	• Item 9: It is important to him to be humble and modest. He tries not to draw attention to himself. • Item 20:Tradition is important to him. He tries to follow the custom handed down by his religion or his family.

conformity	• Item 7:: He believes that people should do what they are told. He thinks people should follow rules at all times, even when no one is watching. • Item 16:It is important to him always to behave properly. He wants to avoid doing anything people would say is wrong.
security	• Item 5: It is important to him to live in secure surroundings. He avoids anything that might endanger his safety. • Item 14:It is important to him that the government ensures his safety against all threats. He wants the state to be strong so it can defend its citizens.
power	• Item 2: It is important to him to be rich. He wants to have a lot of money and expensive things. • Item 17: It is important to him to get respect from others. He wants people to do what he says.
achievement	• Item 4: It is important to him to show his abilities. He wants people to admire what he does. • Item 13: Being very successful is important to him. He hopes people will recognize his achievements.
hedonism	• Item 10: Having a good time is important to him. He likes to "spoil" himself. • Item 21: He seeks every chance he can to have fun. It is important to him to do things that give him pleasure.

stimulation	• Item 6: He likes surprises and is always looking for new things to do. He thinks it is important to do lots of different things in life. • Item 15: He looks for adventures and likes to take risks. He wants to have an exciting life.
self-direction	• Item 1:Thinking up new ideas and being creative is important to him. He likes to do things in his own original way. • Item 11: It is important to him to make his own decisions about what he does. He likes to be free and not depend on others.

List of Figures

List of Tables

Bibliography

[1] C. Change et al. “Intergovernmental panel on climate change”. In: *World Meteorological Organization* 52 (2007).

[2] V. Ramanathan and Y. Feng. “Air pollution, greenhouse gases and climate change: Global and regional perspectives”. In: *Atmospheric environment* 43.1 (2009), pp. 37–50.

[3] S. M. Ogle, L. Olander, L. Wollenberg, T. Rosenstock, F. Tubiello, K. Paustian, L. Buendia, A. Nihart, and P. Smith. “Reducing greenhouse gas emissions and adapting agricultural management for climate change in developing countries: providing the basis for action”. In: *Global change biology* 20.1 (2014), pp. 1–6.

[4] A. Haines, A. J. McMichael, K. R. Smith, I. Roberts, J. Woodcock, A. Markandya, B. G. Armstrong, D. Campbell-Lendrum, A. D. Dangour, M. Davies, et al. “Public health benefits of strategies to reduce greenhouse-gas emissions: overview and implications for policy makers”. In: *The lancet* 374.9707 (2009), pp. 2104–2114.

[5] D. L. Greene, H. H. Baker Jr, and S. E. Plotkin. “Reducing greenhouse gas emissions from US transportation”. In: (2010).

[6] C. Yang, D. McCollum, R. McCarthy, and W. Leighty. “Meeting an 80% reduction in greenhouse gas emissions from transportation by 2050: A case study in California”. In: *Transportation Research Part D: Transport and Environment* 14.3 (2009), pp. 147–156.

[7] J. Delbeke and P. Vis. *Towards a climate-neutral Europe: curbing the trend*. 2020.

[8] W. F. Lamb, M. Grubb, F. Diluiso, and J. C. Minx. “Countries with sustained greenhouse gas emissions reductions: An analysis of trends and progress by sector”. In: *Climate Policy* 22.1 (2022), pp. 1–17.

[9] D. Fiorello, A. Martino, L. Zani, P. Christidis, and E. Navajas-Cawood. “Mobility data across the EU 28 member states: results from an extensive CAWI survey”. In: *Transportation Research Procedia* 14 (2016), pp. 1104–1113.

[10] J. H. Barnes, T. J. Chatterton, and J. W. Longhurst. “Emissions vs exposure: Increasing injustice from road traffic-related air pollution in the United Kingdom”. In: *Transportation research part D: transport and environment* 73 (2019), pp. 56–66.

[11] W. H. Organization et al. *Air pollution and child health: prescribing clean air: summary*. Tech. rep. World Health Organization, 2018.

