DOCTORAL THESIS

A Study on Predictive Data Modeling towards Rural Mobility of Japan for the Sustainable Revitalization of Regional Tourism Activities

By
Ali Bakdur

Advised by:
Fumito Masui
Michal Ptaszynski

KITAMI INSTITUTE OF TECHNOLOGY
GRADUATE SCHOOL OF ENGINEERING

September 2021

ABSTRACT

A methodological framework assesses tourism as a multidisciplinary area of study, presenting different variability, performance, the ritualization of practices, and mediation between participants (Ritchie, B. et al., 2005). Influenced by this brief explanation, this thesis attempts to identify organizational responses to the role of on-demand mobility in Japan's tourism and sustainable rural societies.

Therefore, for the simplicity of explanation, while the scientific concepts of the thesis expand, the initial thematic research phases are divided into two parts: First, to explain the current status of rural mobility of Japan as a happening caused by human activities from the conceptual model of tourism theory's point of view. Second, to predict the fluctuation of prospective on-demand mobility while employing all explicable and obtainable kinds of data within the statistical decision theory applicable domain.

Simultaneously, expanded ideas and methods used in this thesis have inspired an intellectual stimulation enriched with the discoveries of previous social science and informatics research towards emphasizing public transportation as an instrument of the social welfare state for providing mobility to the sustainability of regional communities in Japan.

Although the philosophy of science has principally recognized the distinction between explanation and prediction, there will be significant benefits in using them appropriately in combination with modern scientific methods, especially in reevaluating social issues related to changing conditions (Shmueli, 2010).

In general, regardless of a country's economic diversity, an accepted opinion is that a higher education institute located in a rural area is an indispensable component of regional development and revitalization models by almost any nation due to university students' promotion of mobility. Moreover, adapting to the world's ever-changing values due to globalization policies with suitable regional revitalization models requires solid interaction, communication, and cooperation between higher education institutions, government agencies, and commercial enterprises.

Considering the number of interdisciplinary studies in the literature that try to expose the role of universities in promoting rural mobility as an initiative from different angles, there is limited research focused on students' behaviors, motivations, and challenges (Toprak, 2018).

Perhaps, a hidden barrier of this situation practically occurred because university students, especially in local cities, are often accepted as temporary visitors or tourists from the dominant perception of the host communities (Ergun, 2014). Thus, apart from the importance of a university as an institution in developing and revitalizing a city, university students are neglected in many cases even though they play a prime agent role in economically sustaining various local business establishments.

Therefore, the expanding frontier of this thesis aims to assist in the determination of the regional revitalization policies of Japan within the norms and understandings of international counterparts' strategies. Furthermore, while developing a multidisciplinary integrated approach to assess the impacts of public transportation on community development and sustainable livelihoods, the systems of tourism theory were acknowledged as a primary framework of this thesis to disclose the benefits to the topic during the conduction of the study's experiments (Lohmann, G. and A, Netto., 2017).

The specific community defined in this study was Kitami, a rural city in Japan, and university students who enrolled in Kitami Institute of Technology. Kitami is in the Okhotsk subprefecture of Hokkaido, and about 1,800 students have enrolled in the institute.

The discussion began with an exploratory data analysis of the existing condition in rural areas from the tourism point of view, using internet-based data sources. According to the preliminary data analysis, some of the highlighted results are:

Japan is a unitary island country with a 126.8 million population and one of the most urbanized nations in the OECD (OECD, 2016). Moreover, the population is rapidly aging and has been declining since 2011 (OECD, 2016). Despite nearly 30 years of slow economic growth, the economy is the world's fourth-largest (World Bank, 2021). The economic value created by the tourism industry is exceptional, and the extraordinary contribution of this value mainly comes from domestic expenditures.

For instance, tourism incomes contributed 7.1% of Japan's GDP in 2019, and 81% of this contribution has formed by domestic expenditures (WTTC, 2021).

While domestic tourism expenditures constitute the backbone of Japan's tourism economy, domestic tourism's contribution to GDP has steadily declined since 1998 (JTA, 2016). Furthermore, statistics reveal that the people's traveling trend for touristic leisure inside the country has also declined between 1986 and 2006 (Public Relations Office, 2015) (Public Relations Office, 2018).

Perhaps many reasons prevent the Japanese from traveling for leisure. Nevertheless, comparing only past and present data values may not be the best way to evaluate this occurrence properly. A better understanding of the differentiation in the popular trends of the Japanese lifestyle will have appeared with a precise interpretation and inference of the actual situation of people's daily routines and leisure perspectives.

Inevitably, nothing is static in real life, and there are always unexpected events, which may be the most prominent obstacle to predictive studies' accuracy. Moreover, the parameters of societies themselves are continuously transforming; therefore, the explanation, which tries to comprehend the consequences of happened events, seems to have the upper hand over the prediction, which focuses on estimating the future of the events that have not happened yet. In this context, evaluating these two concepts with an understanding that they complement each other, the organized experiments of this study carried out in two settings:

A preliminary experiment presented in this thesis was a comprehensive big data analytics on the demographic structure of the Okhotsk subprefecture as a representative sample of rural regions of Japan. As a result, the anomalies and sustainable development points from the findings for this rural area were used as probabilistic modeling components within the tourism point of view.

The necessity of interpreting multiple temporal trends with several parameters due to the structure of the experiment enabled this thesis to utilize probability theory as a secondary thematic framework. Principally, Bayesian inference was characterized to improve the conceptual quality of probability from the standpoint of the specific conditions

of the scenario. Technically, a Markov chain Monte Carlo simulation was used parallel with the Bayesian statistic to improve the probabilistic model's reliability. Finally, the outcome was illustrated graphically and evaluated in terms of likelihood.

This thesis's main experiment focused on the probable influence of notorious public bus services (PBS) in rural Japan to revitalize declining social life activities by utilizing an opinion survey to gather data from a community sample. The practical applicability of statistical decision theory on the machine learning domain was appropriate to the tertiary thematic framework of this thesis due to the collected data structure of the experiment.

The accessibility and economic sustainability of PBS have been neglected in Japan's rural regions due to several reasons for many years as a finding from the previous experiment. Besides, either associated with the general hardship of living in the rural areas or not, rural cities are also negatively affecting by the migration, especially among young people, to urban areas experiencing a rapid economic expansion.

The empirical topic of the experiment was chosen as the current and prospective status of PBS in Kitami within the limits of on-demand mobility. The method was to estimate the fluctuation of future on-demand mobility as an entry point to explore the impact of the challenges of living in a rural city in terms of university student's daily life needs. Furthermore, the analyzed inquiry aimed to evaluate the weight of university students' daily needs on the quality improvement of these public services as potential beneficiaries.

As a result, with deliberately selected a two-step verification probe from the peculiarities of the Kitami Institute of Technology students' community, while the initial demand rate for PBS was around 60%, this score increased to 71% in the second confirmation. Moreover, the expected demand rate was over 90% of the F-measure after employing machine learning-based prediction methods to the collected data. Still, the most trustworthy and steady forecast approach accounted for 82% with the weight of other daily living variables.

Finally, the approach provided actual proof regarding the usability of the selected methodology to obtain appropriate data and analyze it correctly.

In conclusion, the outcomes of this unique case-based study are expected to assist local governments and relevant private initiatives in developing foresight and reducing the time it takes to make their decisions. The scientific contentment derived from the compilation of this thesis can be outlined as a humanitarian aspiration for the flexibility to assemble reliable and sustainable responses against the ever-changing requirements of social life.

Key-words: Regional development, university-cities, tourism, mobility-on-demand, student decision, public bus services, descriptive analysis, predictive analysis

ABSTRACT IN JAPANESE (論文内容の要旨)

方法論的枠組みにおいて、「観光」は、当事者間における種々の行動や変動性、習慣、調停などを扱う学際的研究分野とみなすことができる (Ritchie, B. et al., 2005)。

本論文では、この見地に基づき、日本の地域社会における観光と持続可能性に対して、オンデマンドモビリティに期待される役割を明らかにすることを試みる。社会科学および情報科学分野において得られた複数の成果を組み合わせて拡張することで、日本の地域社会の持続可能社会に資するモビリティを提供する社会福祉的手段としての公共交通の可能性について考察する。

議論を簡単にするために、本研究の主要部分を二つに区別して議論を進める．第一に、日本の地域モビリティの現状を、観光理論の概念モデルを用いて説明することを試みる。第二に、説明可能かつ入手可能なあらゆる種類のデータに基づくケーススタディにより、統計的意思決定論によってオンデマンドモビリティの将来的変動予測を試みる。

科学哲学はこれまで説明と予測を区別してきた。この立場に現代科学の客観論的アプローチを持ち込むことは流動性の高い社会問題を客観的に評価する上で大きなメリットとなる [Shmueli, 2010]。

一般に、地域社会の高等教育機関は、国家の経済的多様性に関係なく学生の流動性を促進させる地域開発・活性化モデルにとって欠かせない要素であることが知られている。さらに、グローバル化政策によって変化し続ける世界の価値観に適応し、適切な地域活性化モデルを構築するためには、高等教育機関と政府機関、商業企業との間で緊密な交流と連携が必要となる。

近年、観光情報学という学問領域が注目されるようになっている。これは、社会学的観点による分析研究が中心であった観光に関わる処処の系統的研究に対して、情報科学的アプローチを持ち込むことで新たな切り口からの客観的解

釈を与えたり、データ分析に基づく定量的解釈を与える研究や ICT 応用による技術支援の提案やそれらが観光分野にもたらす効果を議論する分野である。

本論文は、観光情報学に分類される研究である。統計的アプローチと人工知能技術を用いてまとまったデータを精密に分析することにより、日本の地方観光の課題を、巨視的視点からの分析と微視的視点から確率的に説明するとともに、その持続可能性を予測できることを明らかにした。

第一に、OECD が公開している世界各国の経済活動に関するビッグデータを用い、ベイズ統計とモンテカルロシミュレーションを組み合わせて適用することによって、観光の側面から日本の地方経済の現状を確率的にモデリングし、分析を行なった。

第二に、大学生を対象としたアンケートによって地方公共バス運行サービスの利用動向と同サービスに対する意識調査を行い、得られた回答データを詳細に統計解析することで、大学生が持つ公共バス運行サービスへの潜在的需要と持続的変化を明らかにした。

そして、上記課題について考察を行うとともに地域観光産業に与える影響を考察した。本研究の具体的アプローチは確率モデリングによる分析と予測が中心であるが、加えて、人工知能アプローチに基づく定量分析・予測手法を提案し、統計的モデリングとの比較によって同手法が需要率の変化を十分予測できる性能を持つことも明らかにした。

分析結果に対する考察において、将来の持続可能性（サスティナビリティ）に関連する主要な課題を定量的に明らかにした点は、観光を通した地域経済活動に重要な示唆を与えるとともに、観光情報学分野ならびに社会科学的にも貢献してすること大なるものがある。

地方公共モビリティの効果についてはこれまで定量的研究は進んでいなかったため、本研究によって得られた知見は今後の持続可能地域社会を考える上で重要な示唆を与えるものである。

"Misfortune is a kind of talisman whose virtue consists in its power to confirm our original nature; in some men, it increases their distrust and malignancy, just as it improves the goodness of those who have a kind heart."[1]

[1] Balzac, H.D. Le Colonel Chabert.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS

Journal Papers

I. Bakdur, A., Masui, F., & Ptaszynski, M. (2021). Predicting Increase in Demand for Public Buses in University Students Daily Life Needs: Case Study Based on a City in Japan. Sustainability, 13, 5137.

II. Bakdur, A., Masui, F., & Ptaszynski, M. (2018). Big data analytics - towards the enrichment of content tourism for revitalization of Japanese rural area. MATEC Web Conf., 169 (2018) 01008.

Book Chapter

I. Bakdur, A., Masui, F., & Ptaszynski, M. (2020). Predicting University Students' Public Transport Preferences for Sustainability Improvement. Advances in Intelligent Systems and Computing book series (AISC, volume 1251).

ACKNOWLEDGMENTS

Compilation of this academic study would not have been achievable without the universal moral support of some righteous people. Therefore, a significant debt of gratitude is under my obligation towards them.

Foremost, I would like to express my gratefulness to Professor Fumito Masui and Professor Michal Ptaszynski for allowing me to return to their research laboratory at Kitami Institute of Technology after a long hiatus from formal schooling. For their time, generosity, and oversight, I shall always be thankful.

Following that, I would like to thank all of the Kitami Institute of Technology students who took part in the survey field study. Their earnest assistance in eliciting their own opinions enabled me to gather crucial data for this analysis.

Finally, thanks to the inhabitants of Kitami-shi for providing me with the necessary strength and encouragement and several friends who endured this long process with me by offering their thoughts and continuous criticism.

Each one of these contributions is much appreciated.

Ali Bakdur

September 2021

CHAPTER ONE

THE RESEARCH PROBLEM AND ITS BACKGROUND

1.1 The engagement of components: University-cities, tourism, and students' mobility

Although there is no fully coherent theory of the origins and sustainability of cities and other human societies, there has been much research into it from several angles - three streams of theory attempt to explain cities as spatial phenomena (Lynch, 2008).

