UiT THE ARCTIC UNIVERSITY OF NORWAY

FACULTY OF SCIENCE AND TECHNOLOGY
Department of Computer Science

Data Analysis and Nudging for Green Transportation

—

Jemea Lady Limunga

INF-3990 Master's Thesis in Computer Science
May 2019

Declaration

I, Jemea Lady Limunga hereby declare that this thesis in its entirety has been composed by myself and has not been submitted, in whole or part for any previous degree or professional qualification. However, I have been part of the Open Distributed Systems (ODS) research group, working in close collaboration with Anders Andersen and Randi Karlsen as supervisors and Cosmin Radu Crăciun who worked on the Data management part of this implementation. Any other form of information or inspiration gotten from other people's work has been well referenced.

Abstract

The topic of a more sustainable environment has been a core factor to governments and the public for many years. Sustainability in land transportation brings about less emission of greenhouse gases, less pollution, less traffic, a healthier and more active society. Ignorance to sustainability has brought about several environmental and human concerns including global warming. Global warming is the measured increase in the temperature of the earth. Multiple factors have been known to contribute to global warming which fall under anthropological and natural causes. In this thesis, we focus on one of the several human activities which contributes to global warming, road transportation. Several means of road transportation exist with some being more sustainable than others. The goal for this thesis is to introduce a system of information and digital nudging where users of the system are presented with vital information and nudge messages which could influence their transportation choices towards more sustainable options. I approached this proposed solution by doing analysis of data from multiple sources and channeling the results of the analysis in a web application through push notifications. While the data and implementation for this thesis is specific for the municipality of Tromsø, Norway, the concept can easily be extended and adapted to suit any other town or country.

Acknowledgement

I would like to thank Professors Anders Andersen and Randi Karlsen of the ODS research group in the Faculty of Science and Technology of UiT- The Arctic University of Norway, for the idea of Green Transportation choices with IoT and smart nudging and for letting me be a part of the implementation of such an interesting and useful project. They constantly offered guidance and input throughout the process of implementing and documenting this thesis.

I would also like to thank Cosmin Radu Crăciun whom I worked in close collaboration with throughout my studies at UiT and for helping with proof reading of this thesis.

Finally, I express my profound gratitude to my mother Mrs. Julie Mofoke and my husband Gerald Obi for providing me with continuous encouragement and support throughout my studies and through the process of writing this thesis. Their contribution made all this possible.

Table of Contents

List of Figures

List of Abbreviations

ACID – Atomicity, Consistency, Isolation, Durability

API – Application Programming Interface

BEV – Battery Electric Vehicles

CAP – Consistency, Availability, Partition tolerance

CFC – Chlorofluorocarbon

CH_4- Methane

CO – Carbon Monoxide

CO_2 - Carbon Dioxide

DBMS – Database Management System

EF – Environmental Friendliness

EV – Electric Vehicles

GDP – Gross Domestic Product

GHG – Greenhouse Gas

HC – Hydrocarbons

HCFC – Hydro Chlorofluorocarbon

HEV – Hybrid Electric Vehicles

HTTP – Hypertext Transfer Protocol

ICE – Internal Combustion Engine

NewSQL- New SQL

NF_3 - Nitrogen Trifluoride

NoSQL – Not Only SQL

NOx – Nitrogen Oxides

ODS – Ozone Depleting Substances

OLAP – OnLine Analytical Processing

PDS – Photo Display System

PFC – Perfluorocarbons

PHEV – Plug-in Hybrid Electric Vehicles

PM – Particulate Matter

RDBMS – Relational Database Management System

REX – Range Extenders

SF_6 - Sulphur Hexafluoride

SO_2 - Sulphur Dioxide

SQL – Structured Query Language

VOC – Volatile Organic Compounds

VTTSs – Value of Time Travel Savings

List of Tables

1 Introduction

1.1 Motivation

Climate change and global warming are terms which have become major concerns not only to environmentalists, governments and academia but also to the public. In recent years, governments, private organizations, schools and individuals have become more aware of these phenomena and are focusing on protecting and preserving the environment from drastic climate changes. There was a recorded increase in the number and lengths of heatwaves in 2018 and it is predicted that a global increase of 2^0C above pre-industrial temperature levels can lead to the scarcity of water, a greater increase in heatwaves and expose millions of individuals to vector-borne diseases like malaria [2]. Global warming, which is the observed rise in the average temperature of the earth's surface is a plague we all must attempt to put a stop to. It is caused by the increased concentration of Carbon dioxide (CO_2) and other Greenhouse Gases (GHGs) in the atmosphere. These gases trap solar energy in the earth's atmosphere causing a slow but steady warming of the earth's surface. It is important to note that GHGs keep the earth in habitable temperatures [3]. Without these gases the earth's surface would be too cold for most plant and animal species. However, the increase in GHG emission caused by anthropogenic activities is where the problem lies. Research has shown that global temperatures have been increasing by about 0.25^0C per decade over the past 30 years. Mega cities such as Paris, Moscow and Houston are facing even greater increase in temperature [4]. Figure 1 presents the annual mean global temperatures from 1985 to 2015.

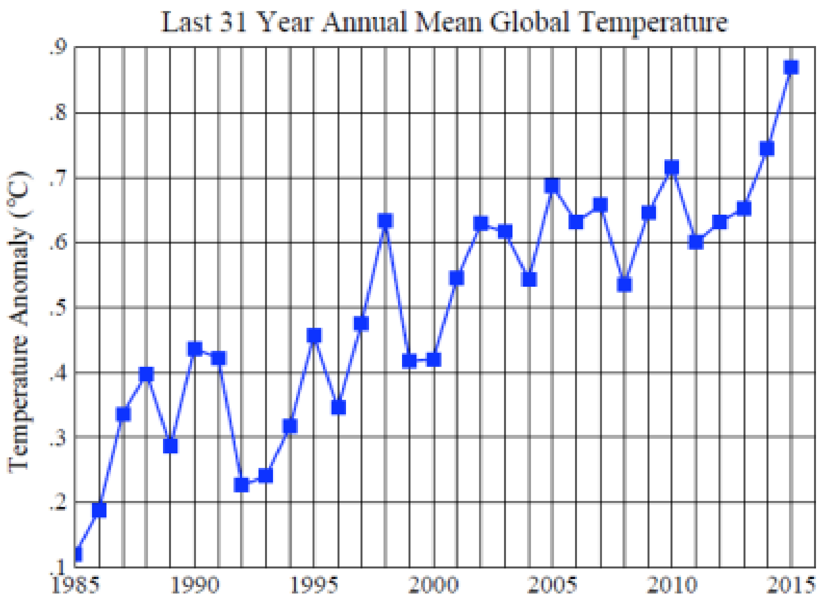

Figure 1. Year to Year climate change from 1985 to 2015 [5]

We see from Figure 1 that the average temperature for the 30-year period has experienced a 0.8^0C increase. If no attempts are made to slow down global warming, the effects could be devastating. Global warming like any other challenge the world is facing can only be properly handled if its true causes are known and understood. Only then can we be pro-active to this phenomenon. What we need to note is: even if all human activities responsible for global warming are stopped, the increase in temperature of the earth surface will not automatically stop. This is because global warming is not only caused by human activities; but also, by natural causes like volcanic eruptions [6] and the almost irreversible damage already caused by our CO_2 emissions. If all GHG emissions are stopped, it will take another 1000 years for the effects of climate change, caused by high CO_2 concentrations to be noticed [7].

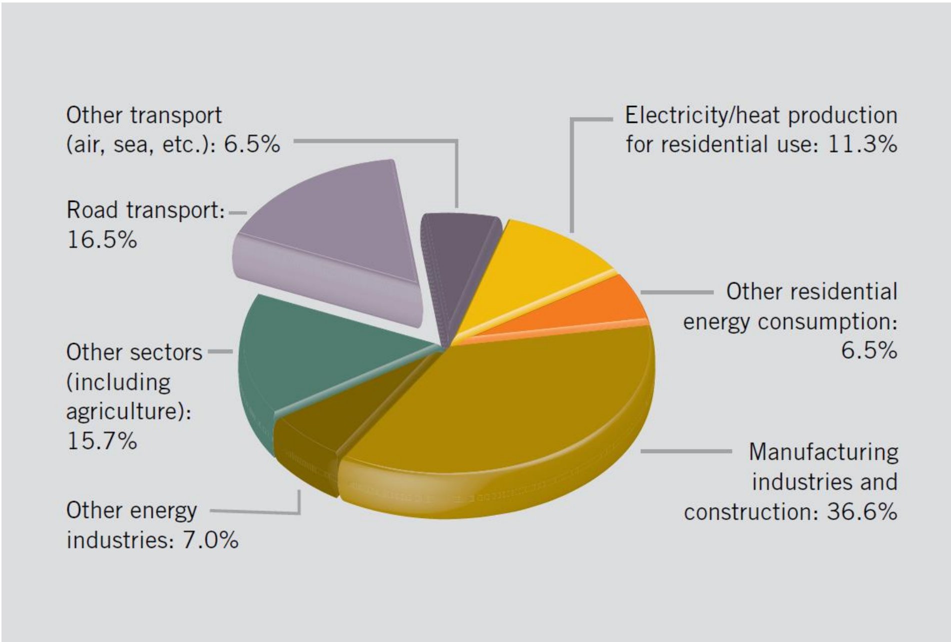

Figure 2. Carbon Dioxide Emission Sources [8]

One of the many sources of human induced global warming is transportation. The analysis in Figure 2 from 2011 highlights the transport sector as the second largest CO_2 emission sector. Within the transport sector, road transport dominates in CO_2 emissions. This sector's dominance in CO_2 emissions together with the effect it had on global mean temperatures as shown in a study done in 2009 [9] makes it relevant for my research.

Prominent human health challenges caused by inhaling harmful exhaust gases are other problems caused by land transportation. Hazardous pollutants from some car exhausts are similar to the compounds contained in cigarette smoke and coke oven emissions which are known to cause lung cancer [10].

1.2 Goal

The transport sector is the second largest contributor to CO_2 emissions as seen in Figure 2 and within this sector, road transport is the largest CO_2 contributor. The overall goal for this project is to address global warming by investigating ways in which harmful exhaust gas emissions can be reduced in the land transportation context, hence reducing GHG emissions. Health concerns related to road transport are addressed, though to a lesser extent.

For specificity, this thesis seeks to tackle global warming and GHG emissions from the road transport sector through an application which works on the nudge theory and aims to forester sustainable mobility. My focus is examining ways in which data analysis techniques can be used in a digital nudge ecosystem. I investigate the complexity of creating nudges and managing nudge personalization. I extend the concept of analysis by demonstrating how data analysis works in the functionality of such a system and I investigate the effects descriptive analysis has on finding patterns in user responses to nudges. This simplifies the nudging process and improves nudge precision.

My hypothesis is: Patterns can be formed from historical nudge data represented through descriptive analysis for a chosen test scenario like day, time or weather condition.

1.3 Problem Context

Road transport means for most cities are made of cars, motor bikes, bicycles, buses and pedestrians. The exhaust pipes of cars, motor bikes and buses give off several gases including CO_2, Particulate matter (PM), Hydrocarbons (HC), Nitrogen Oxides (NOx), Carbon Monoxide (CO) and Sulfur Dioxide (SO_2). Most of these gases are greenhouse gases which are great contributors to global warming and air pollution.

GHG emission from car exhausts is a global issue which will entail multiple solutions with each solution tailored to fit a target group. This thesis focuses on Tromsø and it involves proposing an application which seeks to promote a lifestyle change in its inhabitants towards sustainable mobility. The application needs data from several sources to function at optimum. The data needs to be collected, filtered, analyzed and used to create and personalize nudge scenarios to fit each person's needs.

1.4 Challenges

The proposed solution involves getting people to act upon an issue which is not considered as urgent by the public. This is a challenge because most people only take actions which have direct and immediate effects. Actions taken towards sustainability for the most part entails that people change their default travel methods to more sustainable forms. Changing defaults is a challenging task because it involves breaking a behavior and forming new habits. For this to happen, people must be interested in sustainability. Additionally, handling the heterogeneity of people is also a challenge because people have different needs and preferences. The process of personalizing scenarios for each individual could easily become complex when dealing with larger populations.

Influencing people to make sustainable transport choices involves sending accurate information and nudges. To achieve this, we need to know which data sources affect mobility, how to access the data, analyze it and make conclusions from it. Finding the best methods of analyzing heterogenous datasets, understanding what information the data carries and finding ways to channel this information into the application are issues I attempt to handle in this implementation.

With nudging, the specificity and timing of the message sent is crucial for the system to function at optimal. Knowing whom to nudge, what information the message should carry and when to nudge are relevant issues I face in the implementation of this application.

Knowing if the nudges achieved their set goal is also of relevance for evaluation. Setting up a system where the users can provide feedback will inform us if the application is useful to them. Additionally, this feedback gives us directions in which ways we can improve to reach our goals.

1.5 Proposed Solution and Approach

We propose a mobile application which guides people in their transportation choices. This application is meant to influence them towards more sustainable transportation types which will reduce the emission of GHGs from car exhausts which contribute to global warming. Users of the application get sustainable transportation nudges through push notifications.

I approach this proposed solution by doing data analysis which will help in forming the nudges sent to the users. For the application to work as intended, data from several sources is used. Data analysis is done by first filtering out the data which is not needed, then grouping data elements to create nudge scenarios which are used to create the nudges sent to the users through the application. Analysis is used to personalize each user’s experience. Personalization involves deciding when to nudge, how often to nudge, when to stop nudging for a scenario and how to modify the nudges. Descriptive analysis is the proposed analysis method used for application evaluation. Pattern mining techniques are used to help find similarities in users’ responses to nudges which affect how the nudges can be modified. Achieving this requires user interaction through the application, data collection, data management, data analysis, a constant update of the nudges sent and feedback from the users.

1.6 Results

Using descriptive analysis on nudge history data, patterns in user behavior can be realized over time which simplifies analysis. The representation of the users' nudge history on charts demonstrate how they perceive the nudges and help in user personalization.

A model was used in another test experiment which was aimed to determine how often to nudge based on previous nudge responses. This method can be applied to real nudge data and it helps in deciding if nudges should be sent.

1.7 Contribution

The contribution of this thesis is a web application which seeks to promote sustainable road transportation. It suggests that people use more sustainable forms of transportation like walking, cycling or taking the bus which helps in reducing congestion, reducing pollution and encourages people to be more active in the Tromsø municipality. I worked on implementing the front end of this application.

In this thesis I investigate ways in which data analysis is used in implementing an application for sustainable transportation and nudging. I use data analysis in decision making: deciding when to nudge, what type of nudges to send, how often to nudge a user, when to stop or change nudges for a user and how to personalize users' experience on the application. I demonstrate how analysis can be done on user feedback data, one technique which can be used to determine if to nudge a user or not based on nudge history, how to form text nudges using several data elements, how to evaluate the application and how to find patterns in nudge responses.

I demonstrate the complexity of nudge personalization based on user nudge history and their set preferences.

1.8 Limitations

The focus of this thesis was to demonstrate the importance of data analysis and its relevance in implementing an application for sustainable mobility. This implementation does not focus on privacy and security of the application.

1.9 Outline

The rest of this thesis is organized as follows:

Chapter 2 – Background: I give an extensive background of global warming, data analysis, greenhouse gases, sustainable transportation, nudging, big data, and data analysis. I explain these terms in relation to the context of this thesis; building the ground work and showing reasons for my choice in this topic. I also provide some past projects and studies related to my research and implementation; stating how I drew inspiration from these projects.

Chapter 3 – Architecture and Design: I provide the overall architecture of this thesis, the design of the application and the approach for analysis. I provide further details in the analysis architecture; highlighting the importance of analysis in such a project.

Chapter 4 – Implementation: It describes the main work done to achieve my goals for this project. I start off by giving details of all the tools and technologies used in the implementation. I move on to the analysis and the application implementations. For the analysis, I provide details of the data sources used, the type of analysis done and the way this data is represented. I also demonstrate how I form, present and handle nudge personalization for users. This is followed by the details of the main application components.

Chapter 5 – Experimentation and Evaluation: It contains demonstrative experiments I carried out by showing how patterns can be found using descriptive analysis. Another experiment was carried out suggesting a method which can be used in knowing how often to nudge users.

Chapter 6 – Discussion: It contains discussion of results and how they can be interpreted. It also gives some decision choices and considerations which were made during the implementation of this thesis.

Chapter 7 – Future Work: It provides some work on data analysis and the application I wish to implement in the future to provide continuity for this thesis.

Chapter 8 – Conclusion: I highlight the relevant results and discuss how they tie to my hypothesis.

Chapter 9 – References: It provides all the references and citations used.

2 Background

2.1 Global Warming

In 1896, a Swedish scientist discovered the term global warming as a theoretical concept [11] which was criticized by several other scientists. In the 1950s, some American scientists discovered global warming as a possibility, which might perhaps come to pass in the remote future and by 2001, thousands of scientists discovered global warming as a phenomenon which had started to affect the weather and could become worse [12]. At the advent of the industrial revolution in the 18th century in Great Britain, production gradually migrated from handcraft means to one dominated by industry and machine manufacturing due to an increase in world population and a quest for better living standards [13]. Energy consumption increased drastically due to the use of machines and the main energy sources used to power these machines were fossil fuels. The by-products from burning fuels contributed to environmental problems, especially global warming. Over the years, the topic of global warming has grown exponentially, sparking discussion and research in schools, research groups, governments, non- governmental organizations and individuals. Global warming is the increase in average temperature of the earth's atmosphere because heat from the sun is being trapped between the earth's surface and the atmosphere and is not being radiated back into space. The increase of global temperatures is threatening the environment, impacting human health and challenging political stability [14]. Global warming results in the melting of sea ice and glaciers, increase in extreme weather event such as hurricanes, flash floods, tornadoes, heatwaves, ocean acidification and the list goes on. These environmental changes are followed by diseases spreading at higher rates. Additionally, global warming puts pressure on plant and animal species to migrate [15] and some species are fearing extinction [16] . The causes of global warming are known to be linked to the increase of GHG concentration in the atmosphere. GHGs are released in the atmosphere due to human activities like fossil fuel combustion and natural causes like volcanic eruptions and wildfires. Global warming and the Greenhouse effect raise so many questions and concerns regarding the state of the earth, followed by international panels trying to find solutions to reduce the impact and preserve our planet. Identifying that GHGs are not necessarily the problem; but an over production of them, the question of how much GHGs is a safe amount has been a topic up for debate for years with suggestions and arguments raised. [17] suggests that we need to reduce emissions by one trillion metric tons of Carbon by 2050 to avoid catastrophic climate changes. It further states that there is a projected rise of 2^0C in earth's temperature if the carbon release limit is exceeded. This only mentions CO_2 but excludes other GHGs like NOx, CH_4 and HCFCs which contribute between 10 to 40 percent of global warming. The Kyoto protocol [18], adopted in

December 1997, took effect in February 2015 involving 192 countries. The emphasis of this protocol lays on fixing limits of the involved countries' annual emission of GHGs such as CO_2, CH_4, NOx, HFCs, PFCs, and SF_6. In Norway, an alternative solution of Carbon Capture and Storage is implemented where they attempt to counter CO_2 emissions in the atmosphere by injecting liquified CO_2 1000 meters below the North Sea bed where it will remain for thousands of years [19]. The responsibility of controlling GHG emission is placed on governments and the industrial sectors, but household activities are also great contributors to GHG emissions. Sadly, we cannot reverse the effects of climate change, but we can change our habits to reduce our emissions to prevent extreme case scenarios.