[12] N. Hooftman, M. Messagie, J. Van Mierlo, and T. Coosemans. “A review of the European passenger car regulations–Real driving emissions vs local air quality”. In: *Renewable and Sustainable Energy Reviews* 86 (2018), pp. 1–21.

[13] B. Gu, L. Zhang, R. Van Dingenen, M. Vieno, H. J. Van Grinsven, X. Zhang, S. Zhang, Y. Chen, S. Wang, C. Ren, et al. "Abating ammonia is more cost-effective than nitrogen oxides for mitigating PM2. 5 air pollution". In: *Science* 374.6568 (2021), pp. 758–762.

[14] M. Hachem, N. Saleh, L. Bensefa-Colas, and I. Momas. "Determinants of ultrafine particles, black carbon, nitrogen dioxide, and carbon monoxide concentrations inside vehicles in the Paris area: PUF-TAXI study". In: *Indoor air* 31.3 (2021), pp. 848–859.

[15] Z. Fan and C. D. Harper. "Congestion and environmental impacts of short car trip replacement with micromobility modes". In: *Transportation Research Part D: Transport and Environment* 103 (2022), p. 103173.

[16] D. Schrank, B. Eisele, T. Lomax, et al. "Urban mobility report 2019". In: (2019).

[17] T. E. C. of Auditors. *Urban mobility in the EU*. Accessed: 24-4-2023.

[18] R. Cervero. "Road expansion, urban growth, and induced travel: A path analysis". In: *Journal of the American Planning Association* 69.2 (2003), pp. 145–163.

[19] W. Fan and Z. Gurmu. "Combined decision making of congestion pricing and capacity expansion: genetic algorithm approach". In: *Journal of Transportation Engineering* 140.8 (2014), p. 04014031.

[20] M. J. Nieuwenhuijsen and H. Khreis. "Car free cities: Pathway to healthy urban living". In: *Environment international* 94 (2016), pp. 251–262.

[21] A. May. "Road pricing: an international perspective". In: *Transportation* 19 (1992), pp. 313–333.

[22] J. N. Gonzalez, J. Gomez, and J. M. Vassallo. "Do urban parking restrictions and Low Emission Zones encourage a greener mobility?" In: *Transportation Research Part D: Transport and Environment* 107 (2022), p. 103319.

[23] E. Croci. "Urban road pricing: a comparative study on the experiences of London, Stockholm and Milan". In: *Transportation Research Procedia* 14 (2016), pp. 253–262.

[24] S. Bamberg, S. Fujii, M. Friman, and T. Gärling. "Behaviour theory and soft transport policy measures". In: *Transport policy* 18.1 (2011), pp. 228–235.

[25] G. Möser and S. Bamberg. "The effectiveness of soft transport policy measures: A critical assessment and meta-analysis of empirical evidence". In: *Journal of Environmental Psychology* 28.1 (2008), pp. 10–26.

[26] S. A. Shaheen, H. Zhang, E. Martin, and S. Guzman. "China's Hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing". In: *Transportation Research Record* 2247.1 (2011), pp. 33–41.

[27] S. Cairns, L. Sloman, C. Newson, J. Anable, A. Kirkbride, and P. Goodwin. "Smarter choices: assessing the potential to achieve traffic reduction using 'soft measures'". In: *Transport reviews* 28.5 (2008), pp. 593–618.

[28] M. A. Taylor. "Voluntary travel behavior change programs in Australia: The carrot rather than the stick in travel demand management". In: *International Journal of Sustainable Transportation* 1.3 (2007), pp. 173–192.

[29] S. Fujii, S. Bamberg, M. Friman, and T. Gärling. "Are effects of travel feedback programs correctly assessed?" In: *Transportmetrica* 5.1 (2009), pp. 43–57.

[30] J. Richter, M. Friman, and T. Gärling. "Soft transport policy measures 1: Results of implementations". In: (2009).

[31] A. Semenescu, A. Gavreliuc, and P. Sârbescu. "30 Years of soft interventions to reduce car use–A systematic review and meta-analysis". In: *Transportation research part D: transport and environment* 85 (2020), p. 102397.