- "*Functional Theory*" focuses primarily on cities' existence since it tries to explain why a city assumes the shape and how it works.
- "*Normative Theory*" is concerned with the generalizable relationships between human values and settlement form, such as recognizing a well-organized city when seen one.
- "*Planning Theory*" asserts how difficult it must be to give a public decision regarding city development due to the differentiation of opinions by the multiple parties' who assemble the entire city population.

In particular, planning theory's domain extends beyond the arena of city planning. It has been extensively developed in those other fields, as these understandings apply to all complex political and economic operations. As a result, the "*Planning Theory*" has also known as "*Decision Theory.*"

As defined by Kevin Lynch, a good city form is an urban design idea focused on the relationship between human values and the physical location of a city. However, few models reflect the emergence of a city through the deliberate actions of people and small groups and that people's ability to learn evolves. In reality, especially rural towns often impose their own "*status quo*" rules, assuming some fixed laws rather than upholding the spirit of human productivity.

Nowadays, due to substantial economic expansion, rural areas are experiencing a population migration into industrial centers. The consequences of these occurrences became one of the research fields frequently studied within different scientific disciplines: Rural regions' depopulation trend (Kato, 2014).

The reasons for this tendency of rural cities' residents can practically be summarized: Migration to developed urban cities due to the relative difficulties of living in rural towns, economic expectations, and decreased birth rates in developed countries.

Essentially, perhaps one to consider most, rather than living in a rural city with limited opportunities, a desire to reach more freedom to improve personal comfort zone in the innovative environments of larger cities.

In general, as rural areas' mobility stagnated steadily year by year, academic and government-based institutions started researching to reveal the reasons for this occurrence and the reliable solutions to revitalize regional communities and ensure long-term population sustainability in Japan's rural cities (Chen, B. et al., 2018). At the same time, various regional studies in developed and developing countries have also examined the impact of tourism on the revitalization of rural areas suffering from different socio-economic insufficiencies (Briedenhann, J. and E. Wickens., 2004) (Wilson, S. et al., 2001) (Long, P. and Lane, B., 2000).

As part of this type of endeavor, an extensive study focused on universities' participation in tourism activities, especially describing the tourism consumption of university students to understand the university's role as an integrating element of society (Albino, 2015). Furthermore, another study's objective was to identify the potential contribution of university students to the development of environmentally and socially sustainable communities (Brennan, J. and A, Cochrane., 2019).

Universities are also under increasing pressure to become more globally oriented and provide their students with a global experience associated with their communities' openness.

On the international students' side, selective motivations for participating in an educational exchange include the desire to travel and the possibility for fun or excitement and the host country's climate, natural surroundings, hospitality, and tourism attractions (Llewellyn-Smith, C. and V. McCabe., 2008).

Moreover, university students desire to participate in their host cities' social and cultural lives rather than isolating themselves in on-campus areas (Eder, J. et al., 2010).

From the passenger transportation perspective, historically, transportation began to play a crucial part in developing the economies and globalization of the cultures starting from the second half of the 1950s.

As a result of people's movement, tourism has grown and diversified in response to increased travel habits worldwide over the last six decades; thus, tourism has become one of the world's most populous and rapidly expanding economic sectors. Although there are several technical definitions of tourism, at this point, the applied purpose of tourism is considering the economy as a primary parameter:

"The science, art, and business of attracting and transporting visitors and accommodating and gracefully catering to their requirements and wants are defined as tourism." [2]

An economic approach to the definition of tourism would always be more appropriate for Japanese society (JTA, 2020). The tourism industry formed 7.4% of the total GDP with a direct 38 Trillion JPY contribution in 2016. Domestic spending accounts for 86% of the tourism industry's overall contribution to GDP.

The comparison with the subsidizing of the automotive manufacturing industry in the same year, which accounted for 16.5%, can be demonstrated better the performance of this contribution (WTTC, 2017). Besides, tourism is more environmentally friendly than other manufacturing sectors (UNWTO, 2016).

In conclusion, the formation idea of a city is rationalizing by two denominators: A consensus of different groups of people on living together and geographic reality. Therefore, this thesis focused on university students, prominent members of the people groups who accepted to live in a city, and examined their potential contribution to environmentally and socially sustainable communities by increasing the on-demand mobility opportunities. In addition, it also looked at the impact of tourism on the revitalization of rural areas suffering from different socio-economic insufficiencies in rural Japan was.

[2] R. W. McIntosh. (1977). Tourism: Principles, Practices, Philosophies (2nd Ed.). Columbus, Ohio: Grid Publishing.

1.2 Introduction of the research problem

In Japan's countryside, public bus services (PBS) have steadily deteriorated in terms of accessibility and economic sustainability (Dominik, L. and M, Avary., 2020). For instance, PBS consumption has decreased by roughly 36% over the last 20 years, except for the major urban centers (Kawazaki, M. and Luczak, D., 2020).

Previous research on passenger transportation examined the definition of services, performance, management, the role of private providers, increased efficiency, pricing, and systematic travel user-demand supervision (Schofer, 1983) (Walters, 1979). Modern academic and private studies emphasize passenger happiness to encourage more public transportation by offering advanced passenger transportation systems based on individual transportation preferences (Sakai, 2020) (Hoerler, R. et al., 2020).

On the other hand, most of the data in the field of transportation, which will be helpful for prospective studies to examine mobility-on-demand in rural Japan, come from official statistics compiled by the government or local authorities (Kawasaki, 2015). In addition, user surveys are a standard method to collect data in significant cities, which frequently specifies the rationale for people's movements, starting and ending points and modes of transportation. However, democratizing and expanding passenger transportation data usage, especially in rural areas, to the researchers may give policymakers a more solid foundation for responding flexibly to ever-changing transportation demands (Dickens, M. and M. Hughes-Cromwick, 2019).

Unlike metropolitan cities in Japan, rural towns offer more limited access to PBS; therefore, transportation has over time become primarily dependent on personal cars in the rural areas for a variety of economic reasons.

However, the availability of PBS should not have been neglected in long-term financial prospects, even though most local Japanese adults prefer to drive their cars due to the consequences of economic reasons. Especially today, the other members of the communities who do not have instant access to an individual car experience several difficulties in mobility. Among them, at first glance, students, seniors whose numbers are increasing rapidly, especially in rural areas, non-permanent tenants, and visitors.

Eventually, a personal car-based character of public transportation makes it unsuitable for outsiders but only for locals. Furthermore, a rapidly aging society and migration of young generation to metropolitan areas created a big gap for middle-level jobs in rural communities of Japan. There are several attempts to compensate for this diminished workforce need with foreign laborers. However, foreign laborers (outsiders) will inevitably experience transportation problems until they acquire their cars, resulting from the current public transportation reality.

Beyond that, the rising trend of young people's "*inward-looking orientation*" compels researchers to focus on deeper psychological reasons for the youth population's social expectations of life (Nghiêm-Phú, 2015). While the government addresses social interaction as one of the essential aspects in determining the quality of a student's experience, most universities' administration places a greater emphasis on academics than cultural adjustment (Takamatsu, R. et al., 2021).

As a result, these are some of the significant issues that threaten the long-term sustainability of the young populations of rural communities.

Considering public transportation as one of the basic requirements of daily mobility in urban life, it is a surprise to confirm that there has been no practical solution provided to improve the quality of public transportation services in rural cities of Japan during the last few decades. In such an environment, this study sought to reveal the critical priorities of modern mobility from the daily life needs of rural city inhabitants of Japan.

On-demand PBS analysis was the entry point of the research problem, and as a selective population sample was university students. Although the functional findings are specific to only this local community, the outcomes of this research, when contributed with more similar studies focused on regional mobility in the future, then provide an informative insight base knowledge for improving the quality of life in any community in Japan.

Therefore, the motivation for compiling this thesis was to increase the diversity of transportation data in Japan and explore an individual geographic region's various conditions and people's expectations to promote acceptable solutions for sustainable communities. Moreover, a practical interpretation of the highlighted findings from this

thesis may also increase the young population's sustainability in rural areas of Japan if appropriately responded to by local authorities.

Therefore, this thesis's scientific novelty and contribution to society stem from the desire to accurately describe and present the simple elements of daily life that affect people's on-demand mobility using an opinion survey that consists of a specific set of questions in the appropriate order.

1.3 The geographical background of the studied area

Kitami (Figure 1.1) is the largest city of the Okhotsk subprefecture located in northern Japan. The city's estimated population is over one hundred thousand people, with individuals dispersed sparsely throughout its neighborhoods and winter months primarily being cold and snowy (JNTO, 2021).

Figure 1. 1 Location of Kitami city within Japan

The Sea of Okhotsk is approximately 40 - 45 kilometers north of the city. The arctic nature of the sea creates a widely recognized natural ice drift observed off the shore during the winter season.

Agriculture is a prominent industry in the region, with Kitami being the largest onion producer in Japan. The town is also well-known for its association with winter sports, particularly curling.

Kitami is hosting a national university institute and a private nursing college. Of the technology institute students, only 35% originate from Hokkaido prefecture and have a female-to-male ratio of 14% (KIT, 2019). The campus area of the institute is 2.5 kilometers far from the city center.

The distance between Sapporo[3] and Kitami is around 300 kilometers. Both bus and train services run daily; however, reservations are required in advance. The city is served by Memanbetsu Airport, which is about 20 kilometers from Kitami.

In addition, Kitami is heavily reliant on automobiles for transportation[4] (AIRIA, 2021), which is much higher than the average of the whole Hokkaido prefecture[5] (Statistics Japan, 2021). Therefore, public bus services (PBS) offered frequency is relatively limited; for instance, some bus lines only run 3 or 4 times per day. Meantime, the city's main bus line operates four times an hour and ends around 21:00 on weekdays and 20:00 on weekends (KitaBus, 2021).

Although a society with a dropping birthrate, such as Japan, does not make sense in large cities, it has a massive effect on the local economy in rural communities like Kitami.

As a result, it is critical to engage these young people's resources in this community from mutual benefits aspect with many ways to help the local economy. The concept encompasses economic elements, such as purchasing from local stores, and environmental concerns, such as opting for PBS over private automobiles.

[3] The largest city of Hokkaido.
[4] 152.4 automobiles per 100 people.
[5] 68.67 automobiles per 100 people.

1.4 Aims and objectives of the study

Before starting to evaluate the situation in terms of informatics, it would be helpful to mentioned common mistakes associated with developing informatics-based real-time solutions. These mistakes diverge from the structure of problems, high-level project management methodologies, and poor decisions on low-level technical issues (Stewart, 1999). Eventually, design and implementation are affected highly by these mistakes in real-world cases.

Even though some of the commonly identified mistakes are under-defining system requirements, focusing too much on tool features, ineffective communication between parties, and excluding end-users from decision-making steps. Perhaps the most obvious of these mistakes is the effort to design informatics-based projects that provide solutions to all problems at once, whether to increase social welfare or meet business needs. The impatience of corporate administrations from an extreme perfectionist approach is often the cause of this situation. As a result, many software projects fail in the development stage due to increased costs and unachievable performance expectations. Therefore, when developing a real-time system, instead of focusing on solving all problems with a single product offered, it would be more appropriate to deploy functional modules one by one by testing their functionality with execution time at each step.

Although it is never easy to analyze the already complex social issues at once, it would still be a more objective approach to divide these issues into pieces and examine them orderly. In addition, the findings obtained through the analyses of a specific local region reality would be helpful to be compared with the statistical data at the country level in each step of the used methods. In this way, the proposal solution project for macro-level problems can train from the comprehensive findings of the micro-level survey results.

Therefore, a diagram (Figure 1.2) presents the theories that were the basis for this study and the relationship between their practical application methods. Thus, the practicality of this thesis is designed with a strategy that can provide solid analysis-based insight to the topics mentioned in the research problem in terms of informatics and humanities/social sciences.

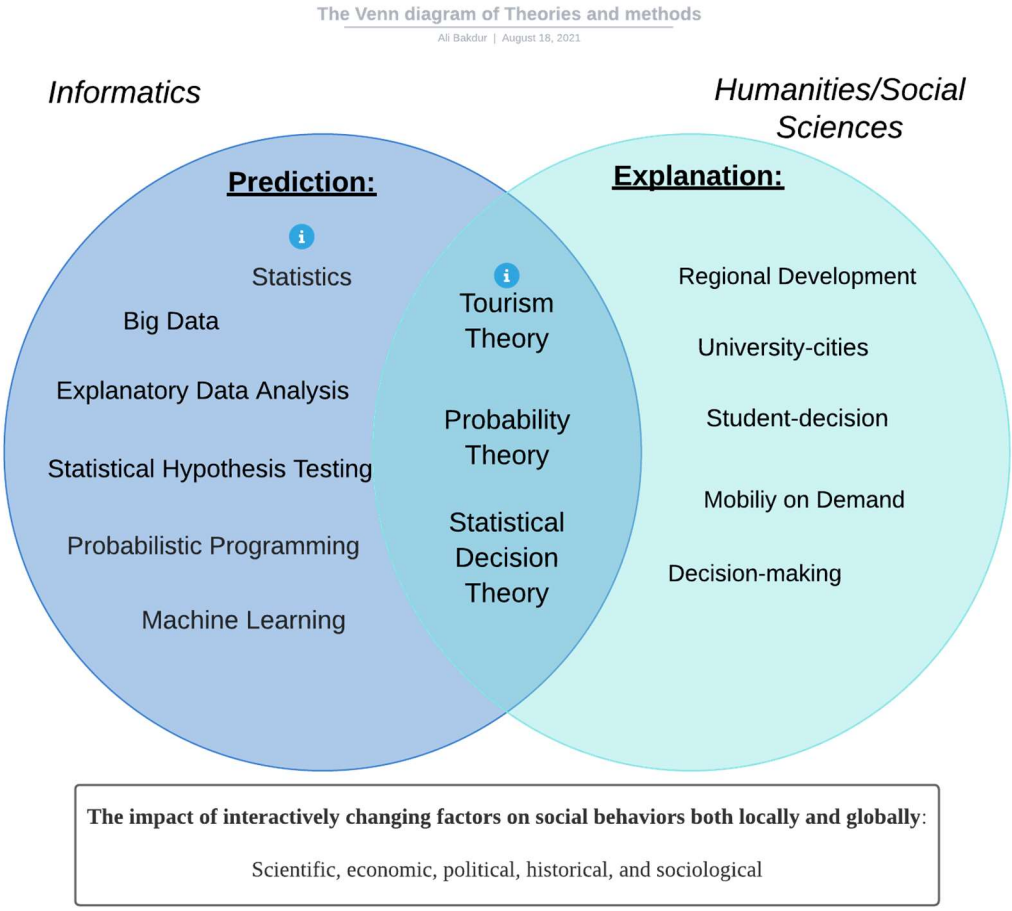


Figure 1. 2 The diagram of used theories and methods

The systems of tourism theory were acknowledged as a primary framework of this thesis. In addition, the idea also utilizes probability theory as a secondary thematic framework. Finally, the practical applicability of statistical decision theory on the machine learning domain was appropriate to the tertiary thematic framework of this thesis.