Greenhouse gases (GHG)

GHGs mainly include CO_2, CH_4, NOx and Fluorinated gases (HCFCs, PFCs, SF_6 and NF_3). They are mainly by-products of fossil fuel (coal, natural gas and oil) combustion, decay of organic waste and agricultural activities. The production of CO_2 experienced an increase by 150,000 million tons in the past half century which has altered the carbon cycle greatly [20]. The carbon cycle is a sequence of events which makes earth habitable by plants and animals. It describes the production, use and recycling of carbon in the atmosphere. Carbon is found in all living things, in the atmosphere and in the oceans. CO_2 is absorbed by plants for their growth which helps to reduce its concentration in the atmosphere hence balancing the carbon cycle. When plants and animals die, they may be converted to fossil fuels like coal and oil which when burned, release CO_2 back into the atmosphere. Due to abundant deforestation, the amount of CO_2 being produced from fossil fuel combustion is too much for the plants left to absorb causing an imbalance in the carbon cycle. The oceans also play an important role in CO_2 absorption but its concentration in the oceans has led to ocean acidification [21]. Fuels are used by households mainly for heating/cooling, electricity generation, transportation and their uses make them hard to avoid.

The fear of an irreversible damage of the ozone layer and the constant increase in the production of GHGs sparked tremendous research and brought about the slogan "Going Green" which are lifestyle changes to address an interrelated set of environmental problems [22]. All efforts made to reduce the emissions of GHGs in various domains like transportation, computing, construction, home, commercial and industrial will reduce the effects of global warming.

2.2 Land transportation and its contribution to global warming

Transportation, which is a fundamental aspect of economic and social growth has become more affordable and more effective with technological innovations over the years. With the focus of this thesis being land transportation, varying means of road transportation have been adopted as time has evolved like the use of horses, walking and driving. Congestion, traffic jams and pollution are just a few of the hazards that come with land transportation. For example, a traffic hold up at one location can be a major bottleneck for large scale transport networks which could incur life threatening and economic damages. Transportation accounts for one third of GHG emissions, two thirds of oil consumption and about half of urban air pollution in the US [23]. The GHG emissions from transport are growing faster than emissions from any other sector.

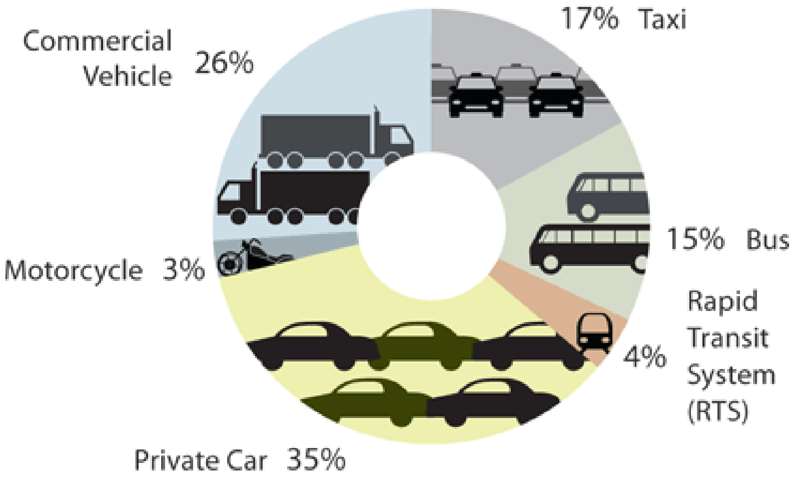

Figure 3. Sectors in land transportation [24]

Within land transportation, several sectors exist, and Figure 3 portrays various transport sectors and how much they are used. These transport types and the percentages used vary between countries and some countries may use more sustainable forms of transportation than others. The differences can be caused by government bans and policies, geographical, cultural or adaptation reasons. Figure 3 shows that private car ownership and use dominates all other sectors in the land transportation domain; followed by commercial vehicles and taxis. The issue of sustainable road transport cannot be properly handled without addressing these sectors.

Green transportation is any eco- friendly transit alternative to driving alone. It is important to note that finding more sustainable means of transportation should not affect the quality of transportation or inconvenience of the travelers. Finding the balance between green transportation and the comfort of the travelers is key and will be a motivation for most travelers to choose to go green.

People choose how to travel based on several factors such as weather, comfort, convenience, finance, distance and travel time. Taking a bike would not be as convenient as commuting by car or taxi; especially in poor weather conditions. Taking a bus however is more environmentally friendly than driving but the bus schedules must match the schedules of travelers for bus rides to be considered an option. More environmentally friendly transportation choices in the long run mean less emission of harmful exhaust gases and pollutants. The gases in the atmosphere cause two types of pollution categorized by primary or secondary sources. Primary pollution caused directly from a primary pollutant gas can impact plant and animal life, like NOx which cause acid rain [25]. Secondary pollution on the other hand is caused when a primary pollutant reacts with molecules in the atmosphere. PM is a secondary pollutant which can cause lung cancer [26] .

2.3 Sustainable Land Transportation

Sustainable Transportation stemmed from the term "Sustainable Mobility [27]" which is *the development that meets the needs of the present without compromising the ability of future generations to meet their own needs*. In the case of transportation, efforts need to be made to prevent and contain the negative impacts caused by transport systems. This is essential to ensure that future generations experience a healthy transportation environment as well. In Europe, 47% of NOx, 39% of VOCs and 66% of CO emissions are produced by motor vehicles [28] while in the US, road dust accounts for over 40% of the particulate emissions [29]. While these numbers fluctuate between countries, several nations are taking steps to control and reduce emissions. However, a lot still needs to be done in this domain for it to become sustainable.

So much work and research has been done in attempts to implement and maintain sustainable means of transportation in different parts of the world [1, 30-33]. [30] provides a carpooling system where people going to similar destinations could share the same car; hence reducing congestion, exhaust gases and other pollutants. [31] suggests the creation of more cycling and pedestrian routes as opposed to car routes. They go as far as implementing taxes, paid parking and tolls for car owners with attempts to make people use more public transportation. Furthermore, they suggest a system where governments can close passages into some cities by private cars after a set pollution limit is reached. [32] suggests that if buses are given priority in traffic, then it can influence people to use buses more than driving. A mobile tool was created and tested in [33] which told users how their currently sensed transport means affected the environment with attempts to make them use more sustainable transport means. Another attempt toward sustainable transportation was the creation of Electric Vehicles (EV) which was aimed to replace Internal Combustion Engines (ICE).

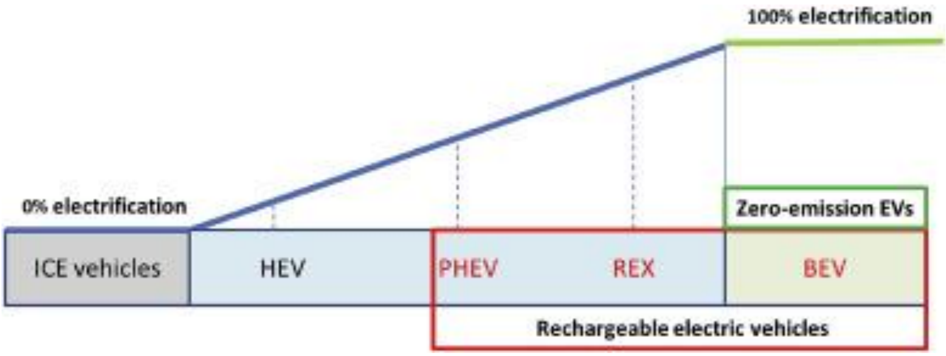

Figure 4. Level of Electrification of Electric Vehicles [1]

EVs are in 4 categories as seen in Figure 4: Battery Electric Vehicles (BEV), Hybrid Electric Vehicles (HEV), Plug-in Hybrid Electric Vehicles (PHEV) and Range extenders (REX). Only pure BEVs are zero emission EVs. Most of the market is dominated by HEVs which cause little or no emission reduction. EVs come with some challenges such as higher costs, lower operating range, and limited charging infrastructure which makes them less widely used. The total benefits of zero emissions by EVs can only be reached by using BEVs which are powered by sustainable energy sources [1].

The duty of maintaining a sustainable environment should be on all citizens. Raising awareness of the situation, informing people of the effects that their transportation choices have on biodiversity and on the environment while presenting sustainable alternatives is a good approach to get people interested in sustainable transportation. Information affects choice and decision-making [34]. People make choices based on the alternative pieces of information they have. In the light of sustainable transportation, presenting people with more sustainable transportation options and educating them on the effects of each transportation option can influence their transport choices.

We investigate a system of information and digital nudging for more sustainable forms of transportation. Transportation is a mobile activity, so digital nudging through a mobile application fits our need. Nudging for this purpose needs to be smart, adaptive and personalized because the factors which affect user transportation choices are independent to each person and they change quickly.

With the target location for our Information and nudge system being Tromsø, we need to know the transport types which exist, how sustainable each transport type is, and the reasons why people choose certain transport types over others. With this information, we can come up with suggestions on how to present information to users, influencing them to take more sustainable transportation choices. The most common transport types in Tromsø are walking, cycling, bus, carpooling, driving and taking taxis. These can be classified by their Environmental Friendliness (EF) as shown in Table 1. EF indicates how the transportation type affects the environment; with a high EF indicating that the transportation type is more sustainable.

Table 1. Environmental Friendliness of various forms of transportation[35]

Type	EF	Discouraging factors	Encouraging factors
Car		Economy (toll, parking, gas), traffic jam	Convenience
Carpool		Inconvenience, traffic jam	Economy, social
Bus		Schedule, traffic jam	Economy, priority in traffic
Bike		Time, effort, exposed to conditions	Economy, health
Walk		Time, effort, exposed to conditions	Economy, health

Table 1 gives us the EF of various forms of transportation, reasons why people will want to use them (encouraging factors) and reasons why people will avoid them (discouraging factors). We see from Table 1 that driving a car alone; equivalent to taking a taxi alone has the lowest EF. Our goal is getting people to use transportation types with higher EFs through our nudge application which does not involve any bans, fines or any other forceful means.

2.4 Nudging

One suggested means of influencing people to make choices which is applicable to the context of green transportation is through nudging. Nudging is a term which was first described in the book 'Nudge: Improving decisions about Health, Wealth, and Happiness' [36] by Richard H. Thaler and Cass R. Sunstein. It is defined as: *Any aspect of the choice architecture that alters people's behavior in a predictable way without forbidden any options or significantly changing their economic incentives.* Nudging is a psychological term which cuts across several factors like economy, politics and behavioral science and the end goal usually is to influence people's decisions for the better. [36] further explains that *a nudge is not a mandate and the intervention must be easy and cheap to avoid.* Traditional mandates like prohibition of public smoking are not nudges but they can be used together to achieve a set goal. Nudges are typically suggestions which are aimed to influence the behavior of people knowingly or unknowingly through presentation. The benefits of nudging are usually for the greater good of the public but could be beneficial to individuals in the long run.

The concept of nudging has been examined in the past in the following contexts:

One implementation of nudging was intended for guiding students towards healthier meal choices in the lunch room [37]. They did this test by strategically placing the healthier food items like fruits and flavored milk in more accessible and visible positions and the less healthy options were placed at the back. At the end of the survey, sales of healthier food items increased by 18% and the consumption amount of less healthy foods decreased by 28%.

[38] gives us another method used by governments to make farmers use more fertilizers to improve yield. They did this by subsidizing costs of fertilizers. This decision affected the government's GDP, but it helped improve the agricultural yield of the farmers.

[39] experimented on ways to make smokers quit by forming reward programs where they were granted some money if they did not smoke for a set time. Financial benefits played a huge role in this experiment because the returns were immediate. When nudging benefits are quick to notice, the responses are quicker and more positive.

Expanding on the definition of nudging, it should not affect people's economy. An increase in the prices of cigarettes, for instance, because they are considered unhealthy is not considered nudging. However, a warning can be placed on the cigarette packet telling people about its disadvantages and the implications of their use of these products. This option is considered a nudge.

In nudging, there are choice architects (those issuing the nudges and creating the environment for the nudge), the choice environment (the topic or ecosystem of the nudge) and the people being nudged. Though the original idea of nudging was coined to be for promoting positive habits, arguments about nudging being manipulative have been raised. The choice architects could have underlined motives which could be personal, financial or political, manipulating the choice environment and the people being nudged. For nudges to be less manipulative, they must be transparent. Transparency here means those who are being nudged must know all the choices available to them, the consequences of each choice and they must have the freedom to choose. Shlomo Cohen in his article '*Nudging and the Informed consent*' for nudging patients [40] stipulates that autonomy of the people being nudged should be respected. He goes along to argue that nudges are ethically permissible even if they take advantage of the patient as long as they have not been deceived and the impact of the nudge on the patient and the environment is positive. Here, he highlights nudge transparency at the expense of manipulation and personalization. This has been argued by [41] who point out the aspect of manipulation; even with the patient's consent as being unethical. They also point out direct and indirect consequences to nudges and argue that whether direct or indirect, manipulation for nudging is wrong. [42] goes ahead to focus on the ethical guarantees which [40] seems to ignore when they assume that the physician's professional judgement will guard against abuse. Here, they emphasize on the use of nudging as a last resort; where there is evidence that nudging is effective in this context, there is evidence of public misinformation leading to an unhealthy lifestyle, and the patient must consent to it. Transparency of nudges raises the problem of effectiveness: how effective can nudges be if people know they are being nudged? In [43] an experiment was carried out to test the effectiveness of transparent nudges. They nudged contributions to carbon emission reduction; where a default value was set, and they exposed transparency to varying levels. By letting the users know how their actions affect the default value and the purpose of the experiment, they prove that transparency does not reduce effectiveness of nudging. [44] gives more insight on nudging but warns against bad nudges. We see in [37] that the manner in which items are displayed can affect consumption. This can be used for the bad as well; like exposing alcohol in shops could increase alcohol consumption. Generally, nudging for more positive outcomes could be the desired intention but knowing how, whom and when to nudge is crucial for the nudges to be effective.

Personalization of people or groups greatly affects how nudges are sent and perceived. Care must be taken in forming and issuing nudges to suit specific groups of people. One method could work for a person or a group of people but not another. So, there is a need to understand the needs and behavior of individuals or groups before nudging them. If the nudge meets a need, then the reaction to it would likely be positive. If the behavior of people from past incidents is studied, then insight is gained about when and how to issue the nudge. Forming nudges to have positive effects on those being nudged is the challenge.

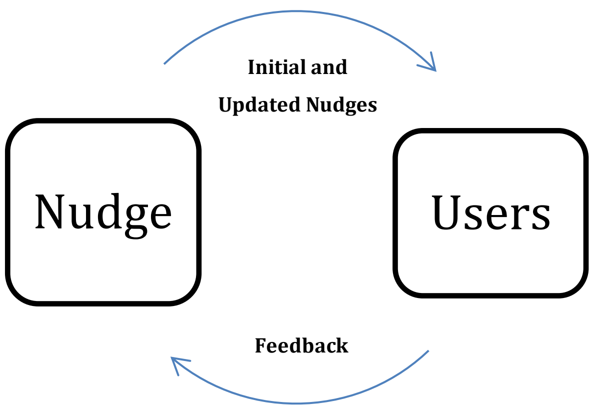

Figure 5.The Nudge Life Cycle

Figure 5 represents the life cycle of a nudge. The nudge is sent to people and their feedback (reaction to the nudge) is collected and analyzed. The nudge is then updated as the user's feedback suggests, and the nudge is resent to the users. This process is repeated but care must be taken so the initial goals of the nudge are not missed.

Digital Nudging

Digital nudging is the use of user-interface design elements to guide people's behavior in digital choice environments that require people to make decisions [45]. In the Digital nudging concept, the choice architects are the software/system designers. Users are presented with some information and a possible nudge through a user interface and they make a choice based on the information they receive. The nudges could come in form of questions, images, text, audio, video or a blend of a few of these technologies. The information presented to the users should be strategic with the goal of the nudge in mind and the users should make choices based on the information they have received. The most common ways to issue digital nudges are through mobile phones. This is because the number of smartphone users is on an increase and 1 out of 4 people own at least 1 smartphone in developed countries [46]. Smartphones however are not the only means by which digital nudging can be done: displaying nudges on public screens, sending emails, sending nudges on the internet which can be accessed by desktops and personal computers usually through pop-ups are other means by which digital nudging can be done. With digital nudging, the user interface plays a huge part in how the nudges will be received. [47] gives us a clear understanding of how presentation affects choice with the saying *'what is chosen often depends upon how the choice is presented'* which extends the concept of the experiment on presentation in [37], but in a digital environment. When presenting a digital nudge, the design must be well thought through and engaging to get and keep the attention of the users. The target audience affects the design as well: an older audience will prefer a simpler design; with bigger and easy to find buttons while a youthful audience will prefer style. However, it is important that the application is easy to use and is engaging regardless of the audience. With digital nudging, personalization can be implemented by getting feedback from the users; by tracking their reactions to the choices presented to them and responding accordingly.