[32] M. Veternık and M. Gogola. "Examining of correlation between demographic development of population and their travel behaviour". In: *Procedia engineering* 192 (2017), pp. 929–934.

[33] P. P. Sarkar and C. Mallikarjuna. "Effect of land use on travel behaviour: a case study of Agartala city". In: *Procedia-Social and Behavioral Sciences* 104 (2013), pp. 533–542.

[34] H. Pan, Q. Shen, and M. Zhang. "Influence of urban form on travel behaviour in four neighbourhoods of Shanghai". In: *Urban studies* 46.2 (2009), pp. 275–294.

[35] J. Anable. "'Complacent car addicts' or 'aspiring environmentalists'? Identifying travel behaviour segments using attitude theory". In: *Transport policy* 12.1 (2005), pp. 65–78.

[36] J. Prillwitz and S. Barr. "Moving towards sustainability? Mobility styles, attitudes and individual travel behaviour". In: *Journal of transport geography* 19.6 (2011), pp. 1590–1600.

[37] M. Paulssen, D. Temme, A. Vij, and J. L. Walker. "Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice". In: *Transportation* 41 (2014), pp. 873–888.

[38] I. Ajzen. "The theory of planned behavior". In: *Organizational behavior and human decision processes* 50.2 (1991), pp. 179–211.

[39] R. Setiawan, W. Santosa, and A. Sjafruddin. "Integration of theory of planned behavior and norm activation model on student behavior model using cars for traveling to campus". In: *Civil Engineering Dimension* 16.2 (2014), pp. 117–122.

[40] I. J. Donald, S. R. Cooper, and S. M. Conchie. "An extended theory of planned behaviour model of the psychological factors affecting commuters' transport mode use". In: *Journal of environmental psychology* 40 (2014), pp. 39–48.

[41] E. Heinen, K. Maat, and B. Van Wee. "The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances". In: *Transportation research part D: transport and environment* 16.2 (2011), pp. 102–109.

[42] I. Ajzen. "Values, attitudes, and behavior". In: *Methods, theories, and empirical applications in the social sciences* (2012), pp. 33–38.

[43] S. H. Schwartz. "Basic human values: Theory, measurement, and applications". In: *Revue française de sociologie* 47.4 (2007), p. 929.

[44] L. Sagiv, S. Roccas, J. Cieciuch, and S. H. Schwartz. “Personal values in human life”. In: *Nature human behaviour* 1.9 (2017), pp. 630–639.

[45] M. H. Bond, K. Leung, A. Au, K.-K. Tong, and Z. Chemonges-Nielson. “Combining social axioms with values in predicting social behaviours”. In: *European journal of Personality* 18.3 (2004), pp. 177–191.

[46] P. M. Homer and L. R. Kahle. “A structural equation test of the value-attitude-behavior hierarchy.” In: *Journal of Personality and social Psychology* 54.4 (1988), p. 638.

[47] Y. Wang, X. Yan, Y. Zhou, and Q. Xue. “Influencing mechanism of potential factors on passengers’ long-distance travel mode choices based on structural equation modeling”. In: *Sustainability* 9.11 (2017), p. 1943.

[48] R. Devika, M. Harikrishna, and M. Anjaneyulu. “Influence of psychological factors in mode choice decision making: A structural equation modeling approach”. In: *Transportation research procedia* 48 (2020), pp. 2821–2830.

[49] A. Anas. “The estimation of multinomial logit models of joint location and travel mode choice from aggregated data”. In: *Journal of regional science* 21.2 (1981), pp. 223–242.

[50] K. Train. “A structured logit model of auto ownership and mode choice”. In: *The Review of Economic Studies* 47.2 (1980), pp. 357–370.

[51] Y. O. Susilo, N. Hanks, and M. Ullah. “An exploration of shoppers travel mode choice in visiting convenience stores in the UK”. In: *Transportation Planning and Technology* 36.8 (2013), pp. 669–684.

[52] J. D. Pineda-Jaramillo. “A review of Machine Learning (ML) algorithms used for modeling travel mode choice”. In: *Dyna* 86.211 (2019), pp. 32–41.