Moreover, this methodological order is also eager to solve many socio-cultural and economic issues of any society. Therefore, the steps of this study are summarized in the following order:

First, a detailed situation analysis of rural reality was conducted to correctly define the origin of problems caused by social order from the internet-based sources' data. In addition, previous studies on local regions of Japan were reviewed in the literature. As a

result, the initial step of the approach utilized big data and probabilistic programming features to illustrate the situation which has not been tried before in this way.

Second, as a result of the discussions made within the situation analysis results, the decided entry point to measure the current interest of this local community was the public bus services, which many previous studies have also examined its structural deficiency. It was also interesting that the current situation of public bus services had been going on for many years in the same condition and seemed to be accepted normal by local people. This situation can be defined as social unresponsiveness, which was also examined inside this thesis.

Third, due to escape from the mentioned traps in developing a socially benefited informatics-based project, this thesis tried to reveal the importance of reliable public bus services in rural areas of Japan as a functional unit of the empirical approaches to promote rural on-demand mobility. Because to move on to the more detailed features of any model, it was necessary first to determine whether the focused community is interested in any improvement on the main topic or not. Furthermore, there would be no point in advancing the study if there was an aversion. This condition was tested with multiple methods during the process.

Finally, at the same time, the objective was to identify the obstacles accurately and reveal the topics that need improvement in the mobility arena for sustainable communities from the tourism perspective. However, on the other hand, it is helpful to carry out such studies periodically, the usability of which is tested by experiments, to evaluate the progression in societies mindset changes.

1.5 The hypotheses of the study

The methodology offers a properly tested combination of sophisticated statistical techniques to interpret the young individual opinions in their aspect. From a reasonable perspective, the findings will be helpful to aid local government decision-makers in reducing adaptation time and improve flexibility in response to student’s on-demand mobility.

This thesis' defined hypotheses to clarify with an experiment regarding mentioned PBS issues were as following order:

- "*To define the impact of daily life needs that will significantly effect on university students' demand for PBS as prospective riders in a rural city.*"
- "*To verify the hypothesis that if an improvement occurs in PBS based on determined by user requirement, the demand on the user side will change.*"
- "*To purpose a basic statistics-based inferential modeling that estimates how much the demand rate fluctuates on the user side if there is an improvement in PBS.*"

Intuitively, decision-makers at all levels show a strong bias against alternatives that aim to modify the long-standing "*status quo*" (Hammond, S. et al., 1998). Therefore, a two-step verification probe was utilized to avoid this trap and reveal a person's actual perceived viewpoint in the question sequence of the survey applied.

1.6 Structure of the thesis

With its governance and strategy dimensions, this thesis is organized into five chapters that emphasize the potential of public transportation services to promote tourism activities and awareness as attractive factors for university students, tourists, and local cities in Japan.

The first chapter, "*The Research Problem and Its Background,*" started with a discussion of the backdrop and nature of the research challenges; the relationship of a city, university, students, and tourism were highlighted as the events' common denominator due to globalization trends affecting towns and universities.

The second chapter, "*Literature Review,*" was separated into three thematic studies: First, an identified variations in modern Japanese living standards and public transportation understanding were in the literature. Second, an appointed summary of the data analysis over sustainable travel and tourism development on a global and Japanese level with comparative reasoning was for the last five years. Third, an examination in the literature of university students' mobility behavior, alternative transportation, university-city

interactions was. In addition, a presentation of the previous studies on big data and machine learning techniques were.

The first and second chapters establish a conceptual link between the university-city and tourism relationship; furthermore, university students participate in this relationship with their daily lives as active agents.

The third chapter, "*Research Methodology,*" was introduced the worldview of this thesis and the reasons for adopting status analysis and case study methodologies as part of the research process. In addition, a brief abstract of the development of relational data, statistical decision theory, statistical hypothesis testing, probabilistic programming, and machine learning principles have been placed among the explained subsections to the technological research infrastructure.

The fourth chapter, "*Experiments,*" was divided into a preliminary experiment and the main experiment:

In a preliminary experiment, the current demographic situation of the region focused on was examined with the application of exploratory big data analytics.

In the main experiment, an inductive research design method was implemented within the machine learning application domain, using participant observation and administering a scoping questionnaire to university students.

The conclusion of the research issues is addressed in the last chapter, "*Key Findings, Conclusions, and Implications.*" The discussion in this chapter were: the consequences of unresponsiveness to an economically developed society, a higher education institute administration prospect, a collaborative framework of university-city interactions, and public transportation demands of university students. It also discusses the limits that developed during the research period, as well as prospective research topics.

The key findings and contributions of the research were summarized in this last chapter as well.

CHAPTER TWO

LITERATURE REVIEW

2.1 The situation overview in Japan

In general, a public transportation system is expected to achieve two objectives together or standalone: Public benefit and profit. Nonetheless, these two objectives do not always coincide hand to hand, depending on the circumstances of countries.

The goal of public transportation policy varies depending on the country's social and economic structure and might be either public benefit or profit. Profit-driven, privately owned, and publicly listed mass transit firms operate most public transportation networks in Japan (Calimente, 2012).

To better perceive the course of the objectives of this thesis, it is worth noting that; the Japanese government has adopted the principle of self-sufficiency in public transportation for its unique economic reasons (Shoji, 2001).

Traditionally, Japanese local passenger transportation systems evolved on railway subsidiaries until World War II (Aoki, 2000). However, unlike many other countries, the railway-based transportation sector transitioned to private automobiles in the late 1960s due to rapid economic and industrial growth.

In the late 1980s, many households' incomes increased, and as a result, they could afford to buy more than one auto car. However, the increasing trend of using private vehicles in these new economic circumstances triggered passengers' desertion of the local railway line, and disused lines soon closed due to unprofitability (Okano, 1994). Simultaneously, the increasing popularity of using private vehicles has also decreased the number of people riding public buses (Kakumoto, 1997).

As a result, many local bus companies began to encounter financial difficulties, eventually leading to the sudden elimination of public bus services (PBS) for residents, despite the lack of a replacement (Imashiro, 1996). In general, if the passenger density per inner-city bus line is less than five, the line is considered unprofitable to maintain.

In Japan, as explained above, public transportation is usually a service provided by private initiatives; simultaneously, people prefer private cars, especially in rural areas. Eventually, this condition resulted in insufficient passenger demand for private-sector public transportation services to be profitable. Due to this vicious cycle, PBS never took place as an independent transportation option in the countryside as a requirement of the welfare state.

On the other hand, many prominent services required to maintain lifestyles in rural regions, in addition to transportation including logistics, and retail, are presently provided by private enterprises. However, due to diminishing demand, a lack of workforce to carry out services, and inadequate long-term profitability projections, many service providers face financial challenges forcing them to exit these sectors (Takumi, 2017).

Although it has not yet fully passed to the stage of applicability, while the sparsely distributed rural settlement structures in Japan are becoming costly for sustain, the government agencies have developed a new urban design and city planning concept dubbed "*compact city*." The idea was to address aging and depopulated societies' problems in rural areas based on people's needs (MLIT, 2008).

Meanwhile, the emerging results of information and communication technologies (ICT) perceive as "*an agent liberator*" for improving people's lives, especially in rural towns from the tourism point of view (Tsuyoshi, 2017) (Ukpabi, D. and H, Karjaluoto., 2017).

Furthermore, content analysis studies have shown that travel, tourism, and hospitality industries use ICT in various functional units' applications (Law, R. et al., 2014). However, in parallel with the compact city concept of Japan, these functional units are usually defining as the quantification of improvement change triggered by a new ICT introduction for the environmental effect evaluation (Nakaniwa, 2006).

Although tourism research in Japan is still accepted in its early stages from the international perspective, ICT-based modernization could help revitalize and raise its standards (Akira, 2017). To put it another way, rather than simply supporting tourism with ICT, the stakeholders should recognize that the time has come to research and develop ICT

to innovate the travel industry for tourism (Bethapudi, 2013) (Farkhondehzadeh, A. et al., 2013).

The total contribution of tourism revenues to the economy, on the other hand, is drawing an inflexible efficiency when compared to the vast industrial sector's earnings in Japan. One reason may be that the cliché market characteristics of domestic tourism determine the industry (WTTC, 2018). For instance, the performance of this stereotypical market accounted for 91.8%[6] of total tourism spending in 2014. However, this rate was 81%[7] in 2019.

The significant difference of these two percentage values comes from government agencies' willingly endeavor during the last five years, increasing the inbound spending share in total tourism expenditures from 8.2 % to 19% (WTTC, 2015) (WTTC, 2020).

As a result, Japan has risen from ninth to fourth among one hundred forty nations on the travel and tourism competitiveness index. In the same period, inbound visitor arrivals in Japan climbed from 20 million to 31 million (Weforum, 2015) (Weforum, 2017) (Weforum, 2019).

However, the present results reveal that the growth rate of inbound tourists to Japan is heavily reliant on the Asian market. Sightseeing tourists have a higher number and percentage growth rate than other sorts of visitors. China has surpassed the United States as Japan's most important source of incoming tourists (Tang, 2020). China's emerging middle-class populations reached over 400 million people (Gustafsson, B. et al., 2017). The reality is that those people create a significant share of the international tourism market globally.

China is also the primary source of international students in Japan, accounting for 43.2% of 279,597 international students as of 2020 (JASSO, 2020). Correspondingly, 80,566 Japanese students used their study abroad opportunities in 2019 (JAOS, 2019).

Unfortunately, the COVID-19 outbreak abruptly limited Chinese tourists and students' arrival, and it seems it will not recover soon (H. Kharas and D. Meagan, 2020).

[6] The total contribution of domestic and inbound tourism spending's to GDP was 2.4% (2014).
[7] The total contribution of domestic and inbound tourism spending's to GDP was 7.1% (2019).

In summary, this section discussed the historical evolution of public transportation and the role of private firms, the influence of rapid economic expansion on society, a new government program to reorganize rural areas with ICT's support, and the economic relevance of tourism income.

Finally, travel and transportation can be considered more broadly as tourism features and a phenomenon caused by human activities to adapt to the changing times, requirements, and technology. Therefore, it is also essential to recognize the significance of public passenger transportation to enhance local communities' social welfare within tourism-focused solutions.

2.2 Sustainability development of travel and tourism (2015 - 2019)

The socio-economic perception of travel and tourism in each country or region is influenced by environmental variables, cultural differences, religious ethics, historical evidence, transportation methods, and other elements.

The geographical location significantly impacts the region's tourism potential; Japan is an island country that requires a different assessment than continental countries. Due to their geographical accessibility advantages, Spain and France, for instance, have more tourist attractions and are well-served by all kinds of transportation.

Table 2.1 compares the sustainability of economies dependent on travel and tourism revenue for specific countries (Spain, France, Germany, and Japan). Japan is the only country that has improved its ranking in the last five years, although the first three places have remained. In Japan, the long-term sustainability of travel and tourism development has substantially increased. Japan has the highest population density, lowest unemployment rate, highest CO_2 emissions, and highest GDP among the rivals.

Here is an excellent example of why quality, not quantity, is so important: The tourist cost-performance value in Japan was roughly comparable to Germany's. Despite this, the number of foreign visitors to Japan is less than a third of France. Table 2.1 also presents the importance of retail sales income in tourism service industry spending.