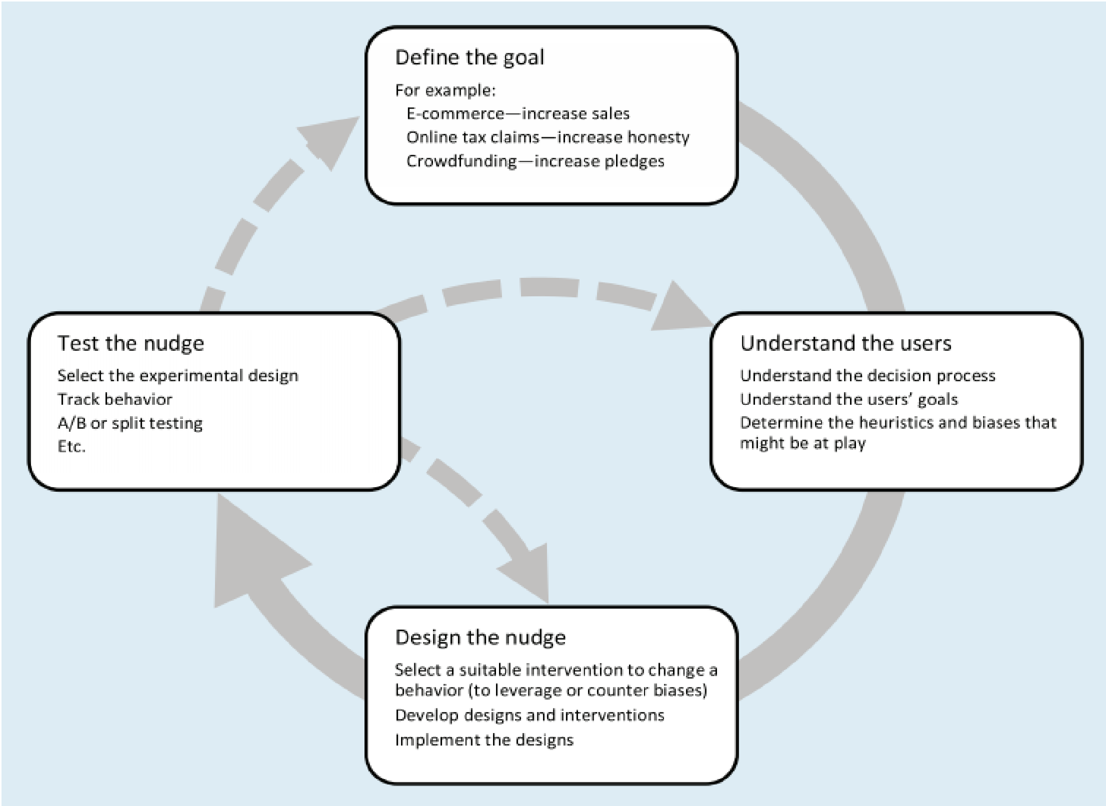

Figure 6. Digital Nudging life cycle [47]

Figure 6 from [47] explains the life cycle of a digital nudge made up of 4 steps: Defining the goal, Understanding the users, Designing the nudge and Testing the nudge.

Step 1 - Defining a clear goal: A good idea of the nudge and its intention should be considered in this step. The ethical implications are to be well thought through and target audience should be kept in mind.

Step 2 – Understanding the users: Knowing the goals of the target audience and how they will benefit from the nudge is important because this will determine if and how they will respond to it. This affects the way the nudge will be designed.

Step 3 – Design the nudge: With clearly defined goals, an understanding of users and their needs in mind, the nudge can be designed. The design of the nudge should be flexible, easy to use, meet the initial goals and be adaptable by the users.

Step 4 – Test the nudge: When the nudge is designed, it must be tested. A/B testing; which is a process used in marketing to isolate and test factors that affect performance of marketing effort [48] is one way to test nudges. Different versions of the same nudge can be sent to different users and their behavior is studied. Feedback from the nudge versions are monitored and the nudge version which receives the most desired feedback gets chosen and sent to all the other users usually through an update. Tracking user behavior directly is another way to go about it; which ensures responding to each users' feedback personally thereby implementing personalization, but it is a tedious process for systems with multiple users.

Just like the nudge life cycle, the digital nudge life cycle entails constant feedback from the users and updates. This is because user demands and preferences are constantly changing, and nudges should change to adapt to these behavioral changes. However, the initial goals of the nudges should not be put aside; but nudges should accommodate change; with the goal always in hind side. Personalization in digital nudges can be implemented by creating user profiles and updating them as time goes by and as the users' preferences change.

An example of digital nudging is seen and explained in [49]. A digital Photo Display System (PDS) was created and installed in homes on users' personal devices. Users chose their favorite pictures and the PDS displayed these images, slowing skipping through them and providing animation effects. The reason for this study was to put the users of the applications in serendipity; each time they see images of past pleasurable moments. This is an example of photo nudging, with extra animation effects like sound.

In another example of digital nudging, experiments were carried out on a used online platform where pop ups were sent to its users with requests to share the web page on other online platforms [50]. The nudge messages were classified into: simple requests, monetary incentives, relational capital and cognitive capital. They concluded that nudges with additional incentives like led to improved sharing while simple nudge requests did not.

These experiments demonstrate that the design, presentation of information and additional incentives are important for the success of nudges.

2.5 Big Data

Big Data is a large set of data from varying sources, made of different structures and is often disorganized. The advances in science and information technology has made the generation of data an easy and fast process. In 2011, the world's data volume was at 1.8ZB (1 Zettabyte= 1e+21 Bytes) and there were predictions of this figure doubling at least every other two years [51]. Social networking and digital marketing are great contributors to data explosion. Averagely 72 hours of videos are uploaded on YouTube every minute and Facebook generates 10PB (1 Petabyte = 1e+15 Bytes) in log data monthly [52].

The details of the definition of Big Data can be broken down into the seven V's: volume ,velocity, variety, veracity, validity, volatility and value [53].

i. Volume: The size of the data is big and usually created from multiple sources including text, audio, video and social networking.
ii. Velocity: The speed with which the data is produced is high, making it harder for traditional analysis methods to be effective.
iii. Variety: The data is of multiple types and sources like audio and video. Data can be structured or unstructured adding complexity to the analysis.
iv. Veracity: Veracity deals with data correctness: how certain are we about the data and how meaningful is the data.
v. Validity: Validity refers to the correctness of the data with respect to the intended use. This explains the use of data based on how it is understood. The same set of data may be valid for one application but invalid in the next.
vi. Volatility: This refers to the rate at which stored data changes over time and how long it is kept (retention period) before it is destroyed.
vii. Value: This is the desired outcome of the data being processed. It describes the quantifiable value which can be gotten by those keeping and using the data. For data to be valuable, it must exceed its cost.

Big Data Classification

The big data value chain is the series of events from big data generation to its usage. This chain is composed of 4 steps including: Data Generation, Acquisition, Storage and Processing (Analysis) [54] which can be broken down into smaller parts as seen in Figure 7.

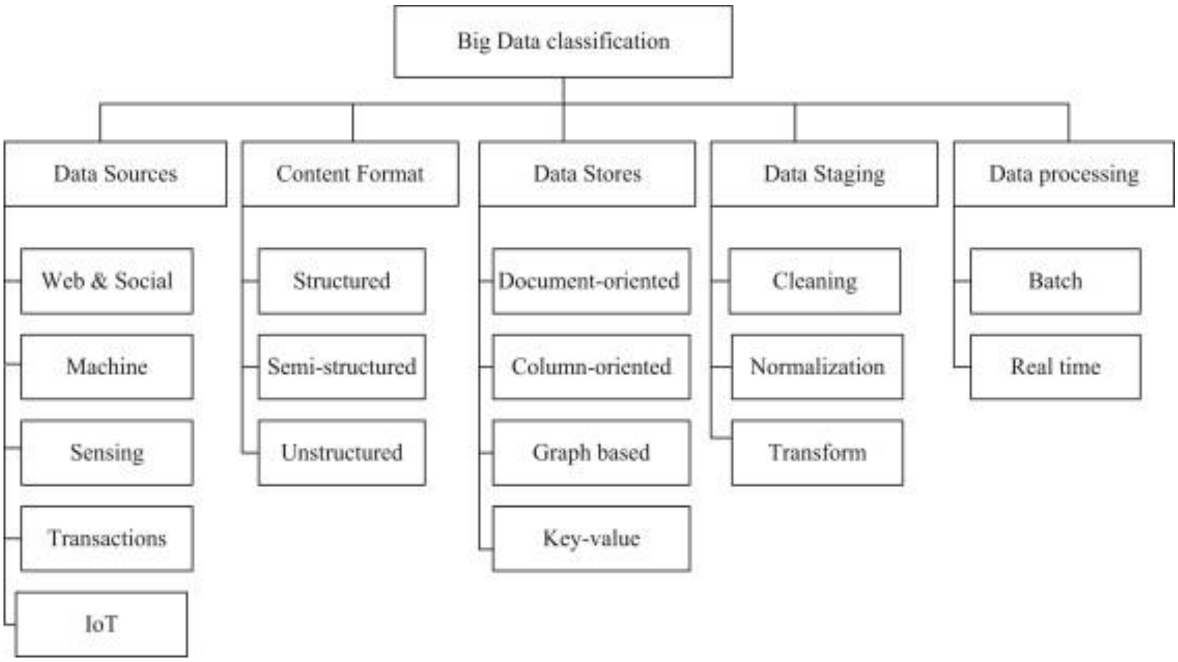

Figure 7. Big data classification [55]

Figure 7 presents a summary of big data with respect to its sources, format, data stores, staging and processing. The analysis and storage steps are detailed in the rest of this chapter. Data staging is also described, though to a lesser extent. Some big data sources include sensors, medical devices, blogs and financial data. The disparity in the data generating sources brought about data of different formats. Some data is structured, semi structured or unstructured leading to a shift in storage from SQL databases to NoSQL databases and recent adaptations made for NewSQL databases. The most common NoSQL database types include: Key-Value databases like Redis, Column-oriented databases like Cassandra, Document databases like MongoDB, and Graph Databases like Neo4j. They are all NoSQL databases but store data in different formats. For any processing to be done on data, it must go through a pre-processing (staging) phase where it is cleaned to remove incomplete and unreasonable data. It is then transformed to a suitable form for analysis and normalized (structured) to reduce redundancy. After preprocessing, the data is analyzed. Processing is done based on the type of data collected. Batch data is data which has been collected and stored while stream data is data which is being developed constantly.

Data Analytics

Data analytics (data mining, data analysis or data processing) which is the last stage of the big data value chain is the part of data science which aims at examining datasets to draw conclusions about the information they contain. Analytics is done through software [54]. Data analytics aims to extract useful information from several data sources, suggest conclusions and/or support decision making. Data collected and used by companies is changing rapidly and Analytics helps discover what has changed and how they can react to the change which helps them increase business value [56]. Most data generated in recent times have big data characteristics (huge size, varying structures, from different sources and can potentially provide value) so big data analytics is required. Big data analytics is the process where advanced analytic techniques operate on big data [56]. Several organizations and companies handle different forms of data daily. While some companies are fully operating on big data and using forms of advanced analytics, some others are not.

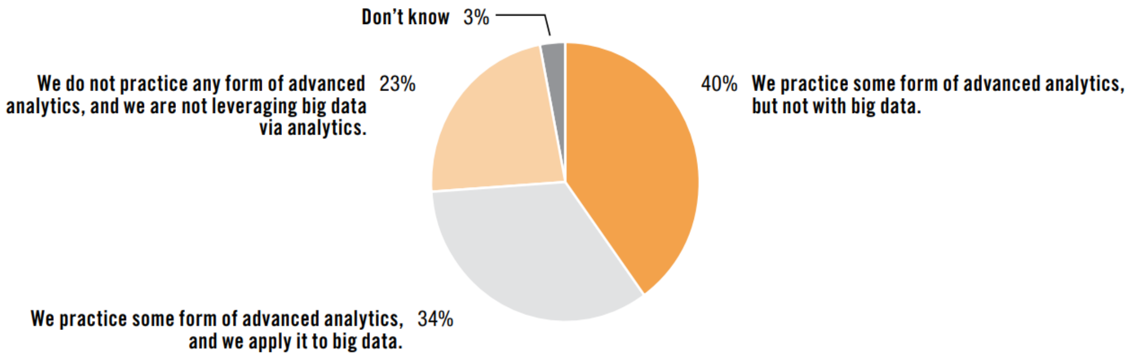

Figure 8. Survey on companies practicing advanced analytics on big data [56]

Figure 8 gives the results of a survey carried out on companies; finding out if they use big data and big data analytic methods. We see that most companies use various forms of advanced analytics but do not necessarily use big data. On the other hand, some companies are neither aware of big data nor advanced analytics. In attempts to find what the benefits of big data analytics are, another survey was carried out with the aim of finding the reasons why companies use advanced analytics. The most common responses involved customer lock-in (better targeted social-influencer marketing, customer-base segmentation and recognition of sales and market opportunities), business intelligence (acquiring of business insights, understanding business change and identification of root causes of cost) and scientific applications (fraud detection, risk quantification and market sentiment trending) [56]. Conclusions gotten from data analytics aim at adding value to the organizations doing the analysis.

Types of Data Analytics

Four types of analytics methods exist and they are: Descriptive, Exploratory (Discovery), Predictive and Prescriptive [57].

i. Descriptive analysis: This tells what happened in the past and presents it in forms like bar charts, pie charts or maps which are easily understood.
ii. Exploratory or Discovery analysis: This method of analysis finds unexpected relationships among parameters in collections of big data. Analyzing data from different sources together can offer insights which individual datasets may lack. An example is how customer feedback affects subscription rate.
iii. Predictive analysis: It is aimed at extracting information from existing datasets to predict future outcomes and trends. The tools used in predictive analysis involve statistical methods, neural networks and machine learning algorithms. Its main use is in marketing; understanding user preferences and adjusting the market to follow suit.
iv. Prescriptive analysis: Prescriptive analysis tells us what to do to achieve a goal. An example is airline pricing to maximize profit.

Tools and techniques for Data analytics

Picking a tool for data analysis is based on the type of data to be analyzed and the desired outcome. Batch data processing is analysis which is done on data which has been collected over time and stored. All the data points in the datasets are analyzed and it could take hours or days to analyze an entire dataset depending on how big it is. Stream or real time data is analyzed as the data arrives. Only approximate results are achieved because the analysis is to be done in real time.

With the processing speed and storage capacities of modern-day computers, parallel processing of data using thousands of inexpensive commodity processors is suitable for analyzing big data [57]. The data is split up, analyzed by several machines and the results are combined, from which conclusions are made. Most big data is stored in data warehouses and can be accessed and analyzed through online platforms. Data warehousing is a process of integrating enterprise-wide corporate data into a single repository, from which end users can easily run queries, make reports and perform analysis [58]. The data in the warehouse can be mined and some data mining techniques are: classification, clustering, pattern mining and outlier detection [59].

i. Classification: Data classification is the process of analyzing data and organizing it in relevant categories for easy location and retrieval. It helps identify sensitive files, optimize storage by eliminating duplicates, discover patterns and trends in data, enhances security and speedup the search process using data indexing.

ii. Clustering: Data clustering is the process of grouping a set of data objects based on their characteristics. Objects in one group must differ in characteristics from objects in another group.
iii. Pattern mining: It is the process of discovering statistically relevant patterns in datasets involving machine learning and statistics techniques.
iv. Outlier detection: It is a process of detecting and removing outliers from datasets. Outliers in data context are data objects which differ drastically from the other data points in a dataset.

OLAP

OnLine Analytical Processing is a computing method that enables users to easily and selectively extract and query data in order to analyze it from different points of view [60]. It allows users to analyze database information from multiple database systems and from multiple perspectives at the same time which helps to provide business intelligence. For OLAP to work, data is collected from multiple sources, stored in a data warehouse, cleansed and organized into data cubes. Each cube is composed of data of similar dimensions like geographic region, age, and time. Each dimension is populated with related data. The types of OLAP operations which can be done on data cubes are:

i. Roll-up: This provides a summary of all the data in a dimension.
ii. Drill-down: This provides a more detailed look at each dimension.
iii. Slice: This provides one level of information which can be displayed.
iv. Dice: Data from multiple dimensions can be analyzed.
v. Pivot: A new view of data can be realized by rotating the cube's data axes.

OLAP software can locate an intersection of dimensions such as: All products sold in a certain region above a certain price during a time frame. This exposes the details of the data held in the warehouse. Descriptive and Exploratory analysis techniques can be used for visualization purposes and Prescriptive analysis can be done to provide solutions for achieving set goals.

Data Staging

Before data is analyzed, it must be staged. Staging involves cleaning the data, normalizing it and transforming it to forms suitable for analysis.

i. Data cleaning: This deals with removing errors, unreasonable, duplicated and inconsistent data to improve data quality [61] .

ii. Data normalization: Data normalization is a process of decomposing tables in a relational database to reduce redundancy and improve data integrity [62].

iii. Data transformation: It involves converting data from one format to another which makes it easier for analysis [55].

Data Storage

Relational Databases

Before data explosion and big data, data was structured and stored in relational databases. Relational Database Management Systems (RDBMS) are a set of programs which enable the creation, update, administration and interaction of relational databases. The Structured Query Language (SQL) is the standard query language used to manipulate data in relational databases; which led them to be called SQL databases. Data manipulation by SQL entails inserting, deleting, querying and changing the data and relationships in the database [63].

Early relational databases were all kept and run on one machine which ensured the data was always in a consistent state. With the emergence of big data, scalability became an issue because commodity hardware often had to be replaced to fit new storage needs [64]. The scalability issues and the growth in partially structured and unstructured data made relational databases unfeasible to work with and new methods of storing data were sought out.