[53] J. Bruin. *newtest: command to compute new test @ONLINE*. Feb. 2011. URL: `https://stats.oarc.ucla.edu/stata/ado/analysis/`.

[54] S. Varotto, R. Krueger, and M. Bierlaire. “Modelling travel behaviour: A choice modelling perspective”. In: (2022).

[55] L. Cheng, X. Chen, J. De Vos, X. Lai, and F. Witlox. “Applying a random forest method approach to model travel mode choice behavior”. In: *Travel behaviour and society* 14 (2019), pp. 1–10.

[56] X. Zhao, X. Yan, A. Yu, and P. Van Hentenryck. “Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models”. In: *Travel behaviour and society* 20 (2020), pp. 22–35.

[57] M. Rokeach. *The nature of human values.* Free press, 1973.

[58] N. T. Feather. “Cross-cultural studies with the Rokeach Value Survey: The Flinders program of research on values”. In: *Australian Journal of Psychology* 38.3 (1986), pp. 269–283.

[59] A. Shoemaker and M. Bolt. “The Rokeach value survey and perceived Christian values”. In: *Journal of Psychology and Theology* 5.2 (1977), pp. 139–142.

[60] P. Suhonen. “Approaches to value research and value measurement”. In: *Acta Sociologica* 28.4 (1985), pp. 349–358.

[61] R. Boudon. “The present relevance of Max Weber’s Wertrationalität (value rationality)”. In: *Methodology of the social sciences, ethics, and economics in the newer historical school: From Max Weber and Rickert to Sombart and Rothacker*. Springer. 1997, pp. 3–31.

[62] S. H. Schwartz et al. “An overview of the Schwartz theory of basic values”. In: *Online readings in Psychology and Culture* 2.1 (2012), pp. 2307–0919.

[63] S. Bergner, C. Palmer, M. Devaney, and P. Kruse. “A framework for antecedents of social entrepreneurial intention: Empirical evidence and research agenda”. In: *Frontiers in Psychology* 13 (2022).

[64] D. A. Ralston, C. P. Egri, E. Reynaud, N. Srinivasan, O. Furrer, D. Brock, R. Alas, F. Wangenheim, F. L. Darder, C. Kuo, et al. “A twenty-first century assessment of values across the global workforce”. In: *Journal of business ethics* 104 (2011), pp. 1–31.

[65] S. H. Schwartz, G. Melech, A. Lehmann, S. Burgess, M. Harris, and V. Owens. “Extending the cross-cultural validity of the theory of basic human values with a different method of measurement”. In: *Journal of cross-cultural psychology* 32.5 (2001), pp. 519–542.

[66] S. H. Schwartz and K. Boehnke. “Evaluating the structure of human values with confirmatory factor analysis”. In: *Journal of research in personality* 38.3 (2004), pp. 230–255.

[67] O. Aride and M.-d.-M. Pamies-Pallise. “From values to behavior: Proposition of an integrating model”. In: *Sustainability* 11.21 (2019), p. 6170.

[68] I. Ajzen. *EBOOK: Attitudes, Personality and Behaviour*. McGraw-hill education (UK), 2005.

[69] A. R. Pratkanis, S. J. Breckler, and A. G. Greenwald. *Attitude structure and function*. Psychology Press, 2014.

[70] J. L. Hale, B. J. Householder, and K. L. Greene. “The theory of reasoned action”. In: *The persuasion handbook: Developments in theory and practice* 14.2002 (2002), pp. 259–286.

[71] C. Lakhan. “The garbage gospel: Using the theory of planned behavior to explain the role of religious institutions in affecting pro-environmental behavior among ethnic minorities”. In: *The Journal of Environmental Education* 49.1 (2018), pp. 43–58.

[72] T. L. Milfont, J. Duckitt, and C. Wagner. “A cross-cultural test of the value–attitude–behavior hierarchy”. In: *Journal of Applied Social Psychology* 40.11 (2010), pp. 2791–2813.