Table 2. 1 Sustainability of travel and tourism in global (2015 - 2019)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2015	1	*ES*	5.3	45	*46M*	22.1	247	*1195B*	*61B*	5.1%	*68M*	*14M*	*54B*	*3M*	14.2
	2	*FR*	5.2	48	*67M*	10.4	300	*2438B*	*95B*	3.9%	*84M*	*27M*	*118B*	*3M*	10.6
	3	*DE*	5.2	44	*82M*	4.6	730	*3361B*	*118B*	3.5%	*35M*	*84M*	*265B*	*5M*	12.1
	9	*JP*	4.9	41	*127M*	3.4	1200	*4389B*	*100B*	2.3%	*20M*	*16M*	*179B*	*4M*	6.7
2017	1	*ES*	5.4	36	*47M*	17.2	253	*1309B*	*72B*	5.5%	*82M*	*17M*	*61B*	*3M*	14.7
	2	*FR*	5.3	21	*67M*	9.4	306	*2586B*	*99B*	3.8%	*87M*	*29M*	*124B*	*3M*	10.5
	3	*DE*	5.3	27	*83M*	3.8	718	*3657B*	*128B*	3.5%	*37M*	*92M*	*290B*	*5M*	12.0
	4	*JP*	5.3	23	*127M*	2.8	1100	*4860B*	*113B*	2.3%	*29M*	*18M*	*198B*	*4M*	6.7
2019	1	*ES*	5.4	49	*47M*	15.3	282	*1419B*	*78B*	5.5%	*83M*	*19M*	*66B*	*3M*	14.7
	2	*FR*	5.4	33	*67M*	9.1	338	*2778B*	*109B*	3.9%	*89M*	*27M*	*135B*	*3M*	10.7
	3	*DE*	5.4	18	*83M*	3.4	620	*3948B*	*139B*	3.5%	*39M*	*109M*	*315B*	*5M*	12.0
	4	*JP*	5.4	20	*126M*	2.4	1030	*4971B*	*119B*	2.4%	*31M*	*19M*	*202B*	*5M*	6.9

Sources:
1. World Economic Forum - The T & T Competitiveness Report (2015, 2017, 2019)
2. The World Tourism Organization (UNWTO)
3. Knoema - World Data Atlas

Records were cross-checked and corrected.

Columns:
1. Year
2. The Travel & Tourism Competitiveness Index
3. Country [Spain (ES), France (FR), Germany (DE), Japan (JP)]
4. The Travel & Tourism Competitiveness Value (0-7)
5. Sustainability of Travel & Tourism development (Rank)
6. Population
7. Unemployment (%)
8. CO2 emissions (MTCO2)
9. GDP (US$ B, in current prices)
10. Travel & Tourism direct contribution to GDP (US$ Billion, in current prices)
11. Travel & Tourism direct contribution to GDP (% of Total)
12. International number of arrivals (Inbound)
13. International number of departures (Outbound)
14. Domestic Travel & Tourism spending (US$ Billion)
15. Travel & Tourism contribution to employment (Indirect)
16. Travel & Tourism contribution to employment (% of GDP)

Another essential fact is that the travel and tourism industry provides only 6.8% of direct and indirect employment in Japan. However, this ratio is much higher in other competitors, making them more vulnerable than Japan against a sudden loss of jobs during sudden market abruptions, like the one related to the COVID-19 outbreak.

The decrease in tourism revenues is always compensable by heavy industry in Japan, while deficiencies impact of Covid-19 has not occurred to a large extent yet. Therefore, domestic tourism revenues are still a different part of the Japanese economy despite the massive industrial sector's income (WTTC, 2021).

From an economic perspective, tourism is an agile business; its features need to improve and adapt depending on time and technology; herewith, Table 2.2 displays the Japanese sustainable tourism development effort on the country level in the same term.

Table 2. 2 Sustainability of travel and tourism in Japan (2015 - 2019)

1	2	3	4	5	6	7	8	9	10
2015	112	9	4.3	57	16	53	75	35	9
2017	113	10	4.4	27	10	45	29	20	5
2019	77	10	4.4	26	6	56	19	18	5

Sources:
1. World Economic Forum - The T & T Competitiveness Report (2015, 2017, 2019)
2. The World Tourism Organization (UNWTO)
3. Knoema - World Data Atlas

Records were cross-checked and corrected.

Columns:
1. Year
2. Ease of hiring foreign labor (Rank)
3. Information and Communications Technology (ICT) Readiness (Rank)
4. Travel & Tourism gov't expenditure (% gov't budget)
5. Marketing and branding effectiveness in attracting tourists (Rank)
6. International Openness (Rank)
7. Environmental Sustainability (Rank)
8. Tourist Service Infrastructure (included Automated teller machines) (Rank)
9. Natural tourism digital demand (Rank)
10. Cultural and entertainment tourism digital demand (Rank)

According to Table 2.2, there was not much change in information and communications technology rank for the last five years - the government expenditure budget for travel and tourism is also almost the same.

A distinguishable increase in marketing and branding's effectiveness attracts tourists similarly to the visa process's simplification for international openness.

Another significant improvement happened regarding the tourist service infrastructure, including a well-known example of international compatibility with automated teller machines (ATM). As a result, an increasing trend for digital tourism demand can also be observed.

Meanwhile, the distribution of the inbound tourists' arrivals in Japan - over the 2016 - 2019 period was as follows (Statistics Dashboard, 2020):

26% of China, 24% of South Korea, 16% of Taiwan (China), 7% of Hong Kong SAR (China), and 27% of the rest of the world.

In summary, this section discussed the characteristics of the Japanese tourism industry, which comprises most of the domestic tourism spending. Thus, even with a great effort to elevate the international tourism prospect to this extent, it is clear that domestic tourism is the principal part of Japan's tourism industry.

2.3 A methodological overview from the research objectives

Analyzing social dynamics to evaluate tourism-related strategies that benefit local communities and rural ecosystems requires a time-consuming process of gathering data and exploring its many components (Ashley, C. et al., 2002). Furthermore, due to each community's unique characteristics, measuring the impact of tourism on communities with different parameters is a challenging assignment (Simpson, 2009). As a result, proposing a unique framework for each social organization separately to grasp the wide range of travel and tourism incentives would be an optimal solution.

Different researchers have examined the phenomenon of rural tourism from the perspective of the development of local industries (Gabor, 2015). In addition, some researchers have tried to identify the importance of social issues for rural communities, the consequences of occurring events (Reid S. , 2007), and consumer behavior pattern changes in time (Reid D. , 2006). Others have looked into the variables that drive tourists to rural areas and the characteristics of rural visitors, destination management, and business performance (Almeida, A. et al., 2013) (Dimitrovski, D. et al., 2012) (Fotiadis, A. et al., 2014).

On the other side, researchers have concentrated on the importance of the passenger transportation sector to promote communities in various ways. The authors have mentioned passenger transportation within the tourism value chain, an essential component in expanding this industry. Tourism incomes in many rural areas are becoming more robust

as transportation service quality improves in response to people's demands and behavior patterns (Puchongkawarin, C. and K. Ransikarbum., 2020).

From a general perspective, rural areas have different transportation needs than metropolitan areas regarding frequency and accessibility. Operating urban passenger transportation networks entails commercial and public-sector firms dealing with various economic, environmental, and sociopolitical concerns. Multi-criteria decision support systems (DSS) are widely proposed in studies to enhance decision-making considering such issues (Camargo-Perez, J. et al., 2015).

To accomplish so, choosing the industry and local economy-related variables to study the parametric prediction of public transit total demand performs better in an extensive system, such as those found in metropolitan areas, where the required data is large enough to examine statistically (Pamplona, A. and A, Oliveira., 2016).

Passenger surveys were also used to assess the quality of public transportation by analyzing several service attributes such as punctuality, network coverage, line connectivity, service operating frequencies, and overall user happiness (Castillo, D. and F, Benítez., 2012). The findings also prove that each person has a different perspective on public transportation.

For public transportation operators, the significant evaluation factors in users' satisfaction are a priority issue. The most identified relevant transportation system features regarding user satisfaction are trip duration, accessibility, fare, net-work connectivity, information, comfort, safety, and employees' kindness (Fellesson, M. and M. Friman., 2011). Aside from that, environmental concerns and long-term sustainability have also been explored (Castillo, M. and F, Benítez., 2013).

In today's world, many municipalities worldwide are currently trying to transform their traditional transportation modes into smart ones to improve sustainable mobility from environmental concerns. However, overcoming the widespread usage of private cars is still a significant challenge in many urban areas (Politis, I. et al., 2020).

Various urban mobility studies suggested multimodal link solutions, as well as public bus travel, in addition to pedestrian movement. Bike-sharing systems are routinely

investigated as a viable alternative to other modes of transportation (Nikiforiadis, A. et al., 2020).

However, depending on the location's demographic, geographic, and climate aspects, bike commuting will not accommodate all mobility needs, particularly in poorly scattered cities or heavily urbanized areas. Nonetheless, a study found that short-distance travel with bike-sharing stations within 500 meters of the house, job, school, or university has a significantly greater likelihood of using the bike-sharing system (Macioszek, E. et al., 2020).

Another solution that has been proposed in the literature is to use the expanded theory of planned behavior to see whether it can explain why people want to use bus-based park-and-ride facilities. According to the findings, attitude, subjective norm, and perceived behavioral control positively impact park-and-ride facilities (Ibrahim, H. et al., 2020).

Within the numerous academic disciplines, there are also researches on university students' mobility. The significant assumptions of these researches can be characterized as universities' involvement in assisting their communities in developing more sustainable and ecological friendly transportation options and students' social lives (Cruz-Rodríguez, J. et al., 2020).

Planning, infrastructure, regional transportation characteristics, educational quality, and cultural traits on student movement are all being investigated. However, one of the outcomes of the studies is a lack of communication and integration amongst administrative agencies in these residential neighborhoods (Azzali, S. and E, Sabour., 2018).

Additionally, conducting a questionnaire is a common standard methodology used in these students' mobility base research because it is a simple way to collect information and experiences from students (Zambon, 2019).

According to a study concentrating on understanding the relationship between mobility and social relationships, students' mobility gives authorization to them to participate more actively in heterogeneous socio-cultural situations that are notably different from their campus living environments (Golob, T. and M, Makarovič., 2018).

Due to these environmental activities, students are more inclined to reject stereotyping and reflect critically on their social settings. As a result, individuals are encouraged to think about their problems, future activities, and personal identity (Archer, 2003).

On the other hand, machine learning methods based on statistical approaches demonstrate significant performance in forecasting human decisions in many social domains, including sales, economics, law, healthcare, marketing, tourist destinations, and customer management (Kleinberg, J. et al., 2017) (Sohrabi, B. et al., 2020) (Sohrabi, B. et al., 2019).

While machine learning has potential, the findings imply that fulfilling that possibility involves integrating these tools into an economic framework and defining the relationship between forecasts and judgments (Rosenfeld, A. et al., 2012).

In addition, the size of training data has a significant impact on the precise accuracy of the models studied. When the data size is minimal, as it is in many decision-making domains of interest to social scientists, purely data-driven strategies have had impressive and underwhelming outcomes (Plonsky, O. et al., 2019).

Furthermore, ethical debates on the end-user decision dilemma improve advanced autonomous vehicles, giving a framework for solving passenger transportation challenges in rural areas (Bonnefon, J. et al., 2016). However, building autonomous ethical machines that promise benefits to change the world by increasing traffic efficiency and reducing environmental pollution is still seems to be one of the most challenging artificial intelligence problems (Gold, N. et al., 2014).

Following a brief literature review, it was discovered that acquiring and processing data from various sources is critical for identifying problems and providing long-term improvements for rural transportation networks. In addition, identifying the components that interact competently with passenger transportation must be addressed correctly to measure a transportation system's performance, as noted in prior studies.

In summary, this section discussed the importance of transportation on tourism revenues, urban mobility, universities' role on student's mobility, and the evaluation of the

predictive models that effectively identified the parts of factors unique from its targeted community in predicting fluctuations in user demand.

The direct user survey approach to collect qualitative data to apply machine learning algorithms were at the forefront of predictive methodologies. Furthermore, it was revealed that to combine the machine learning approach into an economic framework that focuses on the role of university students in tourism activities, particularly their demand side for possible contributions to the development of socially and environmentally sustainable communities.

In addition, people's travel plans and decisions fluctuate under the influence of time and other factors. Generally, there are two types of change: Due to conditions, the majority are more likely to change their travel plans; however, the minority are less likely to do so (Kozak, M. et al., 2007). As a result, the impact of these changes in the decision was also examined using the approach proposed.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 The author's worldview

In a brief description, this study sought to examine the current and prospective status of on-demand mobility from a university student's daily life needs in a rural city of Japan, which hosts a higher education institute. Therefore, suggested methodological stages in this thesis were:

- To utilize descriptive analyzing methodologies within the big data concept to investigate the research problem as a phenomenon to comprehend it better.
- To implement advance statistical approaches and machine learning techniques to explain and predict.
- To combine human science and administrative study with the outcomes to envisage organizational roles of travel within tourism theory.

However, beyond completing these specific assignments, looking from a broader perspective, the adopted three fundamental pillars of knowledge envision the functional units of this thesis when designing a solution to an identified social problem (Figure 3.1).

Figure 3. 1 The pillars of knowledge adopted thematically in this thesis

Indeed, the suggested methodology's allegation of the thesis produces practical and applicable solutions to any social problem encountered in the real world within the flexibility, reliability, and sustainability order from the experiences of the technical mind.

3.1.1 Flexibility

Flexibility is a feature of personal quality that defines a person's ability to adapt to changing circumstances and think about issues and tasks in innovative and creative ways (Runco, M. et al., 1999).