NoSQL databases

The explosion in data brought about storage problems, privacy concerns, scalability of existing databases, analysis concerns and heterogeneity considerations. The need to handle concurrent transactions to the same data resource with very low latency became a major concern [65]. NoSQL which means Not Only SQL databases are storage alternatives to relational databases. They were built to manage concurrency, scalability and the unstructured nature of present data in more optimal ways than relational databases [64]. They are usually non-relational, distributed and scale horizontally. Non-relational means the data in the databases do not necessarily need to be stored on tables; though some NoSQL databases have a table-like structure. Distributed implies the data in NoSQL databases is stored and managed in different machines which eases replication, access and concurrent transactions. They were built to scale horizontally; which addresses the vertical scalability problems in relational databases. Horizontal scalability here implies connecting multiple hardware and software to work as a single logical (cluster-like) unit where more affordable hardware and software can be added or removed as the demand arises [64]. Horizontal distribution involves dividing computation into tasks which are concurrently processed; decreasing processing time and reducing transaction latency.

The distribution in NoSQL databases brings about lower latency but raises consistency concerns. Data usually is replicated and stored in multiple machines, with each machine managing the data it stores. If a transaction is made to one machine, this transaction must be updated on all machines holding copies of the data for consistency to be achieved, else the data will be in an inconsistent state. Concerns arise if there is a network partition between machines holding the data. Data can be accessed on both sides of the partition at the same time which leaves the database in an inconsistent state. This is one of the downsides of data distribution as explained in the CAP (Consistency, Availability, Partition tolerance) theorem [66].

The CAP theorem states that, in the presence of a network partition, a distributed database either guarantees consistency or availability. Users choose databases depending on their needs and priorities.

NewSQL

Even though NoSQL databases have proven to scale to optimal, there are dissatisfactions with the lack of ACID guarantees in NoSQL databases and the lack of a standard query language like SQL. Modern Relational databases which seek to provide NoSQL's scalability benefits and maintain SQL's ACID guarantees are called NewSQL databases [67]. Though they are criticized for not bringing anything new to the database world, they exploit SQL and NoSQL technologies which were formally implemented one-at-a-time.

Cloud Storage

Organizations which run huge data operations usually come across issues such as infrastructure, flexibility and availability problems. Cloud computing serves to hold and manage companies data while the companies handle key business functions. Cloud computing promises reliable software and hardware delivered over the internet and remote data centers [55]. Computation and storage of big data can be done on the cloud and accessed remotely which eliminates the need to own and maintain hardware and software needed for data processing and management. This makes it cost saving for companies using cloud services. Amazon and Google are cloud service providers [66].

2.6 Related Work

This section describes theoretical and implemented work which relates to the context of this thesis.

Green Transportation choices with IoT and Smart Nudging [35]

This paper gives insight of nudging for green transportation from data sensing to the actual nudges being sent. It presents the stages in a clear and concise manner, highlighting challenges that may be faced along the way and suggests methods which can be followed to successfully nudge for green transportation.

UbiGreen [33]

This is a very relevant study which inspired our approach to the topic of Green Transportation and some of the design choices we made throughout our implementation. They built an application called the UbiGreen Transportation display where they used external sensors to track the user's travel method and based on the user's choice of travel, they provided active feedback presenting how their commute choices affect the environment. This feedback was represented through the screen of the user's mobile phone where a participation in more sustainable transportation types led to a display either by a greener fruit tree (tree progression) or an eco-system full of polar bears, fish and seals (polar bear progression) and a lack of sustainable mobility participation displayed either a sparse tree or a lonely polar bear standing on a thin sheet of ice. Their implementation focused on immediate feedback through a change in the user's wallpaper display.

Mobility Choices app [68]

The Mobility Choices application was realized because of the Mobility Choices project. They built an application where the users create profiles, state their preferences based on health, environment, cost, time, their preferred means of transportation, the number of transfers, waiting time and maximum walkable distance. They use this user input data to find the best possible green transportation route to propose to the users. The users update their preferences and the application adjusts based on the updated preferences.

A better day the 100 way [69]

This is a tool which helps to improve several aspects of human life. It focuses on improving sustainability for nutrition, mobility and living. Their concept is using a 100-point daily budget which should be sustainable for all humans, and if exceeded will be at the expense of other humans. They explain that the amount of CO_2 every person emits a day is 6.8 kg to keep the planet in balance. This amount is converted to 100 points which the users should share among their various daily activities including transportation and try not to exceed their 100-point limit. In this way the production of CO_2 by us humans is at a minimum, hence leading to a more sustainable environment. Although the application is still a work in progress, their aim to reduce the greenhouse effect through an application is relevant and related to the work done in this thesis.

E-Nudging- motivational aid in the prevention and treatment of chronic diseases in everyday life

Though this is not directly related to green transportation, it involves using nudging (digital nudging) to achieve a goal and in their case, it is used for the prevention and treatment of chronic diseases. They throw light on some of the pros and cons of digital nudging and they emphasize on design concepts to bear in mind while designing digital nudging applications. They mention simplicity in the design, personalization, adaptation of users to the application, getting users to set and achieve small goals with reachable dimensions as opposed to big goals which can discourage them if change is not seen. Their main problems were handling heterogeneity of user groups and getting the users to stay motivated.

Smoking patterns and sociodemographic factors associated with tobacco use among Chinese rural male residents: a descriptive analysis

This study relates to the analysis done in this thesis. They carried out a survey in a rural area; to discover smoking patterns in men of ages 19 and over. They used a multi-stage systematic sampling procedure to determine the people they were to question and from the questionnaire answers (data gathering process), they used a Chi-square test algorithm for analysis from which they made conclusions about their studies like the total percentage of smokers, the smoking rate of the areas and the money spent in buying certain types of cigarettes monthly. This study depicts descriptive data analysis.

Inspiration from related work

The Green transportation choices with IoT and smart nudging paper inspired the interest in the topic of sustainable mobility. The choice of using nudges and separating the analysis logic from the application for maintainability and extension was inspired by this paper. UbiGreen implemented a mobility tracker which provided users' live feedback to persuade them towards sustainability. We used push notifications as a means of persuasion towards sustainable mobility. The mobility choices app inspired us in the direction of building user profiles and collecting user preferences. A better day the 100 way introduces a system of sustainability in several life aspects by encouraging people to stay within a daily CO_2 emission limit. We use the nudge approach for promoting sustainability. The E-nudging paper provided insight on digital nudge design and personalization which were used in the implementation of the NudgeApp. Smoking patterns study encouraged the use of descriptive analysis which helps present data in forms which are easier to understand.

3 Architecture and Design

This chapter describes my approach to data analysis and nudging in a green transportation eco system. I begin by describing the overall architecture of the project, then move on to the approach used in data analysis, after which I extend the architecture to the various components implemented, and finally I present the design of the web application.

3.1 Architecture of the Nudge Project

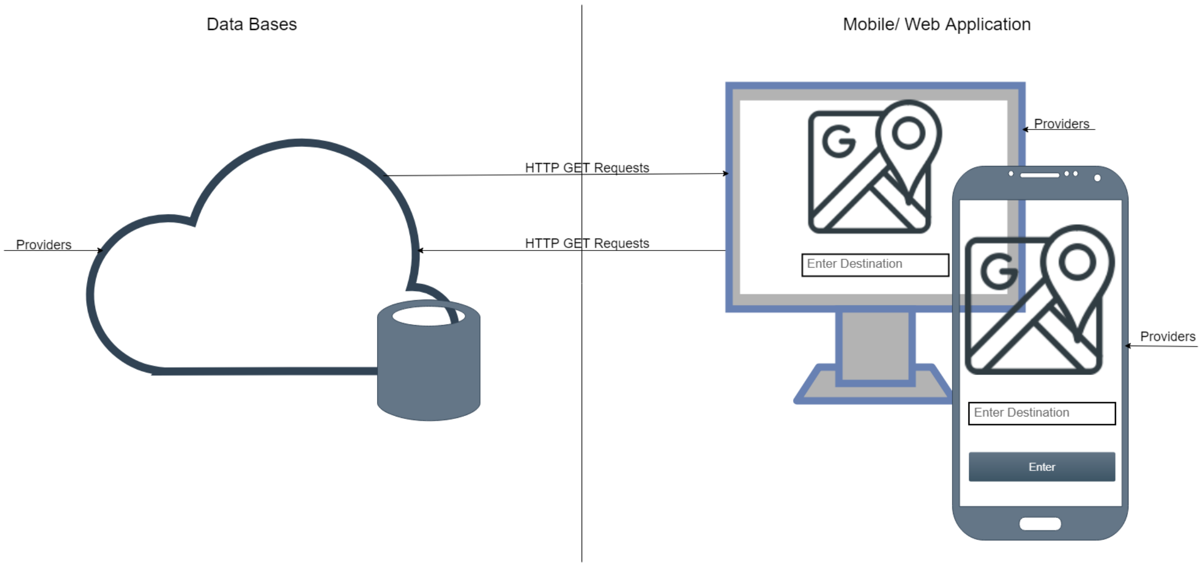

Figure 9. High-Level Architecture of the Nudge project

Figure 9 represents the High-Level Architecture of the nudge project. This architecture comprises all the sections which work together for the implementation to be achieved. We have a web and Mobile application (NudgeApp) represented on the right which serves as a means of communication between us and the users of our application. The left-hand side represents the modules which work together to ensure that the NudgeApp functions at optimal. These modules include the databases, the external providers and the data logic (data analysis and data management). Both sections of the Architecture communicate through HTTP requests.

The external data providers for this project are AccuWeather [70] which provides us with weather data, Tromskortet [71] providing bus data for Tromsø and Google APIs providing direction, maps and location data.

The chosen data sources for this project are very specific for what we aim to achieve. The goal is to nudge for more sustainable forms of transportation, so we need various transport means which we get from using the bus information and Google APIs. With weather being a prominent factor in transportation in Tromsø, it is important to understand how weather conditions affect peoples' transportation choices to get them to make better choices.

Data analysis is done on the data stored in the data repositories. The analysis done helps support decision making (when and whom to nudge) and aims to find and describe patterns in user behavior from their responses to nudges.

3.2 Data analysis approach

The approach to data analysis for nudging towards sustainable road transportation raises issues ranging from what data sources to use, identifying relevant data elements from the data providers, what methods to use to analyze data and why, how the analyzed data is utilized and how analysis helps achieve the goal of nudge personalization.

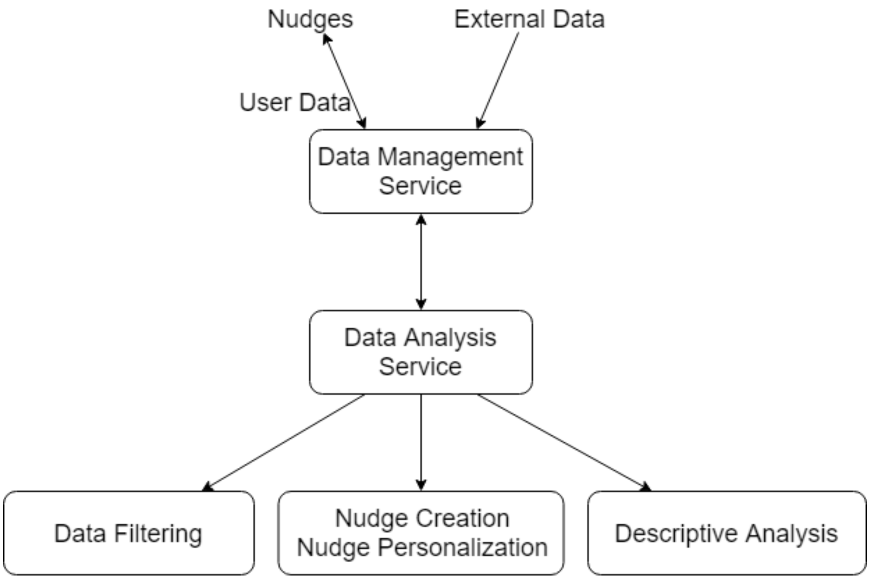

Figure 10. Data Analysis Architecture

Figure 10 represents the Data analysis Architecture of this project. The data sources used for analysis are either from the users of the application or from External sources like weather data and travel data. The Data Management Service [72] communicates with the external providers and the users of the application; getting data, storing it and converting it to forms which are easily understood like converting location coordinates into exact location names and storing them as plain text.

When the Analysis service accesses external and user data, data filtering is done which removes all the unwanted data elements, after which a nudge is created based on the data readings and sent through the data management service to the users. The responses to the nudges are channeled back from the users to the management service and the analysis service can access this data, perform further descriptive analysis and help nudge personalization.

The data accessed by the analysis service helps in taking decisions like whom to nudge, how often to nudge users based on their history, when to stop nudging, how to represent the nudges and how to find user response patterns based on user nudge history.

3.3 Detailed Architecture of the Nudge Project

This section represents a more detailed look at the architecture of the nudge project. The architecture is described in terms of its components, explaining how they work and how our goals are achieved through their functionalities. This is shown in Figure 11.

NudgeApp

The NudgeApp is the application built to forester green transportation which represents the client side of our implementation. It can either be run as a web page application or a mobile phone application. Communication between the servers and the NudgeApp is continuous, enabling dynamic reloads.

Web APIs

The Web APIs serve as a means of communication between the NudgeApp and the components and services running at the backend of the application. All the communication is done through HTTP requests. Requests from the NudgeApp are forwarded to the backend through web APIs. These APIs oversee the conversion of data into forms which are well understood by the services handling requests. The requests are forwarded, and responses are sent back to the NudgeApp through the APIs.

Data Management Service

The data management component is responsible for collecting, storing and managing the data used for this project. All operations involving data use this service. It contains the databases and the logic used to manipulate these databases. It serves data to the NudgeApp and to the analysis service; converting them to their desired formats and sending them to the required components.

Data Analysis Service

The data analysis service works in close contact with the data management component because all the data to be analyzed is stored and managed by the data management service. This service provides the logic used in manipulating data for this implementation. It decides what factors to investigate and what data points it needs for the investigation. The analysis is separated into two components; the Analysis and Nudging parameters.

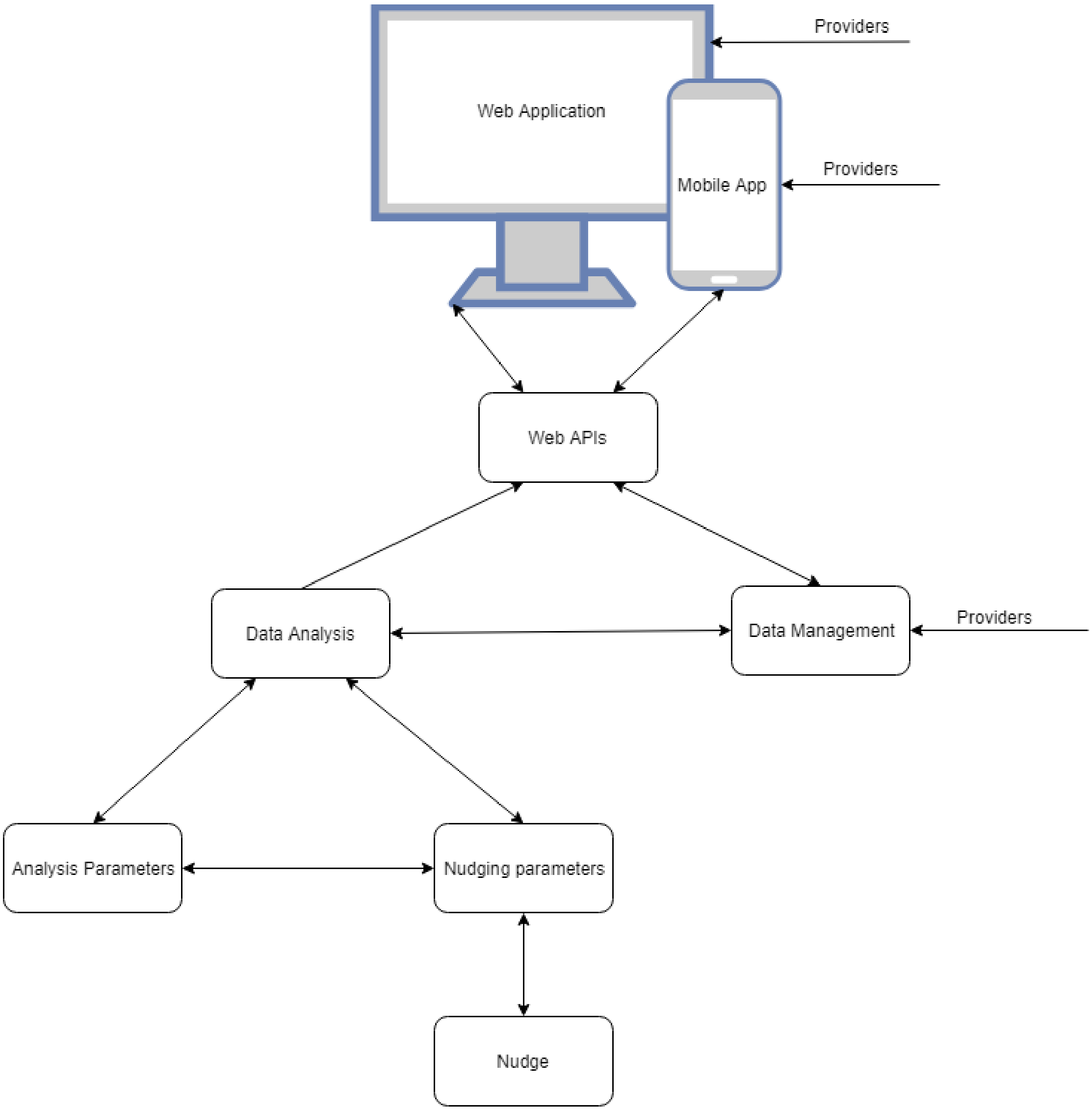

Figure 11. Detailed Architecture with components

Analysis Parameters

This service oversees the provision of descriptive analysis for this project to know how the application functions and what can be done for improvements. This service collects data from the data management component, does some computation and presents the results in forms which are easily understood like a graph. Conditions like the weather conditions where nudges were most accepted or declined or the proposed walking distances which led to the most declined nudges are explored. A representation of these results helps realize patterns in user responses which lead to the overall application improvement.

Nudging parameters

The nudging parameters help decide when to nudge and whom to nudge based on analysis results. This component implements personalization in nudging; deciding which users to nudge for specific conditions and times. This works by checking users' nudge history; seeing how they responded to nudges of specific weather conditions, travel times and distances, and deciding if to continue nudging them for those conditions or to stop.