[73] J. Ko, S. Lee, and M. Byun. “Exploring factors associated with commute mode choice: An application of city-level general social survey data”. In: *Transport policy* 75 (2019), pp. 36–46.

[74] A. Papagiannakis, I. Baraklianos, and A. Spyridonidou. “Urban travel behaviour and household income in times of economic crisis: Challenges and perspectives for sustainable mobility”. In: *Transport policy* 65 (2018), pp. 51–60.

[75] L. Böcker, P. van Amen, and M. Helbich. "Elderly travel frequencies and transport mode choices in Greater Rotterdam, the Netherlands". In: *Transportation* 44 (2017), pp. 831–852.

[76] M. V. Johansson, T. Heldt, and P. Johansson. "The effects of attitudes and personality traits on mode choice". In: *Transportation Research Part A: Policy and Practice* 40.6 (2006), pp. 507–525.

[77] L. Li and Y. Zhang. "An extended theory of planned behavior to explain the intention to use carsharing: a multi-group analysis of different sociodemographic characteristics". In: *Transportation* (2021), pp. 1–39.

[78] W. Chen, C. Cao, X. Fang, and Z. Kang. "Expanding the theory of planned behaviour to reveal urban residents' pro-environment travel behaviour". In: *Atmosphere* 10.8 (2019), p. 467.

[79] W. Li and M. Kamargianni. "An integrated choice and latent variable model to explore the influence of attitudinal and perceptual factors on shared mobility choices and their value of time estimation". In: *Transportation science* 54.1 (2020), pp. 62–83.

[80] X. Hu, N. Wu, and N. Chen. "Young people's behavioral intentions towards low-carbon travel: Extending the theory of planned behavior". In: *International journal of environmental research and public health* 18.5 (2021), p. 2327.

[81] R. Arroyo, T. Ruiz, L. Mars, S. Rasouli, and H. Timmermans. "Influence of values, attitudes towards transport modes and companions on travel behavior". In: *Transportation research part F: traffic psychology and behaviour* 71 (2020), pp. 8–22.

[82] A. Nordlund and K. Westin. "Influence of values, beliefs, and age on intention to travel by a new railway line under construction in northern Sweden". In: *Transportation Research Part A: Policy and Practice* 48 (2013), pp. 86–95.

[83] M. Hunecke, S. Haustein, S. Böhler, and S. Grischkat. "Attitude-based target groups to reduce the ecological impact of daily mobility behavior". In: *Environment and behavior* 42.1 (2010), pp. 3–43.

[84] J. Garcıa, L. Mars, R. Arroyo, D. Casquero, F. di Ciommo, and T. Ruiz. "Personal values, attitudes and travel intentions towards cycling and walking, and actual behavior". In: *Sustainability* 11.13 (2019), p. 3574.

[85] R. Wall, P. Devine-Wright, and G. A. Mill. "Comparing and combining theories to explain proenvironmental intentions: The case of commuting-mode choice". In: *Environment and behavior* 39.6 (2007), pp. 731–753.

[86] T. F. Golob. "Structural equation modeling for travel behavior research". In: *Transportation Research Part B: Methodological* 37.1 (2003), pp. 1–25.

[87] H. Kang and J.-W. Ahn. "Model setting and interpretation of results in research using structural equation modeling: A checklist with guiding questions for reporting". In: *Asian Nursing Research* 15.3 (2021), pp. 157–162.

[88] J. B. Ullman and P. M. Bentler. “Structural equation modeling”. In: *Handbook of Psychology, Second Edition* 2 (2012).

[89] T. Larsson, L. Plonsky, and G. R. Hancock. “On the benefits of structural equation modeling for corpus linguists”. In: *Corpus Linguistics and Linguistic Theory* 17.3 (2020), pp. 683–714.

[90] L. L. Thurstone. “A law of comparative judgment.” In: *Psychological review* 34.4 (1927), p. 273.

[91] R. D. Luce. *Individual choice behavior: A theoretical analysis, new york, ny: John willey and sons*. 1959.

[92] J. Marschak. “Binary-choice constraints and random utility indicators (1960)”. In: *Economic information, decision, and prediction: selected essays: volume I Part I economics of decision*. Springer, 1974, pp. 218–239.