The personal ability adjusts to situational demands and commitment to the actions; when circumstances change or unexpected occurrences force a person to shift the position, perspective, or responsibility.

A dominant consciousness represented in this thesis was flexibility. Therefore, flexibility is also considered the first protocol of communication among intellectual encounters. The intercultural interactions' humanitarian perception due to the offensive and defensive nature of human beings usually manifests itself in two ways:

- Giving without receiving,
- Receiving without giving.

However, combining these two approaches in one pot will only push for win-win success in international relations.

In cognitive science, flexibility is also described as an ability of mindset moving between two or more concepts and thinking regarding various ideas simultaneously, either inversely or recursively (Miyake, A. et al., 2000).

Therefore, it would not be a mistake to express the definition of multitasking in computer science terminology as the applicability of this ability to the human self.

3.1.2 Reliability

The dictionary description of the word "*reliability*" refers to the ability of a morphological or mechanical system to perform its stated purpose adequately for a specified period under the operational conditions encountered (Gertsbakh, 1989). Thus, a

defined system will be reliable if some undesirable events, called failure, do not occur during the system operation.

Reliability can also be defined as fault tolerance, analysis, and design from a computer science perspective. The frequently used function in life data analysis and engineering is the reliability function. The reliability function can theoretically be defined as the probability of *(n)* success at the time, in the space of *R (n)*. This probability is estimated from detailed analysis of previous datasets through reliability testing and reliability modeling (ReliaSoft, 2001).

Performing a proper quantitative reliability prediction for systems may be difficult and expensive if done without testing. Therefore, an author emphasized the importance of initial or system-level testing until failure and learned from failures to improve the system or section efficiency (Barnard, 2008).

The most critical issue of developing a reliability strategy has a culture for commitment in a social organization. It is essential that everyone involved in a social organization, from upper administration to individuals, understands that a well-resulted reliability innovation is necessary to succeed in the system's desired recovery. However, achieving this culture of reliability may be more complicated than it seems. Some organizations may not have the historical understanding that lends itself to supporting a reliability achievements program. This situation mainly occurs where the organization has had a niche market or little or no previous competition experiences (ReliaSoft, 2004).

On the other hand, convincing an organization's top administrator must be the first step toward developing the requisite success of a reliability culture. The implementation of a reliability innovation will be a frustrating and challenging procedure if this does not happen. Furthermore, practical education on the advantages of a well-designed reliability innovation will go a long way toward supporting reliability methodologies and strategies.

Finally, the need to focus on the financial benefits of a substantial reliability innovation, particularly in terms of reduced costs and more significant participant goodwill, is critical.

3.1.3 Sustainability

Sustainability focuses on addressing current needs without jeopardizing future generations' ability to meet their own (Mitchell, 2020). The three pillars of sustainability were economic, environmental, and social - also known as profits, planet, and people.

A more specific definition of urban sustainability in theory and practice responds to the crises of sustainability in the world by going back to basics from today to the past.

In addition, it makes significant contributions to thinking about and acting upon cities circumferences. First, it provides a means of reflexivity learning about urban sustainability in working practically for positive social development and projected change. Second, it challenges the usually taken-for-granted nature of sustainability practices while providing tools for modifying those practices (James, 2014).

3.1.4 Intelligibility with the research topics

The understanding of social research within the heuristic thought of trial and error is oriented to real-world practice, as mentioned in the reliability section. Thus, the implementation of research methods had into consideration how specific phenomena, such as university-city relationships from the student's requirement of transportation as a part of tourism activities, became part of specific political and social-economic contexts affecting human behavior. Furthermore, the idea of city development emerging from a balance between conflict and group collaboration has also been adaptable to organizational theory regarding management systems (Obolensky, 2007) (Stephenson, 2011).

The composition of the research problem and the corresponding methods aimed to highlight the importance of transportation for the sustainability of societies in city policies, where the university should be a stakeholder.

As tourism and social informatics-oriented multidisciplinary researcher with an empirical knowledge claim, deeply concerned with answering the questions "*why*" and "*how*" of the research problem. In this layout, a desire to investigate the motives that lead universities to adopt discourses and actions conducive to tourism activities and explore

collaborative university-city relations in public transportation development - was generated through a personal reflection on past and present experiences as an exchange student.

Therefore the exploratory questionnaire survey allowed a layout of the university's students' reactions to this phenomenon from a fresh perspective. However, the community sample in this research has not been chosen for their dissimilarity or similarity with the other nations' university students. Instead, it was selected to question whether globalization affects the internationalization efforts of universities with local tourism policies.

Otherwise, a close-circled, considered strictly controllable, and out-of-vision type globalization strategy would not lead to any success without addressing the significance of the rise of universities and their students in the public sphere and local policymaking.

In conclusion, while designing the questions in the survey part of the study, the issues mentioned above were taken into consideration one by one.

3.2 Evolution of computation throughout the increasing size of statistical data

The decisions to commercialize the relational database prototype called System R, developed during the 1970s at the IBM research facility in San Jose, California, were made based on a hardware family business case (Harris, H. and B. Nicol., 2013). By 1977, the Oracle Corporation founders created the first commercially available Relational Database Management Systems (RDBMS) product (Preger, 2012).

Ever since the first appearance of RDBMS, it has offered more or less a successful solution as envisioned before for the transactional manipulation of the data overall electronic computers (Codd, 1970).

Since then, the increasing size of recorded data for statistical analysis at the enterprise level has created competition among the database vendors to provide more sophisticated in-database analytics in their products. A significant challenge in this competition was that each new statistical technique must be implemented from scratch in an RDBMS, which leads to lengthy and complex development processes (Feng, X. et al., 2012).

On the other hand, RDBMS's prominent weakness was storing big files in a document format. However, it would have been handier if data has been converted into info format before the storing action. Unfortunately, other requirements and rivals in the field have been postponed to this task for many years.

Indeed it would have also been helpful for the programmers initially. Still, the batch operations have given a speed advantage to the database programmers when large volumes of information needed to be analyzed. Nevertheless, it has eventually brought the sacrifice from performance expectation when the volume of data is getting bigger and bigger.

By the start of the digital era around 2002, data quality and quantity have increased dramatically (Shepherd, 2004). Data has never been before as valuable as then, and not only data but accessing and processing it on time has taken enormous importance.

A corporate response to the increasing data size due to digitization, Oracle Exadata took place at the computation scene in 2008 for processing data at the utmost speed with specified non-open mainframes (Bach, M. et al., 2015).

Thus, with the debut of Oracle Exadata, a centralized and dominant computing trend has been returned to the computation scene (Figure 3.2) for processing data at the highest speed.

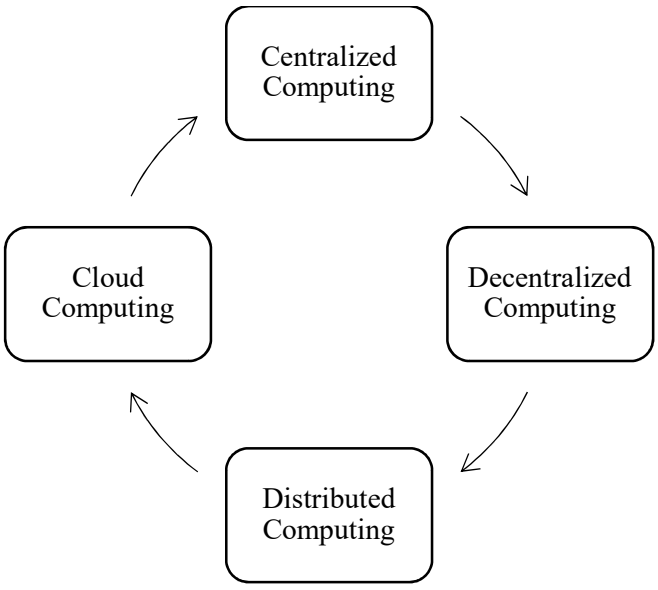


Figure 3. 2 The circle of the computing trends

InfiniBand, parallel processing, and usage of caches have brought severe velocity advantages at the storage servers, for a large amount of data that come from separate supplies by speed (Mietke, F. et al., 2006). However, these specified machines are expensive (Oracle, 2021).

With almost the same attitude as returning centralized non-open mainframes, open-source systems with a creative design and the distributed computing hardware/shareware configurations were developed to use this new increased data concept. As a result, this development has brought a price advantage and created a competitive environment.

Apache Software Foundation has developed an open-source software platform named Apache Hadoop for reliable and scalable distributed computing (Nandimath, J. et al., 2013). On the other hand, the complexity and requirements of hardware/configuration updates were the open-source system's disadvantages. Indeed centralized computing is always more reliable for corporate and government institutions because of cost-performance and security expectations.

Meanwhile, no distributed system is safe from network failures; thus, network partitioning must be tolerated. The CAP theorem indicated that it is impossible for a distributed computer system to simultaneously provide more than two out of three of the following: Consistency, Availability, and Partitions (Fox, A. and E. Brewer., 1999).

Finally, the fast development of social networks and cloud computing has enhanced the large data size concept. Thus, big data applications eventually have found a more proactive usage location inside the computing world. In this manner, the relatively centralized computing and distributed open sources trend has entered a new phase by the dawn of big data.

In technical terms, big data is an idea for massive or complex forms that standard data processing methods are insufficient to process. Data, knowledge, methods, and tools are all called inside the big data concept. However, the data size does not seem enough to determine whether it is considered big data or not.

Additionally, it is necessary to mention that the big data concept would have been meaningless without cloud computing. In terms of the data integration with cloud-based

infrastructure, cloud solutions allow developers that previously required significant investments into hardware to store data to do the same through the cloud at a lower cost.

Big data capabilities can be summarized as follows: Understanding and navigating a relatively significant volume of data sources and capturing data from fast happening events in real-time. Analyze data in its native form, which may be structured, unstructured, or streaming. Manage and store colossal volume and variety of data. Process data in a reasonable time, ingest data at high speed, and bear faults and exceptions.

The types of big data are generally defining as structured data (i.e., MySQL), semi-structured data (i.e., XML), and unstructured data (i.e., NoSQL). On the other hand, the source of big data is defining as databases, log files, sensor networks, social networks, message queues, internet pages, search index data, scientific research records, photograph archives, i.e.,

In the meantime, it is helpful to mention, an essential difference between traditional relational databases (RDBMS) and non-relational (NoSQL) databases is that the NoSQL databases do not support updates and deletes. As a result, many applications do not need to update and delete operations; instead, different versions of the same data are maintained. Updates and deletes are handled using insertion with version control.

Big data analyzes lots of different data types from lots of locations with various types of tools. The main reason why people are investing time to understand big data is to derive value from it. This value characterizes big data. So it is because that data analytics has expanded into the technological fields of machine learning and artificial intelligence.

In addition, the trend to migrate to big data technologies is driven by the need for additional information from an analysis of all electronic data available to a business (Research and Markets, 2021). Thus, in the end, data analysis techniques serve a dual purpose within an organization:

Extensive data analysis is conducted through data streaming as it arrives, and batch data analysis is performed as it grows - to look for behavioral patterns and trends. Thus, as the amount of generated data grows, so will the number of data management techniques.

Furthermore, the more insightful data becomes in terms of speed, scale, and depth, the more it fosters innovation (Manyika, J. et al., 2011).

In conclusion, the technologies that process, manage, and analyze the data are entirely different and expansive, which similarly evolve and develop over time. However, techniques and technologies aside, any form or data size is valuable; alongside the evolving computer-based analysis techniques data harnesses, the analysis still relies on the traditional statistical methods.

3.3 Statistical decision theory

The combination of methodologies in this thesis accepted the mathematical ground of statistical decision theory (STD). STD is concerned with making an optimum decision in obtaining statistical knowledge (data), clarifying some uncertainties in the decision-making process (Berger, 1989). The settlement of the joint distribution function provides the formal definition of STD for a two-class problem:

$$p(X, C_k) \qquad (1)$$

Definition 1:

p denotes the probability,

X denotes an input vector that consists series of values (a set of data),

C denotes a class,

k denotes a constant that takes the value of 1 or 2.

Consider that y is the target variable's correspondent vector (1 or 0) and that $y = 1$ corresponds to class C_1 and $y = 0$ to class C_2. In the process of assigning an input vector X value to an appropriate class (C_1, C_2), the theory is concerned with how to make an optimum decision given tolerable probabilities.

Moreover, STD is a fundamental mathematical theory that finds usage in the discipline of machine learning and social informatics (Trommershäuser, J. et al., 2008). For instance, drawing a sample confusion matrix (Table 3.1) would help visualize the theory in two-class problems. In addition, the performance of machine learning algorithms is also typically evaluated by a confusion matrix.

Table 3. 1 Sample of a Confusion Matrix

	Predicted Positive	***Predicted Negative***
Actual Positive - C_1	*TP*	*FN (Type II Error)*
Actual Negative - C_2	*FP (Type I Error)*	*TN*

In the matrix:

TP denotes correctly classified positive examples (True Positives),

FP denotes incorrectly classified negative examples (False Positives); Type I Error,

FN denotes incorrectly classified positive examples (False Negatives); Type II Error,

TN denotes correctly classified negative examples (True Negatives).

In general, decision theory conceptually concerning these two types of occurrence in a determination problem: Type I and Type II Error. Therefore, SDT also has two key aims in succeeding in its mission: Reducing the probability of making the wrong decision and minimizing the expected loss.