The NudgeApp is comprised of both the Data driven architectural style and the Event driven architectural style. Users enter events in their calendars and our NudgeApp reads the events from the calendar, does analysis and presents the users with nudges. This represents an event driven architectural style in that the application is triggered by an event in the calendar. For the data driven architectural style, analysis is done on data and based on the results of the analysis, nudges can be sent to the users. This decision is based on the data reading.

External providers

The external providers are AccuWeather for weather data, Tromskortet for Tromsø bus service data and Google APIs for location and map data. These external providers either interact with the backend, through the Data management service making the requests or with the front end, through the NudgeApp making the requests. The NudgeApp requests the maps API, the backend servers request the weather and bus data and the directions API is requested by both the NudgeApp and the backend servers.

3.4 Design

The design seeks to separate the data from the logic manipulating the data. This is useful for continuity of this implementation and it gives room for other data sources. This design choice also helps in the maintainability of the entire project because the logic can be changed as the data changes.

NudgeApp Design

The NudgeApp implementation follows a simplistic design; with few features and an easy to use interface. The manner in which things are presented affects choice [47] and as the choice architect for this implementation, the design choice is simplistic; accommodating people of different age groups. Green was chosen as the principal color for information presentation because green is directly opposite red in the color spectrum and red usually is used to enhance negative words [73]. Green on the other hand is known to be linked with positivity, pleasantness, calmness, happiness, growth and natural fertile environments. Since the application is aimed at promoting positivity, health and sustainable transportation, green is the chosen color.

Signup

At signup, information is presented and collected which helps in the functionality of the application. The name, email and home address are collected, validated and stored in the user table of the database. The user transport preferences are stored in the preferences table and the passwords are collected, hashed and stored in the accounts table of the database. The preferred transport mode collected is relevant for nudge personalization.

Login

Users log in using their names and passwords as shown in Figure 12. They can reset their credentials if forgotten or choose to log in with their Google accounts.

The Google account information is stored in the accounts table of the database. Initially, Google login was not the intended design but became indispensable when the users' calendars had to be accessed. To get access to the calendar on the users' mobile devices, they had to log in with their

Google accounts. With Google login, knowledge of the users' transport preferences provided at sign up would be lost.

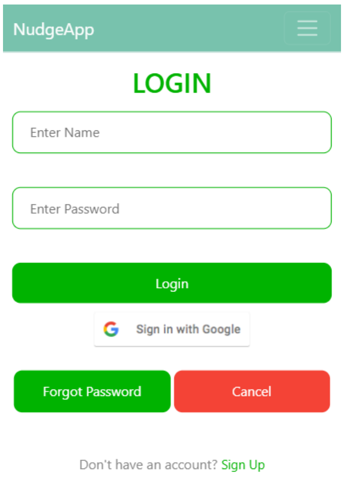

Figure 12.Login Page

Main

When the users successfully log in, they are redirected to the application's home page which contains a map pointing to the users' current location as seen in Figure 13. If the users wish to make a commute instantly, they enter a travel location and on clicking enter, they are redirected to another page holding some travel information. The travel information held on this page is the

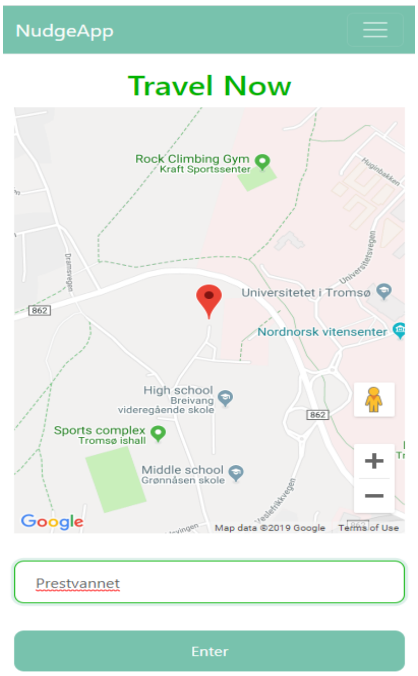

Figure 13. Home Page

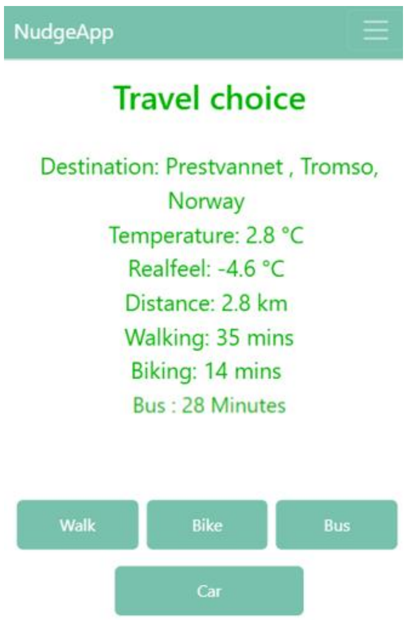

Figure 14.Travel Information

entered destination, basic weather data (temperature and real-feel temperature), distance from present location to chosen destination, walking, biking and bus times; shown in Figure 14.

On clicking either the walk or bike buttons, they are redirected to other pages with maps showing the walk and bike paths respectively from current location to the target destinations. On clicking the bus button, the Tromskortet page opens which indicates the bus stop closest to their present location, the walking distance to get to the bus stop, the bus stop closest to their target destination and the walking distance from the bus stop to the destination. The driving button opens a map showing the driving routes from current location to target destination. The users make their commute choices based on the information given to them.

Preferences

The preferences provided at signup are inaccessible when using a Google account. An adaptation was made where users could provide not only their travel preferences, but also weather temperatures where they would love to be nudged and other weather preferences as shown in Figure 15. Users can adjust these preferences at will and the nudges they receive will be based on the set preferences.

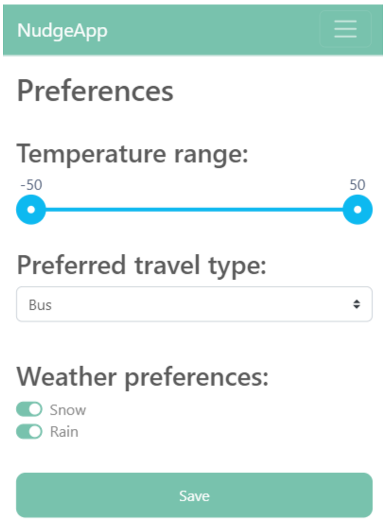

Figure 15. User Preferences

In Tromsø, the main means of transportation are: walking, skiing, cycling, motor biking, taking the bus, driving and taking taxis. The application only uses the transport types ‘Car’ (representing taxis, motor biking and driving), ‘Bus’, ‘Bike’ (for cycling) and ‘Walk’ (representing skiing and walking) because they have the same EF. The reason we only use these modes of transportation in the application is because we are more interested in the impact the transport modes have on the environment, and the transport types in these groups have similar EFs.

Nudge Info

Users of the application are nudged through push notifications. When notifications are sent, the details can be read by opening the notification tab on the app where all the nudge notifications reside as shown on Figure 17. They can see the details of each notification as Figure 16 portrays and either accept or reject the nudges. Their responses to nudge notifications are stored and used for analysis.

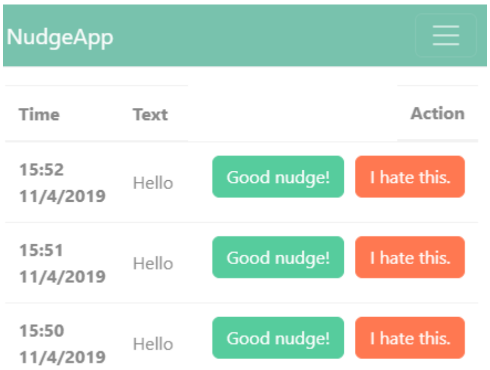

Figure 17. NudgeApp notification page

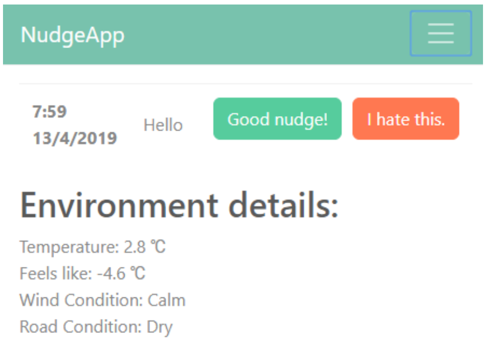

Figure 16. Nudge notification details

4 Implementation

This section details the implementation of my proposed solution to the issue of global warming and other hazards caused by unsustainable forms of road transportation. The implementation is limited to the Tromsø municipality.

I begin by presenting the tools and technologies used in the implementation, then I move to data analysis where I examine what analysis methods and technologies were used and how they work towards my goal. I present details of the main components of the NudgeApp and finally provide some considerations made in the implementation which were not implemented due to time constraints. I mention instances where the data analysis and NudgeApp interact with the data management service through web APIs.

4.1 Technical background, tools and Programming Environment

Programming environment

Visual Studio community 2017

Microsoft Visual Studio 2017 was the latest free version of Visual Studio at the start of this implementation. Visual Studio is a development environment developed by Microsoft [74]. It was the programming environment of choice especially for their support of C# programming language, .NET Core and .NET Framework Tools. Visual Studio ensures code quality by implementing live Unit Tests; this means as the code changes, Visual Studio lets you know if the changes affect your tests in real time.

Backend

ASP.NET Core and C#

The framework of choice for this project is the .Net Core version 2.2 maintained by Microsoft [75]. It is a free, open source, general purpose and cross platform framework which supports the Windows, MacOS and Linux operating systems. It offers full support for the C# programming language.

C# [76] is a general purpose, functional, declarative, multi-paradigm and object-oriented programming language developed by Microsoft within its .Net initiative. This language was chosen because it has full support from the .Net Core and Microsoft Visual Studio.

Frontend

The Angular Framework

Angular is an open source web application framework led by the Angular Team at Google. It is based on traditional web application development techniques like HTML and Typescript. It is cross-platform, so the same code works for multiple operating systems and the code base for the web application can be adapted for mobile applications as well. This promotes code re-usability irrespective of the operating system and scales well for big applications [77].

TypeScript

Typescript is an open source programming language developed and maintained by Microsoft. It is a strict syntactical superset of JavaScript. It is developed for large applications and enhances maintainability by ensuring that all the code is safe at runtime [78]. It also enhances code readability by implementing object-oriented features like classes.

HTML and CSS

Hypertext Markup Language (HTML) is the standard language used for the creation of web pages. Version 5 was used because it is the latest version of HTML [79]. It is used in connection with Typescript and CSS for the Frontend.

Cascading Style Sheets (CSS) [80] is used to add style to the application with the help of Bootswatch, a free Bootstrap [81] styling theme.

Data Analysis

Descriptive Data analysis was used with pattern mining for representing nudge test data and finding similar patterns in the generated data. Exploratory and Predictive analysis are also discussed.

Databases

Microsoft SQL Server 2017

This is a relational database management system developed by Microsoft [82]. We use relations to store all the user data objects and these relations have relationships between them. Microsoft SQL server is chosen for its easy integration with .NET Core; simplifying maintenance.

Oracle Database

This is a database management system developed by the Oracle Corporation [83]. This is used for storing the nudge data. The data stored is used mainly for descriptive analysis and nudge personalization.

Redis
This is a distributed in memory key-value store running on our servers. It is used to cache weather data for a given time which can be used by multiple users to reduce the number of requests made to the external providers.

Other Technologies

Draw.io
The web application was used to create figures used in this thesis [84].

HTTP
Hypertext Transfer Protocol is used in communication between the clients and servers [85].

AWS
Amazon Web Services is a cloud service platform founded by Amazon offering database storage and compute power provided to individuals, companies and governments to help solve scalability problems [86]. It is used for holding instances of the database.

4.2 Data Analysis

To meet the goals of this project; nudge users to choose more sustainable forms of transportation, the nudges must be sent from a place of knowledge. The initial step towards this goal is to understand what factors affect transportation and nudging, what the needs of the people are and how my implementation will be of use and benefit to them and the environment. After deliberations and discussions with members of the research group, it was concluded that the weather conditions, distance to people's destinations, the time of travel, the travel cost of various travel options and the lack of information on alternative travel routes affect how people choose to travel in Tromsø. To get people to change their transportation choices to more sustainable forms, these factors need to be addressed properly. Presenting people with information concerning various travel options, the time and cost of the various options and the weather conditions can greatly affect their choices.

To get the right information to present to people, data from several sources needs to be collected, analyzed and from the results of the analysis, some information can be channeled to people concerning different travel options. The information is sent and presented to the users in a way that their transportation choices can be influenced, because choice is greatly affected by the way it is presented [37]. The presentation highlights sustainable transport types, which can help people with the choice of greener transport means.

The accuracy of the information presented to people is of great importance so that the application is not misleading; hence losing its credibility. The information sent to the users includes a nudge which gives an extra push to people, aimed to influence their choices. When sending the information and nudges, it is important to know whom to send the nudges to, when to nudge, when not to nudge and when to stop nudging for some conditions. This entails the implementation of personalization where each person is handled as an individual and their experience is unique.

The results of the analysis are channeled to people through our NudgeApp; where each user has a profile and they are presented with personalized nudges.

Data Sources

The data sources used for this implementation are classified into User Data and External Data. User data includes all the data collected from the users at signup, their calendar information, their set preferences and the results from their responses to nudge notifications. The external data includes weather data, bus data, direction, maps and location data. The external data is provided through HTTP REST requests to the data management service.

- **External Data Sources**

Weather Data

All weather data is provided by AccuWeather and accessed by the project through API calls. AccuWeather provides varying weather data plans and we use the free Current Conditions and the 12 hours of hourly forecast plans for Tromsø which limits us to 50 API calls daily. Due to this limitation, we implemented a means of caching using the Redis cache, where we access the 12 hours plan and cache it for some time which enables us to make multiple accesses daily for tests. The current conditions plan is used when providing information for instant commutes. People are presented with weather conditions for that hour which can influence their transportation choices. We also use the current conditions plan when issuing nudges based on users' calendar readings and spare time nudges. Spare time nudges are sent willingly to people based on the weather in the weekends and national holidays.

Bus Data

According to Table 1 in section 2.4, taking the bus is environmentally friendlier than driving or taking taxis. The bus schedule information is required to recommend the bus option to people as an alternative to driving or taking taxis. The bus schedules for the Tromsø municipality are provided by Tromskortet. For the bus to be an option, nudges must be sent in a timely manner, matching the users' schedules.

Google APIs

Google APIs are used for accessing the location information, the travel distances and the travel times between different travel points. Users are provided travel durations for walking, cycling and driving using Google APIs. Google Calendar APIs are also used to access users' calendars and based on the readings from the users' calendars, their travel durations and travel distances can be calculated, and nudges sent.

- **User Data**

Initially, user data is collected at signup and stored in the Microsoft SQL database. This data contains the user preferences which help in issuing nudges. However, there is no knowledge of the users' calendars which restricts the functionality of the application. The users have to login to the application with their Google accounts to make their calendars accessible. They can also include their preferences when they are logged in with their Google accounts.

Nudge Data

When nudges are sent to users, either as spare time nudges or nudges based on their calendar readings, the users either accept or reject the nudges. The accept/reject information is stored in the database. This data is used for analysis, helping in implementing personalization and knowing when to modify or stop sending some nudges.

Data Analysis Logic

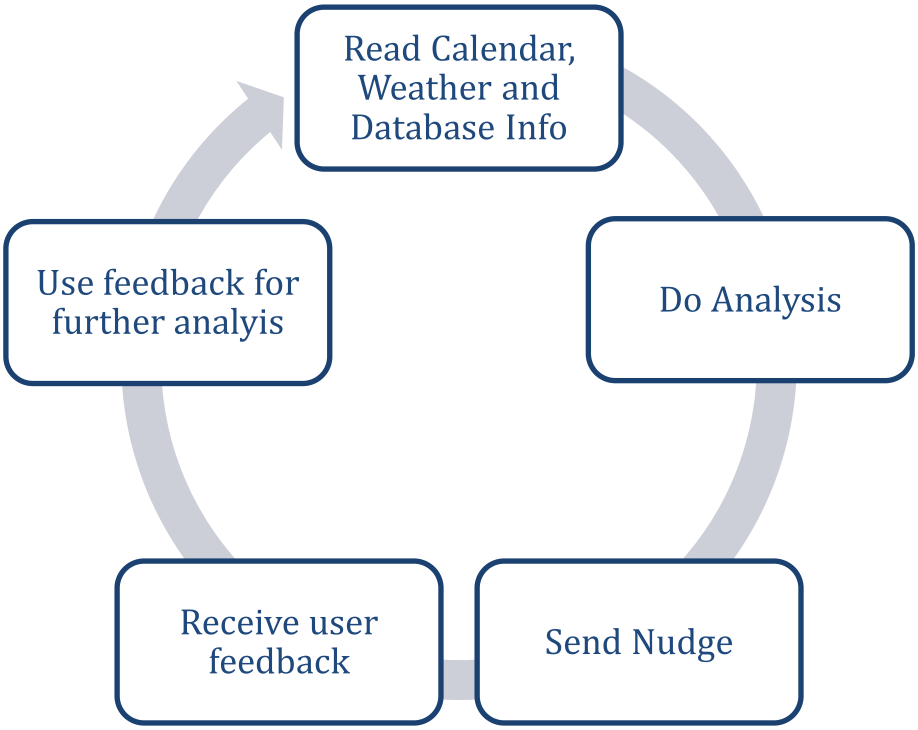

Figure 18. Data Analysis flow

Figure 18 illustrates the flow of the analysis logic; where analysis is done, when and why it is done, and how the results of the analysis affects the overall functioning of the application. The illustration in Figure 18 contains calendar readings, nudge formation, how the nudges are sent to the users, how the users respond to the nudges and how their responses affect the future of analysis and nudging in the application. All the data used by the analysis is provided by the data management service and some of the conclusions of the analysis are sent to the NudgeApp as improved nudges and to the data management service for storage.

The flow in Figure 18 can be explained in two scenarios: user triggered life cycle and event triggered life cycle.

User Triggers

This includes Calendar inputs and Instant commutes made by the users.