[93] D. McFadden et al. “Conditional logit analysis of qualitative choice behavior”. In: (1973).

[94] T. Domencich and D. McFadden. *A Behavioural Model of Urban Travel Demand*. 1975.

[95] M. T. Kashifi, A. Jamal, M. S. Kashefi, M. Almoshaogeh, and S. M. Rahman. “Predicting the travel mode choice with interpretable machine learning techniques: A comparative study”. In: *Travel Behaviour and Society* 29 (2022), pp. 279–296.

[96] V. Van Can. “Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model”. In: *Transportation Research Part A: Policy and Practice* 49 (2013), pp. 149–159.

[97] Y. Kim, T.-Y. Kim, and E. Heo. “Bayesian estimation of multinomial probit models of work trip choice”. In: *Transportation* 30 (2003), pp. 351–365.

[98] L. Shanmugam and M. Ramasamy. “Study on mode choice using nested logit models in travel towards Chennai metropolitan city”. In: *Journal of Ambient Intelligence and Humanized Computing* (2021), pp. 1–9.

[99] M. Elmorssy and T. Onur. “Modelling departure time, destination and travel mode choices by using generalized nested logit model: discretionary trips”. In: *International journal of engineering* 33.2 (2020), pp. 186–197.

[100] M. Ye, Y. Chen, G. Yang, B. Wang, and Q. Hu. “Mixed logit models for travelers’ mode shifting considering bike-sharing”. In: *Sustainability* 12.5 (2020), p. 2081.

[101] J. Guo, T. Feng, and H. J. Timmermans. “Co-dependent workplace, residence and commuting mode choice: Results of a multi-dimensional mixed logit model with panel effects”. In: *Cities* 96 (2020), p. 102448.

[102] D. A. Hensher and W. H. Greene. “The mixed logit model: the state of practice”. In: *Transportation* 30 (2003), pp. 133–176.

[103] M. Ben-Akiva, D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Boersch-Supan, D. Brownstone, D. S. Bunch, et al. “Hybrid choice models: Progress and challenges”. In: *Marketing Letters* 13 (2002), pp. 163–175.

[104] A. Glerum, B. Atasoy, A. Monticone, and M. Bierlaire. "Adjectives qualifying individuals' perceptions impacting on transport mode preferences". In: *Second International Choice Modeling Conference*. CONF. 2011.

[105] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, J. Akinjobi, et al. "Supervised machine learning algorithms: classification and comparison". In: *International Journal of Computer Trends and Technology (IJCTT)* 48.3 (2017), pp. 128–138.

[106] F. Wang and C. L. Ross. "Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model". In: *Transportation Research Record* 2672.47 (2018), pp. 35–45.

[107] C. A. U. Hassan, M. S. Khan, and M. A. Shah. "Comparison of machine learning algorithms in data classification". In: *2018 24th International Conference on Automation and Computing (ICAC)*. IEEE. 2018, pp. 1–6.

[108] J. Brownlee. *Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end*. Machine Learning Mastery, 2016.

[109] S. Wang, B. Mo, S. Hess, and J. Zhao. "Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: an empirical benchmark". In: *arXiv preprint arXiv:2102.01130* (2021).

[110] J. Chen and S. Li. "Mode choice model for public transport with categorized latent variables". In: *Mathematical Problems in Engineering* 2017 (2017).

[111] M. Stöckl. *Mathematical Modeling of Discrete Sports Behaviors as Continuous Topologies – Applications in Golf and Field Hockey*. en. 2013.

[112] W. Bilsky, M. Janik, and S. H. Schwartz. "The structural organization of human values-evidence from three rounds of the European Social Survey (ESS)". In: *Journal of cross-cultural psychology* 42.5 (2011), pp. 759–776.

[113] R. Mohammed, J. Rawashdeh, and M. Abdullah. "Machine learning with oversampling and undersampling techniques: overview study and experimental results". In: *2020 11th international conference on information and communication systems (ICICS)*. IEEE. 2020, pp. 243–248.