3.3.1 Reduce the probability of making the wrong decision

The first step reduces the probability of making the wrong decision; the input space is divided into decision-making zones (R_k), one zone for each class, (R_k) is assigned to a class (C_k). However, while doing this operation, an inappropriate error occurs when an input vector value belonging to R_1 is appointed to C_2 or a vector value belonging to R_2 is appointed to C_1. The equation is given below use to calculate this error:

$$p(inaccuracy) = \int_{R_1} p(X, C_2).dx + \int_{R_2} p(X, C_1).dx \qquad (2)$$

Definition 2:

X denotes an input vector that consists series of values (a set of data),

C_1 and C_2 denote the two classes,

R_k denotes decision zones,

k denotes a constant that takes the value of 1 or 2.

In summary, a given input vector of X must be assigned to the classes that have provided a smaller integral in the equation for reducing the weight of inaccuracy.

3.3.2 Minimize the expected loss

The second step minimizes the expected loss, considering that each new value of a given input vector of X is assigned to class C_j. In contrast, the correct class is C_k, and an incurred loss occurred, L_{kj}, the k, j coordinated elements of the confusion matrix. The equation is given below use to calculate the average loss function:

$$E\ [L] = \sum_{k} \sum_{j} \int_{R_j} L_{kj}\ p(X, C_k).\, dx \qquad (3)$$

Definition 3:

L_{kj} denotes an occurred loss value,

X denotes an input vector that consists series of values (a set of data),

C_k denotes the actual correct class,

R_j denotes decision zones,

j and k denote the constant of zones that take the value of 1 or 2.

The optimum solution to minimize the average loss function: The imprecision for the correct class is stated through the joint probability distribution $p(X, C_k)$ for a given input vector of X.

3.3.3 Coherence with the research topics

Predictive modeling, especially in the machine learning classification realm, is tried to estimate the future probability of an occurrence in advance from the chronicle set of records related to that occurrence.

Therefore, the researchers desperately require data and different relational data features to conduct predictive analysis. For instance, while conducting a survey study, many additional questions are implemented to collect more information regarding the subject issue. In this way, the weight of the further questions over all the final resolutions can be defined clearly.

When there is not enough historical data to satisfy a researcher, a survey is commonly used to gather opinions on a selected topic. As a result, every survey has a collection of questions for various analysis reasons.

The fundamental objective of the case study in this research is to ask a proper set of questions about the mobility-on-demand choices of university students in a rural town while respecting their sensitivities. As a result, there was no political or misleading substance in the construction of the questions.

3.4 Statistical hypothesis testing

A statistical hypothesis test quantifies how plausible it is to witness two data samples, considering that they have the same distribution (the null hypothesis).

This null hypothesis can be tested by applying appropriate statistical calculations. If the test results indicate insufficient evidence to reject the null hypothesis, any observed variation in model scores is purely coincidental.

On the other hand, if any practical difference in model scores is actual, the test result infers enough evidence to reject the null hypothesis. Hypothesis testing entails the following three steps:

1. To create two hypotheses, one of which must be correct.
2. To gather evidence (data).
3. To create an analysis strategy and develop decision-making criteria.

Testing hypotheses requires the formulation of gathered data and process them with several kinds of statistical methods. When the null hypothesis is accurate, the statistical significance level (α)[8] is a threshold probability of error for rejecting the null hypothesis, also known as Type I error. Conversely, the Type II error rejects the null hypothesis, concluding that there is no effect when available (Luminousmen, 2020).

The statistical significance level (α) is commonly employed at a lower limit since the smaller the chance of making errors. In other words, when differences are judged substantial yet random, it is permissible to commit a statistical error[9].

The p-value is another name for this likelihood.

[8] α is usually expressed as a proportion. Thus, if the confidence level is 95%, α equals 0.05.
[9] From the researcher's perspective.

The statistical criteria for the calculation are affected by whether the samples are independent or dependent. For instance, suppose data or samples were taken from non-metric (nominal or ordinal) scales. In that case, it may be too small to assure that the population's data were drawn have a normal distribution. Therefore, non-parametric elementary methods such as the chi-square test are used to ensure data reliability. Of course, the size of data, either small or large, does not provide enough evidence for the significance of the statistics; however, the overall tendency usually reveals.

In the main experiment, McNemar's test (McNemar, 1947), a particular type of chi-square statistic, was utilized due to the data structure obtained after the survey field study.

3.4.1 McNemar's test

The survey design implemented a two-step verification probe to compare the responders' opinions (positive or negative) in two different stimuli. Initially, these two stimuli were measured with two separate survey questions: The first inquiry (FI) and the second inquiry (SI).

A simple pattern was applied to a (2 × 2) contingency table sample (Table 3.2) to clarify the case. The table tabulates the possible outcomes of two experiments on the same group of N people.

Table 3. 2 Sample of a Contingency Table (2×2)

	Trial 2 Positive	***Trial 2 Negative***	**Row Total**
Trial 1 Positive	*A*	*B*	*A + B*
Trial 1 Negative	*C*	*D*	*C + D*
Column Total	*A + C*	*B + D*	*N*

The cells '*A*' and '*D*' are identical[10]. Conversely, cells '*B*' and '*C*' contain discordant[11] results in the table (Smith, P. and G. Ruxton., 2020).

The critical question was whether the totals in these two discordant cells varied enough to indicate that they elicited separate reactions. The answer could be given by a p-value, which is the probability of the observed difference in these two discordant values in terms of the null hypothesis testing paradigm.

[10] The frequency of positively or negatively answered both stimuli.
[11] The frequency of positively to one stimulus but negatively to the other.

McNemar's test is a statistical test developed to provide this likelihood (Lu, Y. et al., 2016). The test was calculated using both '*B*' and '*C*' values in the formula below:

$$X^2 = \frac{(B-C)^2}{B+C} \approx \frac{(|B-C|-0.5)^2}{B+C} \quad (4)$$

Definition 4:

X^2 denotes a chi-square distribution with 1 degree of freedom,

B denotes the positive reactions to the first stimulus but an adverse to the second.

C denotes the adverse reactions to the first stimulus but a positive to the alternative.

McNemar's test confirms two dichotomous variables' marginal homogeneity. It is a simple case of the chi-square statistic. Thus, it is preferable to analyze paired binomial data to conclude a meaningful change in the data between two stimuli.

In addition, Cramér's V^{12} (φc) value can also be calculated to understand the influential association between two dichotomous variables. Cramér's V correlation ranges from 0 to 1 (Cramér, 1946).

3.5 Probabilistic programming

Probabilistic programming (PP) is a programming paradigm that specifies probabilistic models and performs autonomous inference. As a result, PP makes it possible to identify and fit Bayesian statistical models in various ways.

Although, Bayesian statistics requires making it simple to design and use statistical distributions, samplers, and transformation functions. PyMC, an open-source PP framework written in Python with multiple features, provides a clear, readable, and powerful syntax similar to the natural syntax used by statisticians to construct models. (Salvatier, J. et al., 2016).

In addition, it features next-generation Markov chain Monte Carlo (MCMC) sampling algorithms such as the No-U-Turn Sampler (NUTS) (Hoffman, M. and A. Gelman., 2014).

[12] Other naming is Cramér's phi.

Multiplatform compatibility, an expressive yet clean and accessible syntax, simple connection with other scientific libraries, and extensibility are only a few benefits of PP in Python Software Foundation (Behnel, S. et al., 2011).

In conclusion, PP thus democratizes statistical modeling by considerably lowering the mathematical understanding and time required to build novel models and gain unique insights into any data.

3.6 Machine learning

Predictive modeling illustrates how a data mining method can use an inferential model to anticipate new or future observations. The purpose is to expect the output value *(y)* for recent statements based on their input value sets in particular *(X)*. Usually, in predictive modeling, a group of classifiers is used to train with the dataset to predict whether a randomly selected record performs good or bad performance.

Predictive modeling includes classification, which is an essential aspect of the data science process. When the link between a response variable and an associated set of predictors (inputs) is of interest while the response variable is categorical, a standard supervised statistical learning problem is specified. A dataset to develop an accurate classifier that generates a class prediction for any new observation with an unknown response is a well-known challenge in classification problems (Beaulac, C. and J. Rosenthal., 2019).

A logistic regression (LR), a support vector machine (SVM), a random forest (RF), and a multi-layer perceptron classifier (MLP) were among the classifier methods employed in the main experiment for comparison.

LR is well-considered a standard approach for binary classification in the context of low-dimensional datasets. A low-dimensional data condition usually occurs in scientific fields such as medicine, psychology, and social sciences, focusing on prediction and explainability (Couronne, R. et al., 2018). By plotting the probability scores of the dependent variables, the LR classifier tests the link between a categorical dependent

variable and continuous independent factors. Then, the statistic that best explains an association with yes or no answers is used to create LR models[13] (Connelly, 2020).

SVMs create a decision boundary in the data space to distinguish two classes (Piri, S. et al., 2018). For instance, the SVM classifier constructs a maximum-margin hyperplane in transformed input space and splits the class representative samples when maximizing the interval to the nearest dividing samples (Shmilovici, 2006).

RFs are a machine learning technique aggregating numerous decision trees in an ensemble[14] (Breiman, 2004). An RF classifier's goal is to take a set of high-variance, low-bias decision trees and turn them into a low-variance, low-bias model. Thus, RF eliminates the conflict that can lead to errors in decision trees by aggregating the multiple outputs of individual decision trees. The importance (weight) of each variable can likewise be reliably assessed using RF.

An MLP, unlike prior classification methods, performs classification using an underlying neural network. Artificial neural networks (ANNs) attempt to learn tasks (solve problems) by emulating brain behavior. Specifically, neural networks memorize patterns between features to fit as closely as possible to the desired output. The brain comprises a vast number of specialized cells called neurons that learn brain activity patterns. MLP is known for its high levels of performance. However, just as explaining the action of individual neurons in the brain is challenging, neural network-based models are ill-suited for explanatory modeling, especially when the training data size is small (Cao, Y. et al., 2013) (Kingma, P. and J. Ba., 2015).

After administering machine learning algorithms to the dataset, implemented a combined performance measuring strategy. As a result, the more consistent outcomes with real-life circumstances could be released with the most realistic evaluations.

In addition, the dataset is envisaged to be split into the train (80%) and test set (20%) to avoid overly optimistic estimation accuracy when calculating performance measures.

[13] No response indicates missing data.
[14] This is commonly referred to as "ensemble learning."

The predictive accuracy, F-measure *(F_β)*, the area under the curve (AUC), Cohen's kappa, and cross-validation (CV) were the performance metrics used to evaluate each algorithm in the experiment.

Predictive accuracy is commonly used to evaluate the performance of machine learning algorithms. However, accuracy is not appropriate when the data is unbalanced, and the costs of specific errors vary significantly. Furthermore, real-world datasets frequently contain a large number of "*normal*" cases, with only a tiny percentage of "*abnormal*" or "*relevant*" ones (Chawla, N. et al., 2002). Finally, accuracy is inadequate when considering a user preference bias toward the minority (positive) class examples because of the least represented influence.

F-Measure is a widely used performance metric that provides more information about a classifier's prediction capacity than simple accuracy. For instance, the β in F_β assigns different weightings to Precision and Recall (β = 1, 2, 3 …). Consequently, this β was chosen as equal to 1 to calculate the F_β score in the second experiment.

The receiver operating characteristics (ROC) and the accompanying area under the *ROC* curve (AUC) are two valuable measurements, especially in imbalanced class domains. However, when compared to the majority class, there are fewer actual occurrences. Furthermore, because ROC curves do not provide a single-value performance assessment, AUC is used. Thus, the AUC enables the average evaluation of the best model. Nonetheless, it is not biased against minorities (Branco, P. et al., 2015).

The reliability of data gathering is a critical factor in determining the proposed machine learning model's overall effectiveness in the real world. For instance, Cohen's kappa statistic is commonly used to determine interrater reliability, which indicates how reliable the data is.

Cohen's kappa was created to account for the potential that raters will estimate at least some factors due to uncertainty. A kappa variable is a form of the correlation coefficient that cannot be interpreted immediately; however, a squared value of the correlation coefficient, called the coefficient of determination (COD), can be interpretable directly. The COD is the sum of variation in the dependent variable that can be explained

by the independent variable (McHugh, 2012). The kappa, like other correlation measures, can range from -1 to +1.

Finally, cross-validation is a prominent algorithm selection approach. CV's fundamental idea is to split data once or numerous times to measure the risk of each method. Then, each algorithm is trained using a portion of the data (the training sample), and the remaining data (the test sample) is used to determine the algorithm's efficacy.

The algorithm with the highest effectiveness is then chosen during the CV process. Because of its simplicity and universality, the CV is a widely used approach (Arlot, S. and A, Celisse., 2010).

"Between the idea

And the reality

Between the motion

And the act

Falls the Shadow"[15]

[15] Eliot, T.S. The Hollow Men.

CHAPTER FOUR

EXPERIMENTS

4.1 Preliminary experiment

4.1.1 Introduction

Okhotsk subprefecture (Figure 4.1) was chosen as a micro-level sample community in this preliminary experiment to gain an insight into the current state of rural Japan.