- **Calendar Events**

This cycle begins when a user enters an event in their calendar. The NudgeApp reads the calendar event which has a location and start time of event. This information is sent to the analysis service which decides if a nudge should be sent. If the event's start time is within one hour, the analysis service makes a request to the data management service requesting weather data, the user's preferences and their nudge history. The weather data is read, and the user's nudge history is checked for previous nudge information of similar weather conditions. If no previous nudging history under those conditions exist, then a nudge is sent, considering it to be the first nudge under those weather conditions. If previous nudge history exists for that user under similar weather conditions, a decision is made which determines if another nudge should be sent. If the accepted nudges are more than the rejected nudges for those conditions, a nudge is sent. If the rejected nudges are more than the accepted nudges for similar weather conditions, the nudge responses from all users for the same conditions are requested and evaluated as well. If the accepts are more than the rejects, the accept rate is checked which determines if the nudge should be sent. The probability of sending nudges based on general nudge data is explained in chapter 5.

Nudges are sent to the users through push notifications where they can accept or reject the nudge on the application. The result of the feedback is stored in the database which will be used for further analysis, enhancing nudge precision and personalization. This process is explained in Figure 19.

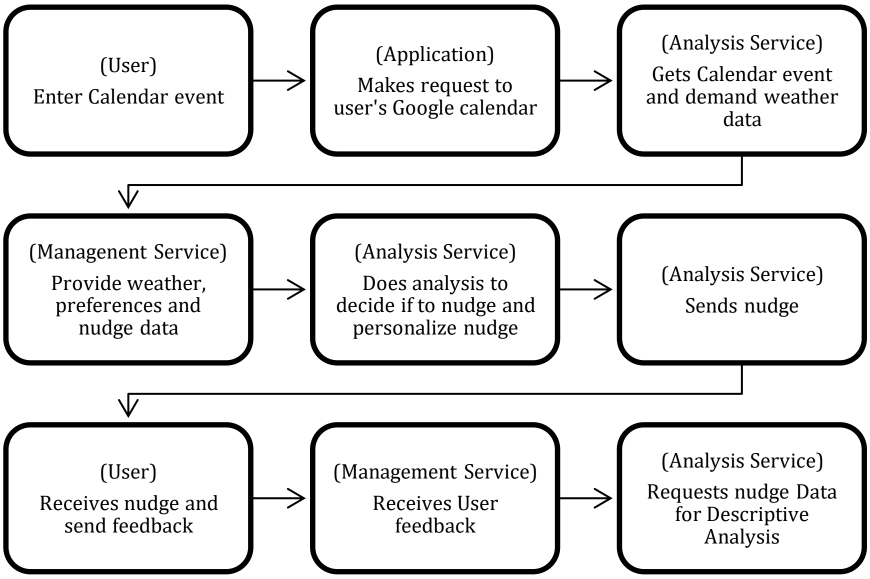

Figure 19. Calendar Triggered nudge flow

- **Instant Commutes**

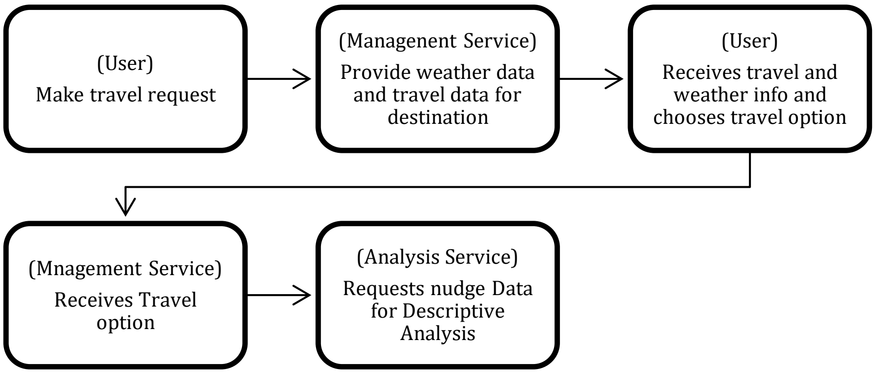

Figure 20. Instant commute nudge flow

This flow begins by users entering travel destinations as explained in section 3.4.1. Although there is no analysis done here, their choices are used for descriptive analysis. When a user makes a travel request, the data management service provides the weather data and the travel information for the entered destination. The user receives the information and chooses a travel option. The response is stored on the database and can be used for analysis as shown in Figure 20. The presentation of travel information poses as a nudge in this case.

Event Triggers

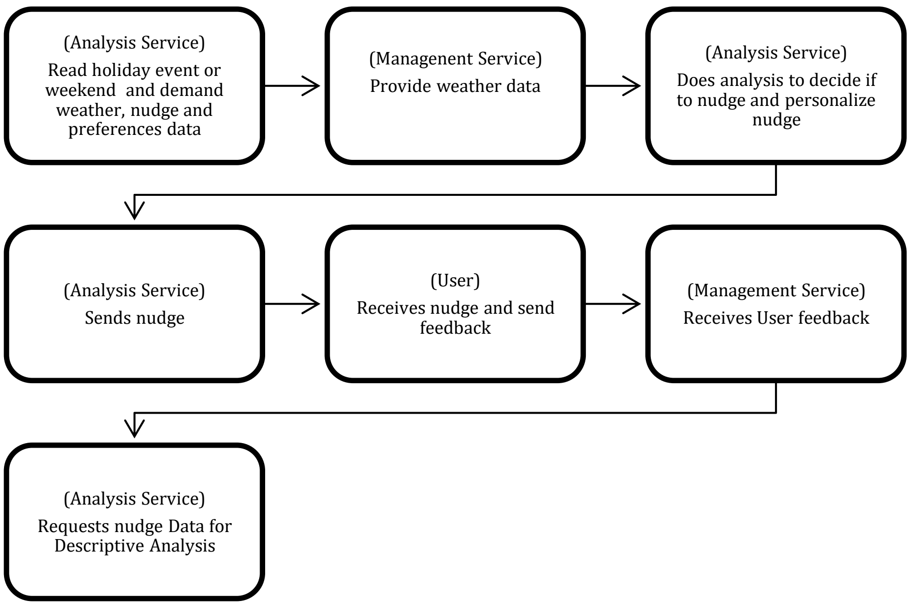

Figure 21. Spare time nudge flow

A task scheduler runs daily at 10 am which checks for national holidays or weekends (Saturday and Sunday) in users' calendars. If it is a weekend or a holiday, weather data is requested and analyzed and based on the weather conditions, the user's preferences and nudge history, a nudge decision is made. The flow continues with the user accepting or rejecting the nudge as shown in Figure 21. These event-triggered nudges which we call spare-time nudges are sent to improve health and to create a more social environment. If a person or group of people are convinced on going on a walk, hike or picnic on a holiday or weekend when they had no intention of doing this, these physical activities help to improve health and enhance a more social environment.

Weather data filtering and Conversion

When the weather data is accessed from AccuWeather, detailed weather information is provided. Not all these data elements are needed for the analysis and for the application to function. The data is filtered, requesting only what is necessary. This knowledge is made possible by the goals of the application and it makes analysis easier to handle; by only investigating data points which are used as opposed to all data points. Data points like *WetBulbTemperature* or *DewPoint* are never used; hence filtered out and weather conditions which affect transportation like the *Temperature, RealFeelTemperature, PrecipitationType* and *PrecipitationProbability* are kept. These data elements are further simplified by converting some of them into more readable and simplified forms. The *PrecipitationType* can be converted into *Snow, Rain, Ice* or *NoPrecipitation* which will make more sense when sent to users.

Nudge Creation

The nudges sent to users of the application are Text Nudges which are sent through Push notifications. The nudges are predefined text nudges which are based on weather conditions. When the weather data elements are converted to their desired forms, weather conditions are placed together to create nudges. For nudging people to take walks for example, weather conditions like the *Temperature*, *RealFeelTemperature* and *PrecipitationType* can be used to determine the state of the roads. The data readings can determine if the roads are *slippery*, *wet* or *dry.* For all these conditions, varying text nudges are created and stored. When there is a nudge trigger, if the decision to nudge is reached, the correct predefined nudge for that weather condition responds with the nudge information sent through a push notification. For nudging people to take a walk on a Saturday, a nudge example presented on the notification can be: '*The skies are clear today, a warm jacket and boots will be good companions for a walk*'. This message carries basic information about the weather (clear skies) which takes *PrecipitationProbability* into account and provides this information with a nudge suggesting that the user takes a walk. It also takes the *Temperature* into account; hence suggesting a jacket.

User personalization

The entire data analysis logic explained in section 4.2.2 is aimed at giving users a personalized view and feel of the application, which helps in improving the nudges sent and evaluating the entire application. This section explains how data analysis helps to implement personalization for calendar entered events, holidays and weekends as shown in Figure 22. This implementation works based on the user personalization settings, the weather data, the nudge history of one user and the nudge history of all the users.

Two scenarios exist with user personalization: user set preferences and no set preferences.

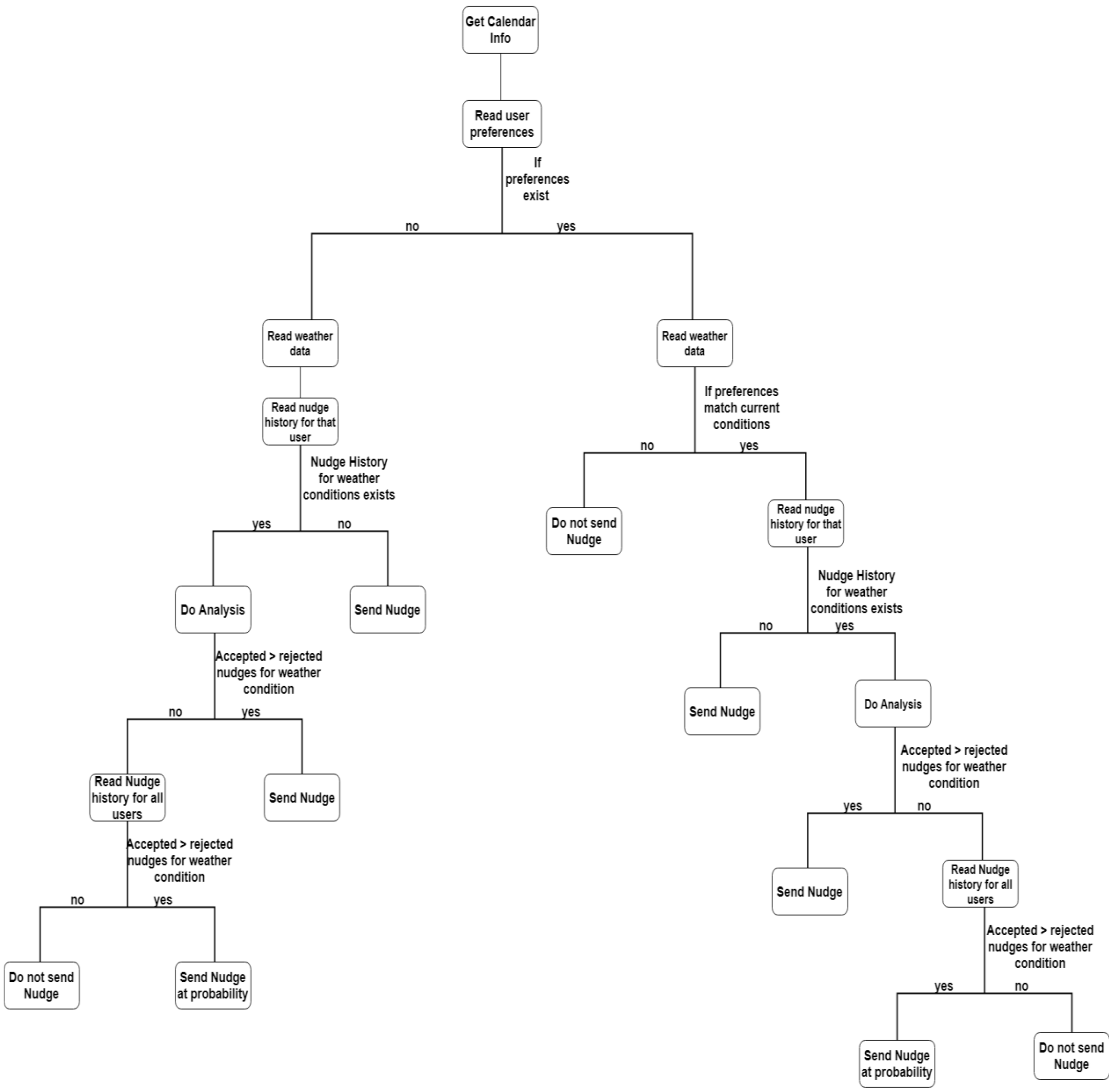

Figure 22. Nudge Personalization

User set preferences

The analysis service gets a calendar reading for a user. It requests the user's preferences, weather data and nudge history from the data management service. The analysis service compares the user's set preferences to the weather data. If the preferences do not match, no nudge is sent, but if the preferences match the weather conditions, it checks for the user's nudge history. If the user has no nudge history for the weather conditions, a nudge is sent assuming it is the first nudge under those conditions. If the user's nudge history exists, analysis is done by checking if the user had more positive responses to nudges under those weather conditions. If more of the responses were positive, a nudge is sent. If most of the responses were negative, analysis is done on all the nudge responses under those same conditions. If most of the responses were negative, no nudge is sent, but if most of the responses were positive, a nudge is sent depending on the response percentage explained in chapter 5.

No set preferences

If the user has not set any preferences, the weather data and the user's nudge history are requested. If no nudge history exists, the nudge is sent as the first nudge under those weather conditions. If nudge history exists, analysis is done on the nudge history of the user and others which helps in the nudge decision. If the user's previous reaction to nudges of similar conditions were positive, a nudge is sent else the reaction of all users for nudges under similar conditions is checked. If their responses were positive, a nudge is sent at a probability, else it is not.

When to Nudge for user calendar events

When sending nudges, timing is of optimum importance. Driving with no traffic is the least time-consuming commute choice as seen in Table 1 in section 2.4. Knowing when to nudge so the users have enough time for environmentally friendly commutes is crucial. If the nudges are sent too early, they might be forgotten [35] and if they are sent too late, there might not be enough time for sustainable transport choices. For this implementation, nudges are sent 1 hour before the start time of event. This time is chosen based on the application locality (only for Tromsø) and the decision not to nudge for more than 30 minutes of walking or biking. Calendar events set by the users are checked whenever the application is in use. If an event's start time is within one hour, the analysis service decides if a nudge should be sent based on other data readings.

Descriptive Analysis

Descriptive analysis is analysis which shows or summarizes data in a meaningful way for it to be well understood [57] and to find patterns which might emerge from the data. The data can be represented in charts, maps or histograms. Information like the average number of accepted nudges, the most accepted nudges, the day with the most rejected nudges can be deduced from the represented information. If the average number of nudges sent for a weather condition is known and the accepted nudges for that condition are less than the average, a decision can be made to stop sending or change that nudge. If information is collected and analyzed long enough, patterns can be realized from these representations like some days in the week having almost 100% rejects. These observations can be used to predict future reactions and responses, make the application more interactive and work towards achieving the goal of sustainability and to simplify future analysis.

4.3 The NudgeApp

This section describes the NudgeApp, its main components and functionalities, how it channels analyzed data to the users, how data is retrieved for further analysis and how all this translates to nudging for sustainable mobility. This prototype consists of a web application with the front end built with Angular 6.1.

The frontend of this application is made up of components which represent the displayed pages and services which work between components; linking them to each other, getting data from the backend servers and transferring data from one page to the other as needed.

Organizing the application into several components improves code maintainability, reusability, reading and testing. Individual pages can be modified, added or deleted without affecting the interface or functionality of the entire application. Ideally, each component holds a single idea which makes unit testing easy; ensuring that each page is responding as it should.

The following sections describe how the main pages function, their connection to other pages and how they interact with the back end through the helper services.

Signup and Login

The application uses both local authentication for signup and login and Google authentication for logging in with users Google accounts.

Local Authentication

The initial implementation of the application involves a signup and login. For the signup, user credentials like name, email, preferences, and password are collected. A random password salt is generated which is used together with the password to generate a password hash, and they are stored on the Accounts table in the database. The *UserService* helps transfer the user credentials to the database.

When the user provides their credentials at login, the password is hashed using the password salt stored in the database and if the hashes match, a token is created and sent back to the NudgeApp, and all the user's HTTP requests are made to the backend servers. The *AuthenticationService* oversees the communication between the login component and the backend servers.

The local authentication method gives us access to the users preferred transport means which guides us when nudging (Personalization) but its functionality is limited because we do not have access to the users' calendars. We can only provide instant commutes and issue spare-time weekend nudges in this case. This problem is solved by letting the users login with their Google accounts which gives us access to their Google calendars.

Google Authentication

If the users decide to log in with Google, their information is received by the frontend as a token. This token is sent to the backend servers which communicate with Google servers, checking if the provided email and password match. If the credentials are authenticated by Google, an update is made on the database with information that the user has registered with their Google account and an OK is returned to the frontend. The communication between the frontend and the backend servers is done by the *AuthenticationService*. With users' Google login, their calendars can be read, and nudges given.

Location Service and Travel Information

After a successful login, the users are redirected to a page containing a map pointing to their current location. The user's location services on their device must be enabled for the application to work. We use Google maps geocoder to get the exact location of the user which is displayed on the map.

The user inputs their travel destination as shown in Figure 13 and the destination, travel distance, travel duration for various transport choices and basic weather information are displayed. The weather data in this case is provided by the Current Conditions plan. The destination is provided as plain text, passed as a query to the Google maps API where the distance between the location and the entered destination is calculated and displayed.

We get the travel times for walking and cycling using Google's Distance Matrix Service. The current location is provided through the APIs and the entered destination is provided as an address, together with the travel type: either walking or cycling. Google responds with a list of travel results from the current location to the entered destination. The first result is selected and displayed to the users.

For the bus data, we use Tromskortet, Tromsø's local bus service. This service is handled in the backend by the data management Service. The nudge app backend servers communicate with Tromskortet's servers to get the bus schedules. The user's read location and entered destination are sent to Tromskortet which responds with the closest bus stops for their present location and their entered destination together with the total travel duration.