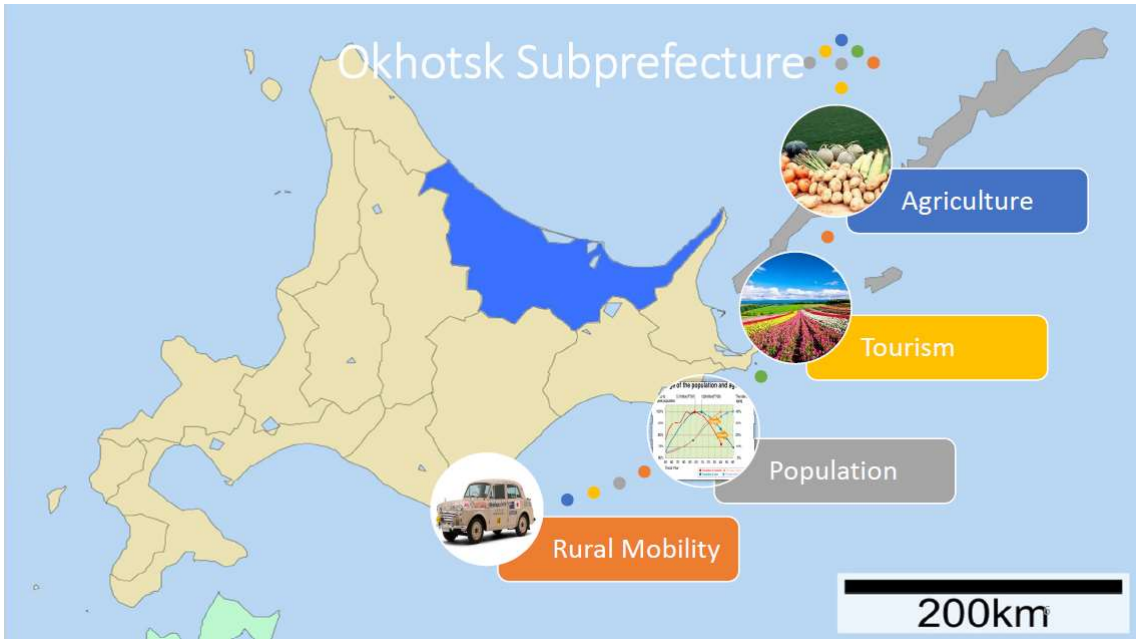


Figure 4. 1 Okhotsk subprefecture with its pros and cons

The Okhotsk subprefecture located in Hokkaido, the second-largest island of Japan, consists of 3 cities: Abashiri, Kitami, and Monbetsu.

The name of the subprefecture is derived from the Sea of Okhotsk. The climate is much colder than in other parts of Hokkaido during the winter; the area is also famous for drift ice sightseeing.

In addition, agriculture is an important economic industry. However, compared with other regions in Japan, the population density in the Okhotsk subprefecture is very low.

In general, Hokkaido's population is aging rapidly and shrinking faster than the rest of Japan due to constant migration to Japan's metropolitan areas, particularly among the younger generation.

Besides being the most prominent agricultural area in Japan, Hokkaido is also a popular destination for domestic and international tourists. Although there was a boom in international tourism arrival during the last five years, the tourism industry in Hokkaido is also dependent on domestic tourism activities.

Meanwhile, Japanese participation in tourism activities has steadily declined from 1998 until 2015. Understanding this happening requires in-depth data and analysis based on scientific, demographic, economic, political, historical, and sociological elements.

Moreover, public transportation consumption has also steadily decreased by roughly 36% in Japan's countryside over the last 20 years.

There must be many reasons that trigger regressions on this scale in such an economically developed society. To begin with, the summarization of the economical parameters of Japan would be appropriate:

Even though Japan has experienced almost 30 years of slow economic growth[16], it is still the world's fourth-largest economy, with a nominal GDP of 590 trillion JPY[17] in 2020 (World Bank, 2021). Japan is also the first very highly developed country in Asia.

In addition to these pieces of information, the tourism industry generated a total impact of 38 trillion JPY[18] of Japan's GDP in 2016, and it makes Japan the 3rd country globally regarding the actual contribution of tourism to the GDP after the United States and China (UNWTO, 2019).

Based on its direct, indirect, and induced GDP impact, tourism generated 7.4% of Japan's GDP in 2016. However, perhaps of Japan's island nature, 86% of the tourism industry's total contribution to the GDP consists of domestic expenditure (WTTC, 2017).

[16] The notorious 'Japanese asset price bubble,' an economic bubble in Japan which burst by the end of 1991, can be given as the reason for the country's economic decline.

[17] 5.328 Trillion USD.

[18] 343 Billion USD.

On the other hand, although Japanese tourism expenditure has reduced, its economic value was still very high in 2016. However, depending on the records, it has lost a great deal since 1998 (Statistics Bureau, 2016). As a cause and effect relationship with the decline of domestic tourism expenditure, the participation rate in domestic tourism activities in Japan also declined between 1986 and 2006 (Figure 4.2).

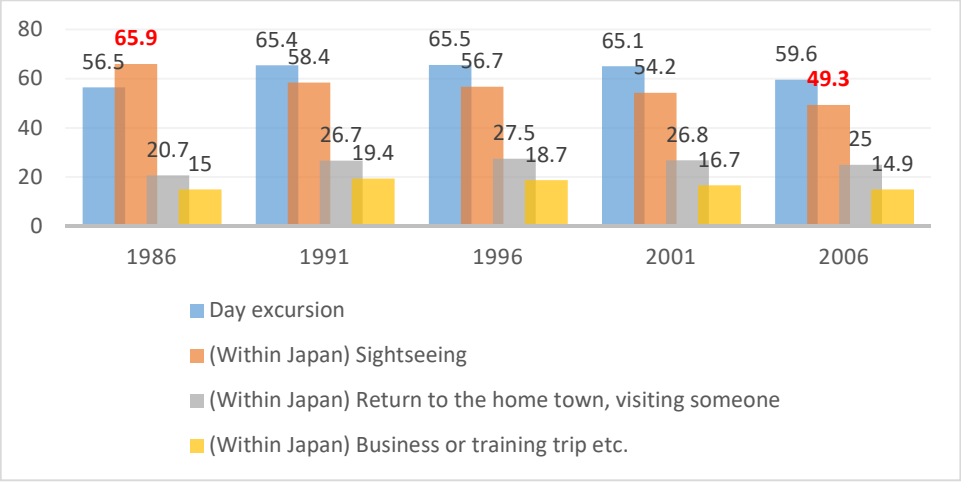


Figure 4. 2 Japanese behavior change towards excursions (1986 - 2006)

The obtained data from the official statistics office of Japan shows that participation in sightseeing activities has dropped from 65.9% to 49.3% of the overall population.

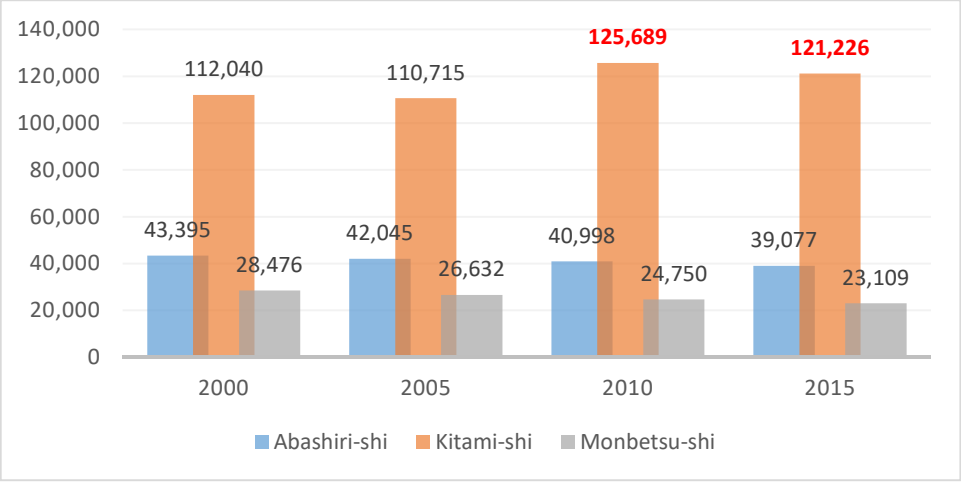


Figure 4. 3 Population census in Okhotsk (2000 - 2015)

A sample of a statistical data anomaly can be seen in Figure 4.3: A sudden increase in the population census of Kitami in 2010 while the general trend is in decline.

The definition for this anomaly was, three independent districts of the region, namely Tanno, Rubeshibe, and Tokoro, were merged into the city of Kitami in 2006. Due to excessive population loss. However, population decline continued again in the following census even after this population injection to the community.

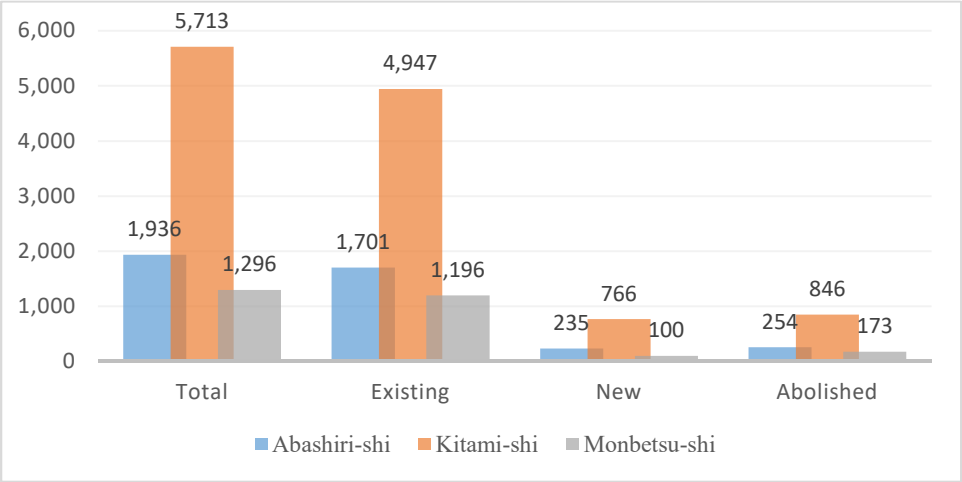


Figure 4. 4 Establishment's existence in Okhotsk (2014)

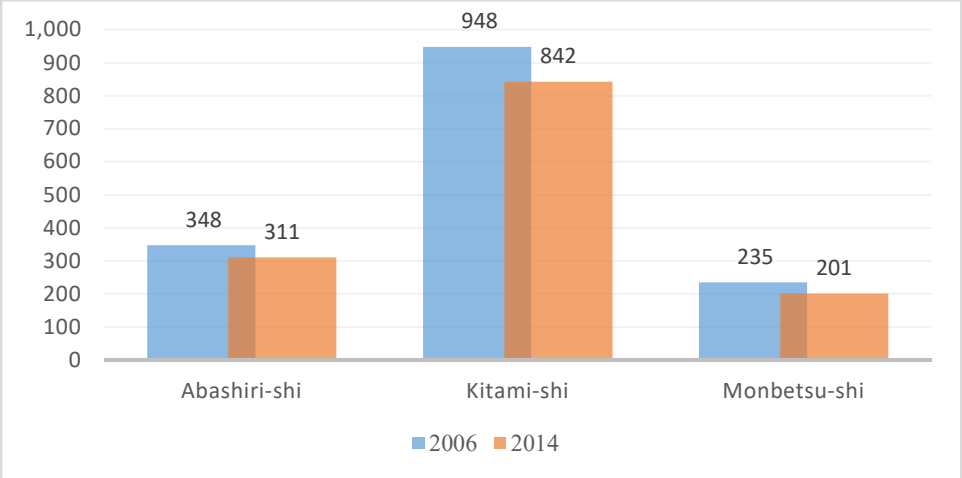


Figure 4. 5 The hospitality industry in Okhotsk (2006 - 2014)

As seen in Figures, like a rapid negative population growth rate in Okhotsk, all business establishments, including tourism-related businesses, were also slowing down.

In summary, with this analysis, the character of Okhotsk statistical data, shown respectively, imitates the similarities with Japan's general characteristics.

Thus, although each geography region's demography structures are different in Japan, this analysis results contributed a knowledge-based sample for other researchers who focused on the regional studies. Therefore the objective of the experiment was to collaborate with the local authorities to construct better living habits for the rural communities and improve their quality of life based on scientific proof earned through the completion of the experiment.

4.1.2 Methodology

The practical steps of this experiment have been revealed after conducting preliminary observations on tourism and demographic indicators in documented statistics and prebuilt survey data in the Okhotsk subprefecture of Japan.

Although tourism informatics methodologies combine several kinds of technology and knowledge, such as big data analytics, probabilistic programming, and machine learning, the observation was essential to building a statistical model from the characteristics of the local community-focused. A reliable mathematical construction of a social model always depends on factors and parameters discovered based on observations.

The summary of the experiment's stages: The data from historical and current sources were processed using a statistical model supported by Bayesian statistics and combined with a Monte Carlo simulation (Scott, L. et al., 2016). In addition, a distributed computing system with Apache Spark was utilized for the performance issue during the technical process. Finally, the findings were illustrated with a set of probability distribution graphs and evaluated in terms of likelihood.

The Bayesian statistics were used in this experiment is a mathematical procedure that applies probabilities to statistical problems. Bayesian inference provides users with the tools to update their beliefs in the evidence of new data. If any incident depends on various conditions, Bayesian interference combined with the Monte Carlo method would be appropriate as an observation technique. In this way, any social problem with a probabilistic interpretation can be solved using Monte Carlo methods.