Push Notifications and user feedback

Our actual nudges are sent to the users through push notifications. These nudges are either user triggered (from calendar input) or event triggered (spare time or holiday events). We implemented a push notification service which forwards nudge messages decided by the data analysis service to the users. Users grant or refuse notification permissions when they first log in. If they granted permissions, nudge notifications are sent to them thereafter which they can either accept or reject. Their responses are forwarded to the database where further analysis is done. A daily task schedular reads all calendar events and the analysis service decides if a nudge should be sent.

4.4 Considerations and Suggestions for Analysis

During the implementation of this thesis, some considerations were made which were not implemented.

Additional Data sources and Analysis methods

Descriptive analysis with pattern mining was a choice for this implementation because the purpose was to find patterns in users' responses to nudges which will help in simplifying analysis and providing more insight on when to nudge. Prescriptive analysis can be used with the goal of getting people to walk more. This will entail a more detailed look at patterns and understanding why people make certain choices. When these details are known, historical data can bring about a solution to the goal of getting people to walk more. Predictive and exploratory analysis are plausible solutions to data analysis, which can be used with the goal of finding patterns, creating personalized nudges and finding solutions towards sustainable mobility.

Using data classification analysis on nudge data can help simplify analysis by placing nudge data into several classes like Accepts, Rejects, and even finer classes containing weather conditions and travel duration. An algorithm like the decision tree algorithm can be used to map the incoming nudge data into its appropriate class, after which the various classes can be analyzed, and conclusions drawn. Classification analysis gives room for additional data sources to be investigated, enhancing scalability. For additional data sources, classification can be used together with pattern mining gaining more details about the data.

For multiple data sources, OLAP techniques can be used which will provide details of the information contained in the data.

Transportation tracking and Smart Nudging

For the present implementation, users manually put in calendar information, and they are nudged based on the calendar input and occasionally get spare time nudges. Users also must manually accept or reject nudges for us to get information about their travel choices. With rejected nudges, we do not know how they traveled. Their defaults can only be detected by analyzing data and deducing this information from their most accepted nudges which again might be inaccurate. A more interactive suggestion and extension to data analysis and the application will be to use smart nudging techniques. Smart nudging will entail tracking the users to get information about how they travel, deducing their work place, school or home and sending them nudges multiple times in a day suggesting sustainable forms of transportation based on their tracked location and the transportation means they used to get there. With smart nudging, the suggestions given throughout the day need to follow the users' transport means they used while leaving the house. Suggesting users to take the bus after work when they took the car to work makes no sense and can make the application unreliable [35]. With smart nudging, the users will not necessarily need to update their calendars and accept or reject nudges.

Other Nudge types

The implementation now uses text nudges containing weather information. As explained in section 2.6, nudges receive more positive feedback when they come with additional incentives and not just plain messages. A nudge to walk can include the number of calories burned for walking that distance as opposed to just weather information. Information about filled up parking lots or traffic on roads usually used by people can influence their travel choices. The nudges sent can also be extended to include information about bus schedules or the amount of money saved on fuel. Any additional incentives which makes users feel like they have immediate rewards will have more positive responses than plain text nudges.

Value of Travel Time Savings (VTTSs)

An important consideration to be made when designing and issuing a nudge should be the value of every minute spent on transportation or the Value of Travel Time Savings (VTTSs) [87]. For time sensitive activities, where a delay in sustainable transportation commutes can cause a loss of clients or miss an appointment, people can be nudged to carpool or just drive or take taxis. Nudging for sustainable transportation should be done for more relaxed and less time sensitive activities like friendly visits. These time-sensitive and leisure activities can be separated using classification mining.

5 Experimentation and Evaluation

This section provides a view of descriptive analysis done using nudge data. It also provides a model used to determine if nudges should be sent based on nudge history. Unit tests are done to certify the correctness of the created text nudges; ensuring that the weather condition readings which create a nudge are correct.

The data used for Descriptive analysis and Nudge probability are generated and this evaluation is for demonstrative purposes. The push notification service was set to send 100 notifications a day for 14 days (2 weeks) and the notifications were responded to daily, and the results were saved on the database.

Descriptive Analysis

This represents analysis done on data from user nudge responses. One test user was populated with nudge data as shown in Figure 17 in section 3.4, and the responses to the nudges were stored in the database. The results obtained from the database are represented in Figures 23 and 24; for weeks one and two respectively.

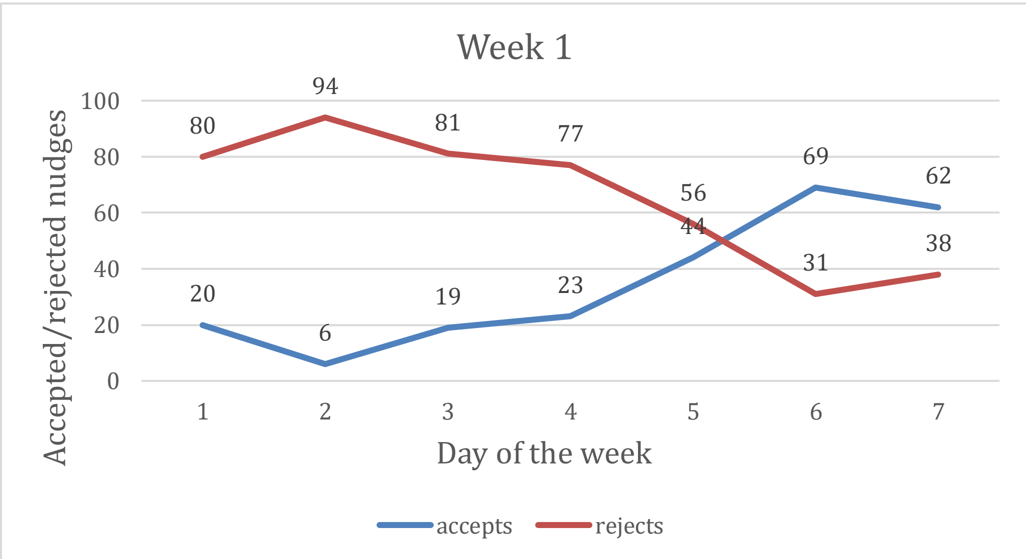

Figure 23. Descriptive Analysis Week 1

From this representation, some observations can be made, like the day with the most accepted or rejected nudges, the average response to the nudges and the percentage of the accepted or rejected nudges. This information is relevant for adjusting future nudges, but the information is not enough for finding patterns. To find patterns, the same situation needs to be tested over time and the results for each test analyzed.

Figure 24 represents a second week of running nudge tests. The results can be used together with the results represented on Figure 23 to find a pattern.

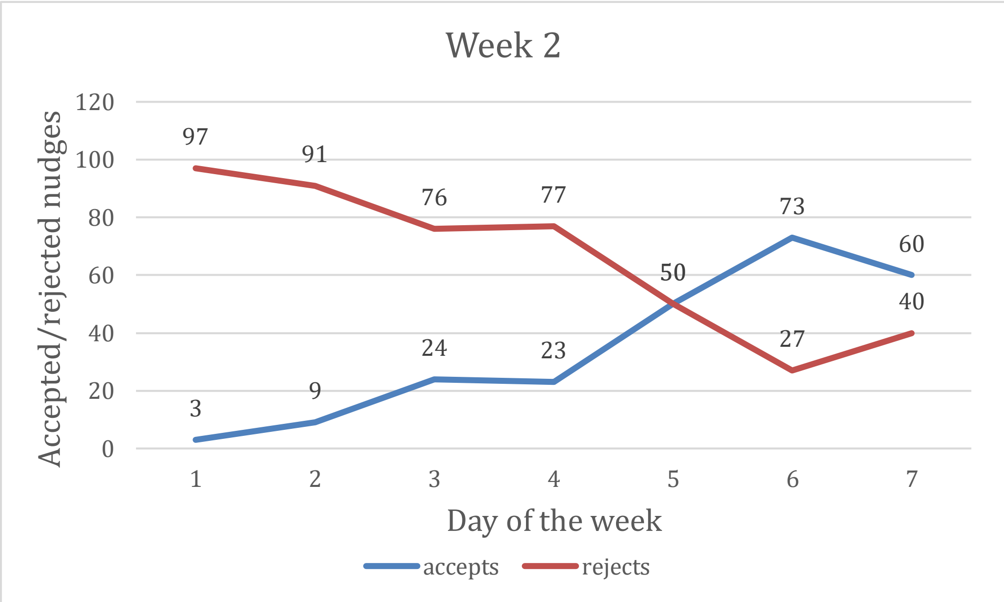

Figure 24. Descriptive Analysis Week 2

Probability of Nudging based on Nudge History

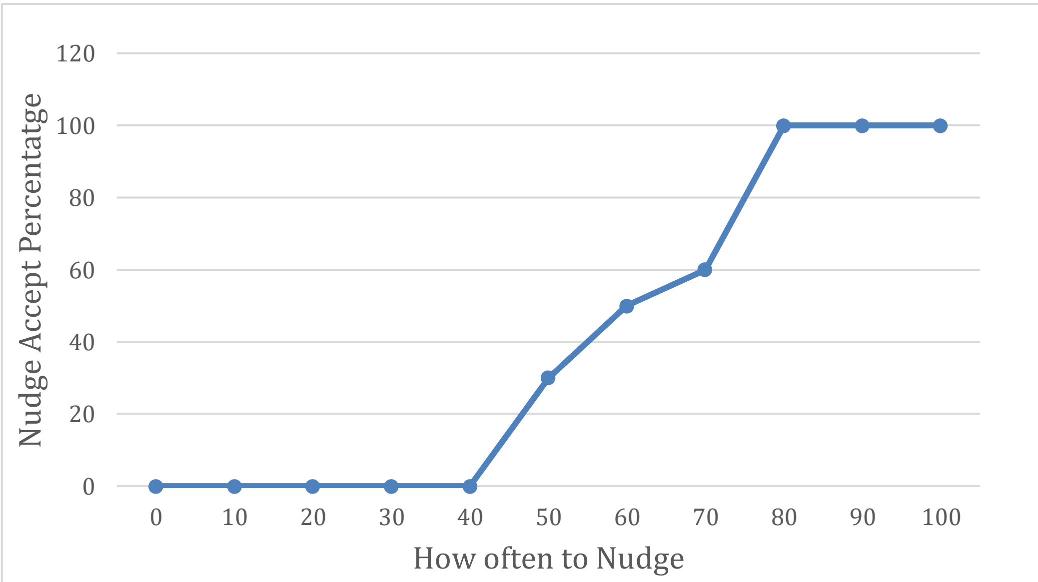

Figure 25.Nudging Probability

When deciding if a nudge should be sent to a user for a condition where their rejected nudges were more than the accepts, the overall reaction to the nudge under the same condition is to be checked as explained in section 4.2.3. If previous nudges for that condition had over 80% of accepts, future nudges for the same condition are always sent. If the accept rate for a condition is less than 50%, nudges for this condition are never sent; the nudges are changed. If the accept rate is between 50% and 70%, the nudge is sent as often as the percentage (if the accept rate is 60%, nudges for that condition are sent 60% of the time); not all the time they arise as Figure 25 demonstrates.

This method was devised so that undesirable nudges should be sent less, and users are not overcrowded with nudges.

6 Discussion

This section entails the discussion of results, some design choices and considerations which were made during the implementation of this thesis.

6.1 Discussion of Results

Finding patterns

Data represented on graphs is easily understood and relevant conclusions can be drawn from it. It is also easy to find patterns when data is represented on graphs. Figure 26 represents the accepted nudges for both weeks. This was extracted from Figures 23 and 24.

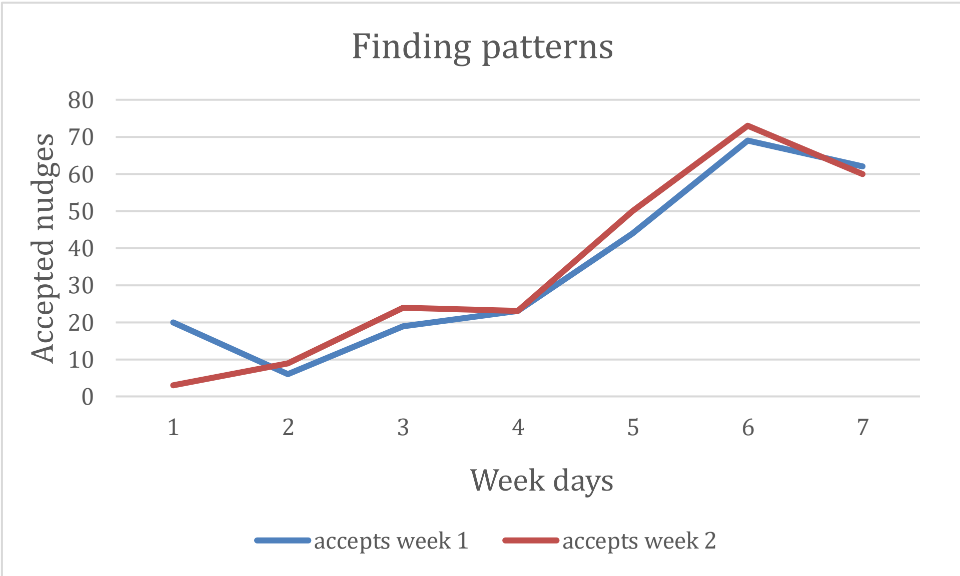

Figure 26. Demonstrating patterns

It is seen that Monday and Tuesday have the lowest accept rates while Saturday and Sunday have the most accepts. After this observation, the nudge can be examined; together with user behavior, finding reasons why the nudge responses on some days are drastically low. If the nudges on Mondays and Tuesdays over time experience similar results, a decision can be made to stop nudging on those days.

Other conclusions from nudge data

From the readings on Figures 23 and 24, other conclusions about nudges can be made. Like: Total number of nudges sent for two weeks was 1400, total number of accepted nudges were 485, and rejects were 915 with the accept average for both weeks being 34.64%. The accept average for week 1 was 34.71% and 32.57% in week 2. This implies a 0.14% drop in the accept rate.

The statistics presented for the test nudge data generated for demonstrative purposes show how real nudge data from multiple users can be evaluated and how conclusions can be drawn from nudge data.

6.2 Design Choices and Implementation Considerations

Age, sex and health considerations

These conditions were ignored willingly; not only because the application is optional for its users but also because users can state personalized preferences and over time, based on their nudge history, undesirable nudges will be stopped, and their preferences will be acted upon. Users are also presented with several travel options to choose from. Though sustainable transport choices are highlighted, other less sustainable forms like driving routes are provided; hence a nudge application.

Environmental Friendliness

For simplicity sake, we use the EF grouping done in [35]. We place walking and skiing in the same group, driving and taking taxis in the same group. All the transport modes in an EF group are assumed to have the same effect on the environment. This prevents us from representing a long exhaustive list to the users and emphasizes that our focus is the effect of the transportation choice on humans and on the environment and not necessarily the type of transport.

Choosing to Stop or Modify the Nudges over time

Humans form habits or do certain actions effortlessly with time. This indicates that at the start of forming a new habit or exercise, they need to be reminded and given an extra push; else they fall on their default actions [36]. As time progresses while performing a new habit, it becomes less difficult and the new habit might become their default habit. Unfortunately, as [88] describes, there is no particular time frame which when exceeded we are sure a habit has been formed. All we know is when humans perform certain tasks often, they tend to make that their default. This also applies to our NudgeApp. If the user accepts nudges over a time frame and the accepted nudges were not their default, then we can assume a habit is being formed and we can stop or at least reduce the number of nudges sent for that action or start nudging them towards a new and more sustainable travel choice. If the user however rejects most of the nudges, we are left to decide if we should modify the nudges or stop so as not to be intrusive or annoying.

Trusting the users’ responses

When we send nudges, we propose a transportation type for the users. When the users accept or reject the nudge, their responses are saved along with the proposed transport types. For simplicity, we trust the users are honest about their choices. From this, we can tell if the users preferred mode of transport has changed after analysis.

Application evaluation

Though more accepted nudges represent a positive feedback for the application, it does not represent the success of the application. For the application to be considered successful, users' preferred transport types should change to transport types with higher EFs, or they should use transport types of higher EFs for leisure activities as discussed in section 4.4.4. This change in preferences represents the success of the application.

Why consider nudging users with previous negative responses

In some cases, even when a user's nudge history shows negative responses to a condition, they still get nudged. This decision is taken based on the overall response to nudges under the same conditions. For some people, changing their defaults can be difficult, so we keep nudging them, though to a lesser extent just to give them a slight push and reminder. These nudges can potentially help in changing their defaults or at least get them to use sustainable transport choices.

7 Future Work

Some considerations and suggestions for continuation of data analysis and nudging were explained in section 4.4. This section contains more suggestions for continuation of the implementation.

7.1 Including Several other Data sources

For continuity, more precision and usability of the application, we will include several other data sources like traffic data, parking lot data, road maintenance and road blockage information and extend the bus data to give us delays. These data sources are relevant to us because with them, we can make more precise nudges. Having information on road blockages or delayed buses can be relevant to users of our application so they know to take other routes or walk or bike to their destinations. Parking lot data is also very relevant because users can know which parking lots are filled up and the knowledge of this can influence them to take other means of transportation. Incorporating these other data sources will make the nudges more precise and the application more useful.

7.2 Including City events data for spare time nudges

We implemented spare time nudges which we send to users now based on weather conditions on the weekends and on national holidays. We nudge users to take walks, go skiing, go to the park based on what the weather conditions are. This functionality can be extended to include events in the city which the users may or may not have been aware of. We can send information of these events to the users and nudge them to take sustainable forms of transportation for the events.

7.3 Exploiting the health aspects and cost savings of sustainable transportation

The constant emissions from motor vehicles have been known to cause air pollution and especially known to be a great contributor to Particulate Matter (PM). PM is composed of solid and liquid particles which cause premature death in people with heart and lung diseases, heart attacks, decreased lung function, and aggravated asthma [89]. These health issues pose a threat to longevity and the quality of life. Reducing emission of PM and other GHGs will slow down these respiratory diseases and infections and improve the quality of life. More sustainable forms of transportation like walking and cycling also have health benefits including a reduction in cardiovascular diseases. Studies have shown that the adult population needs at least 30 minutes of moderately vigorous physical activity on most days of the week [90] and even up to 90 minutes daily to fight obesity.

[91] gives us cost related concerns in vehicle ownership and usage including vehicle license fees, fuel taxes, vehicle use fees, emission fees, road tolls, parking fees and fuel prices. Using sustainable transportation means like walking, cycling and taking the bus eliminates these extra expenditures.

7.4 Mobility tracker and Smart nudging

Providing tracking possibilities on users' devices will make the application more interactive and in the long run, smart nudging will be possible. Users will not need to manually input calendar data. The application will learn their travel behavior and after analysis, smart nudges will be sent to the users based on the learned travel information.

7.5 Detailed feedback to users of the application

The choice of giving user's feedback as explained in [33] says that feedback and personalization can positively impact environmentally responsible behavior. A study in [92] saw a 5 to 12% savings in electricity consumption when users were given feedback on their electricity usage. After weekly analysis of user travel data, feedback can be sent to them detailing their transportation choices throughout the week and the effects these choices have on them and the environment. We can extend this by converting the distance driven to the amount spent on fuel and the distance walked/cycled to the number of calories burned. This will be possible if we are able to track the users' transportation and use the data for further descriptive analysis.

7.6 Run on mobile device

Using NativeScript on the Angular platform, the application can be run on mobile devices

7.7 Extend implementation out of Tromsø

The implementation of the nudge app is dependent on the Tromsø municipality. This limitation was run into because Google APIs have no data about the Tromsø municipality bus service, so we had to use the local bus service, Tromskortet. Several countries, cities and towns have different transportation options which would have been complex to integrate into the application in the time we had. We decided to limit this implementation to Tromsø to test the concept of digital nudging in an urban environment which can be extended to include other towns and countries.

8 Conclusion

This thesis examines ways in which data analysis can be used through a mobile application for enhancing sustainability in road transportation. Our NudgeApp is just a means of channeling our ideas to the users. This implementation is just one solution to combating excess CO_2 and other GHG production caused by road transportation, hence global warming. We also look at the health concerns caused by these gases and the health and financial gains which can be achieved by using sustainable transport methods.

For the type of data we had and the way it was stored, I decided to leverage Descriptive analysis, just to demonstrate as a proof of concept how analysis can be done in such a situation, and I showed how patterns can be detected using data represented on graphs. The results of these analysis are used to offer personalized nudges and to demonstrate how the application functions.

The possibilities of this concept of nudging for green transportation are endless, and they can be extended and adapted to suit several transport types. Since the concern is global warming, climate change and health, simple life practices can be employed which enhance sustainability and reduce the overuse of natural resources. Little actions in everyday life can help reduce our GHG emissions such as taking the stairs instead of elevators for physically fit people, unplugging devices when not in use, carpooling or turning of lights when not in use. These simple changes, together with our proposed solution can help create a safe, pollution free and sustainable environment for us all.

9 References

[1] A. Ajanovic and R. Haas, "Dissemination of electric vehicles in urban areas: Major factors for success," *Energy,* vol. 115, pp. 1451-1458, 2016.

[2] Y. Xu, V. Ramanathan, and D. G. Victor, "Global warming will happen faster than we think," ed: Nature Publishing Group, 2018.

[3] N. R. Council, *Climate Change: Evidence, Impacts, and Choices: Set of 2 Booklets, with DVD*. National Academies Press, 2012.

[4] S. M. Papalexiou, A. AghaKouchak, K. E. Trenberth, and E. J. E. s. f. Foufoula‐Georgiou, "Global, Regional, and Megacity Trends in the Highest Temperature of the Year: Diagnostics and Evidence for Accelerating Trends," vol. 6, no. 1, pp. 71-79, 2018.

[5] J. Hansen, M. Sato, R. Ruedy, G. A. Schmidt, K. Lo, and A. Persin, "Global temperature in 2015," *GISS, NASA, NY. URL: http://data. giss. nasa. gov,* 2016.

[6] A. Robock, "Volcanic eruptions and climate," *Reviews of geophysics,* vol. 38, no. 2, pp. 191-219, 2000.

[7] S. Solomon, G.-K. Plattner, R. Knutti, and P. Friedlingstein, "Irreversible climate change due to carbon dioxide emissions," *Proceedings of the national academy of sciences,* vol. 106, no. 6, pp. 1704-1709, 2009.

[8] W. International Energy Agency. (2011). *Health and sustainable development* [Online]. Available: https://www.who.int/sustainable-development/transport/health-risks/climate-impacts/en/.

[9] T. Berntsen and J. Fuglestvedt, "Global temperature responses to current emissions from the transport sectors," *Proceedings of the National Academy of Sciences,* vol. 105, no. 49, pp. 19154-19159, 2008.

[10] R. G. Cuddihy, W. C. Griffith, and R. O. McClellan, "Health risks from light-duty diesel vehicles," *Environmental science & technology,* vol. 18, no. 1, pp. 14A-21A, 1984.

[11] S. Arrhenius, "XXXI. On the influence of carbonic acid in the air upon the temperature of the ground," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* vol. 41, no. 251, pp. 237-276, 1896.

[12] S. R. Weart, *The discovery of global warming*. Harvard University Press, 2008.

[13] B. K. Bose, "Global warming: Energy, environmental pollution, and the impact of power electronics," *IEEE Industrial Electronics Magazine,* vol. 4, no. 1, pp. 6-17, 2010.

[14] M. Z. Jacobson, "Review of solutions to global warming, air pollution, and energy security," *Energy & Environmental Science,* vol. 2, no. 2, pp. 148-173, 2009.

[15] J. R. Malcolm, A. Markham, R. P. Neilson, and M. Garaci, "Estimated migration rates under scenarios of global climate change," *Journal of Biogeography,* vol. 29, no. 7, pp. 835-849, 2002.

[16] C. D. Thomas *et al.*, "Extinction risk from climate change," *Nature,* vol. 427, no. 6970, p. 145, 2004.

[17] D. J. S. A. Biello, "How Much Is Too Much?: Estimating Greenhouse Gas Emissions," 2009.

[18] U. Nations. (1998, May 2019). *Kyoto Protocol ro the United Nations Framework Convention on Climate change* [Online]. Available: https://unfccc.int/resource/docs/convkp/kpeng.pdf.

[19] V. Pham, F. Riis, I. Gjeldvik, E. Halland, I. Tappel, and P. Aagaard, "Assessment of CO2 injection into the south Utsira-Skade aquifer, the North Sea, Norway," *Energy,* vol. 55, pp. 529-540, 2013.

[20] G. S. J. Q. J. o. t. R. M. S. Callendar, "The artificial production of carbon dioxide and its influence on temperature," vol. 64, no. 275, pp. 223-240, 1938.

[21] O. Hoegh-Guldberg *et al.*, "Coral reefs under rapid climate change and ocean acidification," *Science,* vol. 318, no. 5857, pp. 1737-1742, 2007.

[22] J. A. Lorenzen, "Going Green: The Process of Lifestyle Change 1," in *Sociological Forum*, 2012, vol. 27, no. 1: Wiley Online Library, pp. 94-116.

[23] D. Sperling, D. Salon, and N. J. A. P. A. S. R. Lutsey, "Reducing greenhouse-gas emissions from transportation," no. 557, pp. 85-95, 2009.

[24] N. C. C. Secreteriate. (2017, May 2019). *Transport* [Online]. Available: https://www.nccs.gov.sg/climate-change-and-singapore/reducing-emissions/transport.
[25] J. Hanania. (2015). *Primary pollutants* [Online]. Available: https://energyeducation.ca/encyclopedia/Primary_pollutant.
[26] B. Afework. (2019). *Secondary pollutants* [Online]. Available: https://energyeducation.ca/encyclopedia/Secondary_pollutant.
[27] World Commission on Environment and Development. and G. H. Brundtland, *Our common future*. Oxford: Oxford University Press, 1987, pp. xv, 383 pages.
[28] M. P. Walsh, "Motor vehicle trends and their implications for global warming," in *Transport policy and global warming p. 69ff, The European conference of ministers of transport (ECMT), Paris*, 1993.
[29] D. L. Greene and M. Wegener, "Sustainable transport," *J Transp Geogr*, vol. 5, no. 3, pp. 177-190, 1997.
[30] R. Wash, L. Hemphill, and P. Resnick, "Design decisions in the RideNow project," in *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, 2005: ACM, pp. 132-135.
[31] F. M. M. Cirianni and G. Leonardi, "Analysis of transport modes in the urban environment: an application for a sustainable mobility system," (in English), *Sustainable City Iv : Urban Regeneration and Sustainability,* vol. 93, pp. 637-+, 2006.
[32] S. Basbas, "Sustainable urban mobility: the role of bus priority measures," (in English), *Sustainable Development and Planning III, Vols 1 and 2,* vol. 102, pp. 823-834, 2007.
[33] J. Froehlich *et al.*, "UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits," (in English), *Chi2009: Proceedings of the 27th Annual Chi Conference on Human Factors in Computing Systems, Vols 1-4,* pp. 1043-1052, 2009.
[34] J. Dijs-Elsinga *et al.*, "Choosing a hospital for surgery: the importance of information on quality of care," *Med Decis Making,* vol. 30, no. 5, pp. 544-555, 2010.
[35] A. Andersen, R. Karlsen, and W. Yu, "Green Transportation Choices with IoT and Smart Nudging," in *Handbook of Smart Cities*: Springer, 2018, pp. 331-354.
[36] R. H. Thaler and C. R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.
[37] A. S. Hanks, D. R. Just, L. E. Smith, and B. Wansink, "Healthy convenience: nudging students toward healthier choices in the lunchroom," *J Public Health-Uk,* vol. 34, no. 3, pp. 370-376, 2012.
[38] E. Duflo, M. Kremer, and J. Robinson, "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," (in English), *Am Econ Rev,* vol. 101, no. 6, pp. 2350-2390, Oct 2011.
[39] C. R. Sunstein, "Nudging Smokers," (in English), *New Engl J Med,* vol. 372, no. 22, pp. 2150-2151, May 28 2015.
[40] S. Cohen, "Nudging and Informed Consent," (in English), *Am J Bioethics,* vol. 13, no. 6, pp. 3-11, Jun 1 2013.
[41] J. S. S. Blumenthal-Barby, "On Nudging and Informed ConsentFour Key Undefended Premises," (in English), *Am J Bioethics,* vol. 13, no. 6, pp. 31-33, Jun 1 2013.
[42] E. Bell, V. Dubljevic, and E. Racine, "Nudging Without Ethical Fudging: Clarifying Physician Obligations to Avoid Ethical Compromise," (in English), *Am J Bioethics,* vol. 13, no. 6, pp. 18-19, Jun 1 2013.
[43] H. Bruns, E. Kantorowicz-Reznichenko, K. Klement, M. L. Jonsson, and B. Rahali, "Can nudges be transparent and yet effective?," (in English), *J Econ Psychol,* vol. 65, pp. 41-59, Apr 2018.
[44] T. M. Marteau, D. Ogilvie, M. Roland, M. Suhrcke, and M. P. Kelly, "Judging nudging: can nudging improve population health?," (in English), *Bmj Brit Med J,* vol. 342, Jan 25 2011.
[45] M. Weinmann, C. Schneider, and J. vom Brocke, "Digital Nudging," (in English), *Bus Inform Syst Eng+,* vol. 58, no. 6, pp. 433-436, Dec 2016.

[46] A. Aly, S. Macdonald, L. Jarvis, and T. M. Chen, "Introduction to the Special Issue: Terrorist Online Propaganda and Radicalization," (in English), *Stud Confl Terror,* vol. 40, no. 1, pp. 1-9, 2017.

[47] C. Schneider, M. Weinmann, and J. vom Brocke, "Digital Nudging: Guiding Online User Choices through Interface Design," (in English), *Commun Acm,* vol. 61, no. 7, pp. 67-73, Jul 2018.

[48] E. Dixon, E. Enos, and S. Brodmerkle, "A/B testing of a webpage," ed: Google Patents, 2011.

[49] T. W. Leong, R. Harper, and T. Regan, "Nudging towards serendipity: a case with personal digital photos," in *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, 2011: British Computer Society, pp. 385-394.

[50] N. Huang, P. Chen, Y. Hong, and S. Wu, "Digital Nudging for Online Social Sharing: Evidence from A Randomized Field Experiment," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[51] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile networks and applications,* vol. 19, no. 2, pp. 171-209, 2014.

[52] J. Manyika *et al.*, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[53] M. F. Uddin and N. Gupta, "Seven V's of Big Data understanding Big Data to extract value," in *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education*, 2014: IEEE, pp. 1-5.

[54] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *Ieee Access,* vol. 2, pp. 652-687, 2014.

[55] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of “big data” on cloud computing: Review and open research issues," *Information systems,* vol. 47, pp. 98-115, 2015.

[56] P. Russom, "Big data analytics," *TDWI best practices report, fourth quarter,* vol. 19, no. 4, pp. 1-34, 2011.

[57] V. Rajaraman, "Big Data Analytics," (in English), *Resonance,* vol. 21, no. 8, pp. 695-716, Aug 2016.

[58] A. K. Pujari, *Data mining techniques*. Universities press, 2001.

[59] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[60] M. Rouse. (2019, May 2019). *OLAP (online analytical processing)* [Online]. Available: https://searchdatamanagement.techtarget.com/definition/OLAP.

[61] E. Rahm and H. H. Do, "Data cleaning: Problems and current approaches," *IEEE Data Eng. Bull.,* vol. 23, no. 4, pp. 3-13, 2000.

[62] Wikipedia. (2019, May 2019). *Database Normalization* [Online]. Available: https://en.wikipedia.org/wiki/Database_normalization.

[63] A. Silberschatz, H. F. Korth, and S. Sudarshan, *Database system concepts*. McGraw-Hill New York, 1997.

[64] V. Sharma and M. Dave, "Sql and nosql databases," *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 2, no. 8, 2012.

[65] J. Han, E. Haihong, G. Le, and J. Du, "Survey on NoSQL database," in *2011 6th international conference on pervasive computing and applications*, 2011: IEEE, pp. 363-366.

[66] J. Pokorny, "NoSQL databases: a step to database scalability in web environment," *International Journal of Web Information Systems,* vol. 9, no. 1, pp. 69-82, 2013.

[67] A. Pavlo and M. Aslett, "What's really new with NewSQL?," *ACM Sigmod Record,* vol. 45, no. 2, pp. 45-55, 2016.

[68] Interreg. (2014, May 2019). *Mobility choices app* [Online]. Available: https://www.interreg.org/aktuell/mobility-choices-app.

[69] J. Strozinsky. (2017, May 2019). *A better day the 100 way* [Online]. Available: https://eingutertag.org/en/the-idea.html.

[70] A. APIs. (2017, May 2019). *Norway weather* [Online]. Available: https://www.accuweather.com/en/no/norway-weather.

[71] T. fylkestraffik. (2019, May 2019). [Online]. Available: https://www.tromskortet.no/?lang=no_NO.
[72] C. Radu, "Data Management and Nudging for Green Transportation," Master's degree, Computer Science, UiT, Norway, 2019.
[73] S. Gil and L. Le Bigot, "Seeing life through positive-tinted glasses: color–meaning associations," *Plos One*, vol. 9, no. 8, p. e104291, 2014.
[74] Microsoft. (2019, May 2019). *Visual Studio 2019* [Online]. Available: https://visualstudio.microsoft.com/vs/.
[75] Wikipedia. (2019, May 2019). *.NET Core* [Online]. Available: https://en.wikipedia.org/wiki/.NET_Core.
[76] Microsoft. (May 2019). *C# Guide* [Online]. Available: https://docs.microsoft.com/en-us/dotnet/csharp/.
[77] Wikipedia. (2016, May 2019). *Angular (web framework)* [Online]. Available: https://en.wikipedia.org/wiki/Angular_(web_framework).
[78] Wikipedia. (2012, May 2019). *TypeScript* [Online]. Available: https://en.wikipedia.org/wiki/TypeScript.
[79] W3C. (2017, May 2019). *HTML 5.2* [Online]. Available: https://www.w3.org/TR/html52/.
[80] Mozilla. (2019, May 2019). *CSS: Cascading Style Sheets* [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/CSS.
[81] Bootstrap. (2019, mAY 2019). *Bootstrap* [Online]. Available: https://getbootstrap.com/.
[82] Microsoft. (2017, May 2019). *What you'll love about SQL Server 2017* [Online]. Available: https://www.microsoft.com/en-us/sql-server/sql-server-2017.
[83] Oracle. (2019). *Oracle Database 18c (18.3)* [Online]. Available: https://www.oracle.com/technetwork/database/enterprise-edition/downloads/oracle18c-windows-180000-5066774.html.
[84] draw.io. (2012, May 2019). [Online]. Available: https://www.draw.io/.
[85] W3scools. (May 2019). *What is HTTP?* [Online]. Available: https://www.w3schools.com/whatis/whatis_http.asp.
[86] Amazon. (2019). *AWS* [Online]. Available: https://aws.amazon.com/.
[87] E. Anagnostopoulou, E. Bothos, B. Magoutas, G. Mentzas, and A. J. I. W. P. P. T. Stibe, "How to not be Annoying: Adjusting Persuasive Interventions Intensity when Nudging for Sustainable Travel Choices," 2018.
[88] P. Lally, C. H. Van Jaarsveld, H. W. Potts, and J. J. E. j. o. s. p. Wardle, "How are habits formed: Modelling habit formation in the real world," vol. 40, no. 6, pp. 998-1009, 2010.
[89] K.-H. Kim, E. Kabir, and S. Kabir, "A review on the human health impact of airborne particulate matter," *Environment international*, vol. 74, pp. 136-143, 2015.
[90] R. J. Shephard, "Is active commuting the answer to population health?," *Sports medicine*, vol. 38, no. 9, pp. 751-758, 2008.
[91] C. Kennedy, E. Miller, A. Shalaby, H. Maclean, and J. Coleman, "The four pillars of sustainable urban transportation," *Transport Reviews*, vol. 25, no. 4, pp. 393-414, 2005.
[92] C. Fischer, "Feedback on household electricity consumption: a tool for saving energy?," *Energy efficiency*, vol. 1, no. 1, pp. 79-104, 2008.