A mighty class of Monte Carlo techniques is the so-called Markov chain Monte Carlo (MCMC) algorithms. Moreover, Monte Carlo methods are often necessary for the implementation of optimal Bayesian estimators. The three main applications of Monte Carlo methods can be list as "*optimization,*" "*numerical integration,*" and "*generating draws from a probability distribution*" (Kroese, P. et al., 2014).

The primary mechanism of any simulation is feeding the analysis model with repeated random sampling inputs to get numerical output and assuming that a repeating sampling input, depending on modeling complexity, may require high processing power and time; this can be achievable with Apache Spark components[19].

Apache Spark is an in-memory framework that is an alternative to MapReduce[20]. Spark has a higher data processing engine for large datasets and in-memory computing. In addition, the Spark Core is designed to scale up from one to thousands of nodes.

PyMC[21] was used as a Python package for probabilistic programming, which focuses on advanced MCMC and various fitting algorithms. Because of its versatility and extensibility, PyMC utilizes to solve a wide range of challenges.

Due to the difficulties in providing all this technical infrastructure from a single source, this experiment was carried out to process faster using the Big Data University Virtual Lab Environment, thanks to IBM cloud computing (IBM Skills Network Labs, 2018).

In addition, the methods and probabilistic programming examples of Cameron Davidson-Pilon have influenced the conceptual progression of this experiment (Davidson-Pilon, 2015).

4.1.3 Case study and results

Before proceeding to the case study assignment, it would be helpful to mention that an essential element of statistical knowledge and, eventually, probabilistic programming is random variables which are divided into two sub-categories:

[19] Apache. Apache Spark URL: https://spark.apache.org/
[20] Apache. Apache Hadoop URL: https://hadoop.apache.org/
[21] PyMC3. Probabilistic Programming in Python URL: http://docs.pymc.io/

- Observed random variables: Have sample data values defined by prior distributions that can come in handy in likelihood distributions.
- Unobserved random variables: Have no sample data values defined by prior distributions.

While discussing the random variables, more specifically probability programming and distribution, at least two variable types need to be mentioned: Discrete random variables and continuous random variables.

Discrete random variables find usage alongside the probability mass function (PMF) when the output of the function data value is finite numbers. Examples include discrete uniform distribution, Bernoulli distribution, binomial distribution, Poisson distribution, geometric distribution.

Continuous random variables, a probability density function (PDF), define continuous random variables, and the output of the data value can be defined as infinite numbers. Examples include normal distribution, uniform distribution, Beta distribution, and Gamma distribution.

- **Problem statement:**

Japan has a very distinguished domestic tourism spending. Its total contribution to the Japanese GDP is remarkable (7.5 % by 2015). However, the fact is, it has been declining gradually since 1998[22].

"*What would be the future of the Japanese domestic tourism industry?*"

- **Data:**

Japanese domestic tourism expenditure data were obtained between 1995 and 2015[23] (Statistics Bureau, 2016). The bivariate numerical data was converted to a line graph in the following Figure 4.6. It is noted that the shape of the graph has revealed a continuous decline.

[22] From 33.5 Trillion JPY in 1998 to 21.7 Trillion JPY in 2016.
[23] Local currency in Trillion JPY.

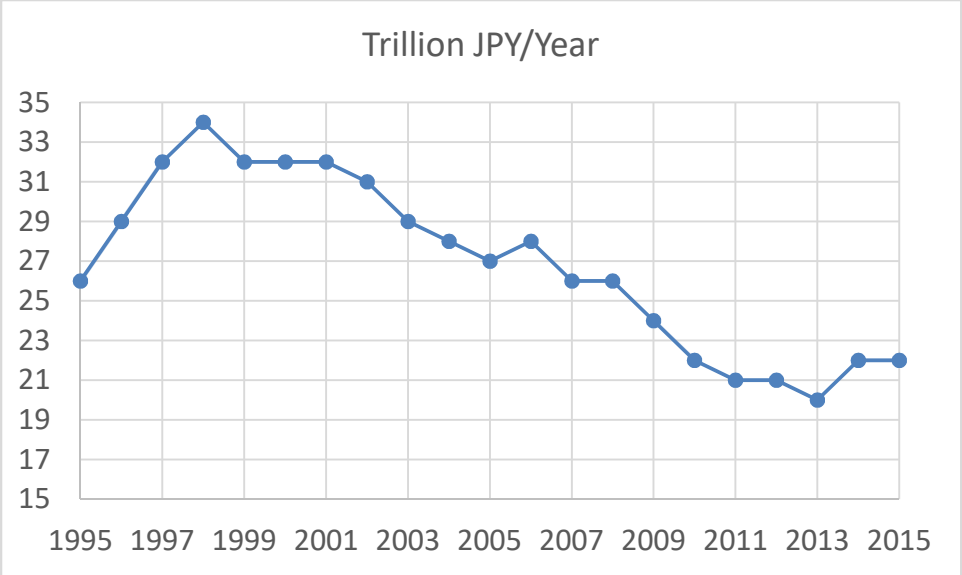


Figure 4. 6 Domestic tourism expenditure/year regression graph

The initial conditions of the dataset were for intuitive perspective. Since obtaining a dataset sample, it is possible to define random variables in a discrete random form for this experiment.

- **Model:**

A significant advantage of the Bayesian interpretation is that it can model uncertainty about events without long-term frequencies (Murphy, 2012).

While, a dataset with discrete values has been obtained in this experiment, which gives an observed random variable opportunity for the model processes. This condition makes it possible to use a Poisson distribution for PMF to modeling, which is the best fit for the features of obtained data[24]. The equation of Poisson distribution is given below:

$$Poi(x|\lambda) = e^{-\lambda}\,\frac{\lambda^x}{x!}, x = 0, 1, 2 \ldots \qquad (5)$$

Definition 5:

x denotes events that occur at times,

λ denotes event occurrence rate,

e denotes the constant number equal to 2.71828 (Euler's Number).

[24] Prior beliefs.

From the equation, the probability is dependent on λ. The value of λ determines the shape of the probability distribution. The calculated plot will have two different values of λ using the Poisson distribution.

Two very different λ's would reveal two different distributions to determine a turning point in the event. In other words, if there was any unusual behavior on the observed random variables regression line graph, the mean of which is our distribution governed by two different values of λ, one before the turning point and one after the turning point. Thus, τ (tau) in the definition represents the turning point that distribution was looking for.

The Python programing language and its probability programming specified package PyMC was used to simplify and solve the problem. PyMC allows building a probabilistic model in a short time, as seen in Figure 4.7.

```
import pymc as pm

def tourism_model(data):

    alpha = 1.0 / data.mean()
    lambda_1 = pm.Exponential("lambda_1", alpha)
    lambda_2 = pm.Exponential("lambda_2", alpha)
    tau = pm.DiscreteUniform("tau", lower=0,      upper=len(data))
    observation = pm.Poisson("obs", lambda_, value=data.values,
             observed=True)
    model = pm.Model([observation, lambda_1, lambda_2, tau])
```

Figure 4. 7 Sample Python code block for building a probabilistic model

- **Inference:**

After creating a model definition, the MCMC sampling algorithm could start to use in the following order:

1. The observed dataset was introduced to the model.
2. The MCMC fitting algorithm was assigned to sample the model[25].
3. The sampling process produces the required set of λ_1, λ_2, and τ (tau) data distributions.
4. Finally, a block of codes inside the model created graphs to interpret this likelihood.

[25] The model was sampled 50000 times with random values.

Again, the critical purpose of this research was to provide a more user-friendly output for individuals in the community who do not have a technical background.

- **Result:**

In Figure 4.8, the first graph (λ_1) shows the distribution before τ, while the second graph (λ_2) shows after τ.

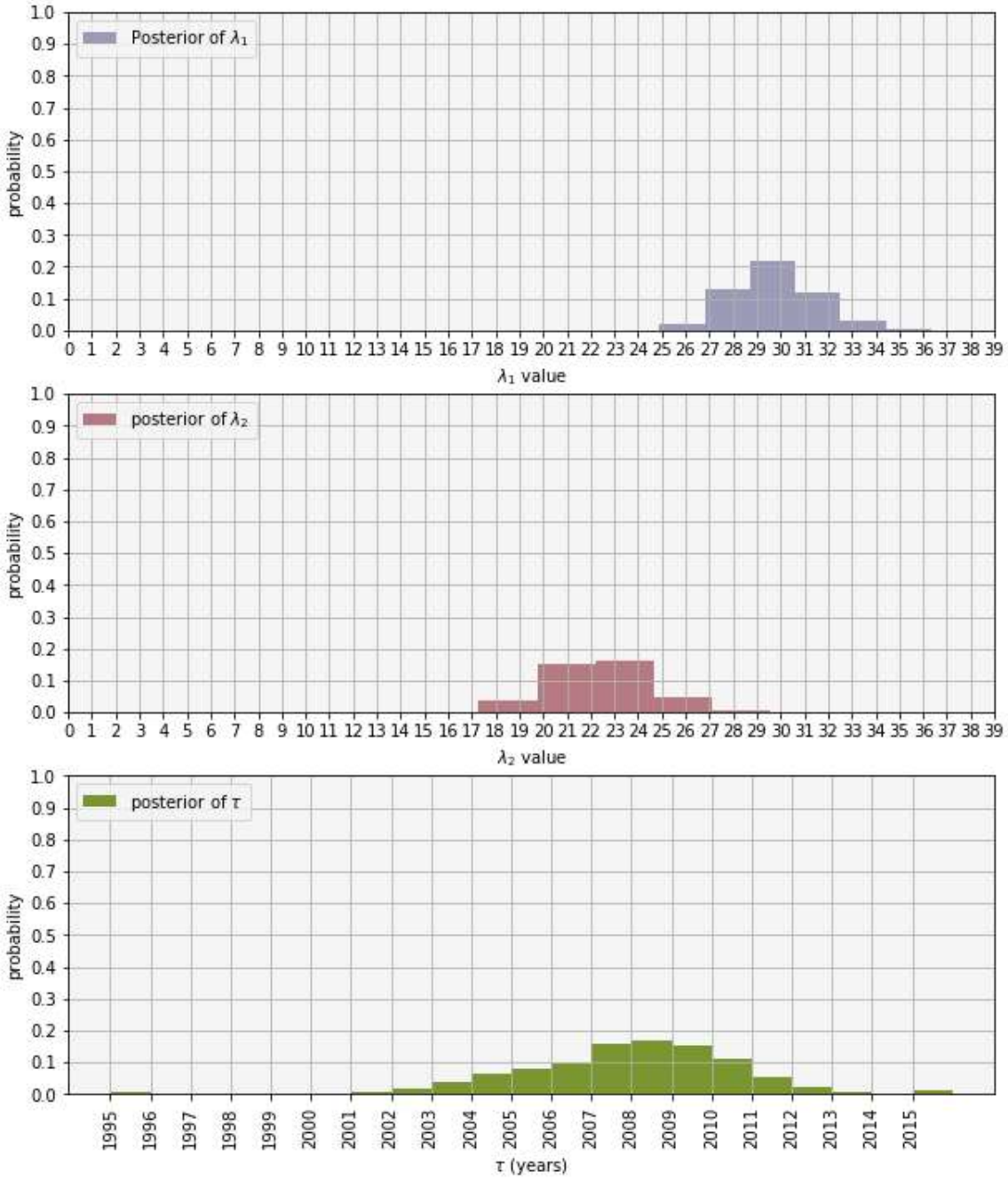


Figure 4. 8 Posterior distributions of λ_1, λ_2, and τ

It should be noted that the τ graph has not had any turning point because preliminary data shows a continuous decline in domestic travel and tourism expenditure. Therefore, since there was no change, λ_1 and λ_2 revealed similar possibilities in different probability points.

The posterior distribution of domestic travel and tourism spending changed between two groups of values, first a realistic 22 up to 23 trillion JPY, second an optimistic 28 up to 30 trillion JPY for the next year, 2016.

Indeed, domestic travel and tourism expenditures were 21.7 trillion JPY in 2016 (WTTC, 2017). Another critical point is that this trend has changed positively from 2015, which was a perfect year for tourism in Japan.

4.1.4 Conclusion

This experiment has triggered an awareness to increase the efficiency of the community's everyday life living in the Okhotsk area of Japan.

Furthermore, based on Bayesian statistics, the results could help understand better both prefectural[26] and country-level[27] human experiences.

The value distribution of people's surroundings is changing at every moment, and the impact it creates requires it to be measured as quickly as possible for clear comprehension in modern society. Never before has data been as valuable as today, but processing the data in time is crucial everywhere in the world.

Big Data enforced various processing solutions, including data cleaning and analysis, to derive value from it. Nevertheless, most people need big data to be presented user-friendly to understand it clearly and make decisions.

Nevertheless, the main interest was to support the Okhotsk subprefecture of Japan and the people who live there through the data collected and processed in this experiment.

[26] Hokkaido.
[27] Japan.

4.2 Main experiment

4.2.1 Introduction

Following the purpose of this thesis from the beginning, this experiment gathered the information through a 2019 survey titled "*Student daily life after school hours*" at Kitami Institute of Technology, which has about 1,800 undergraduate students (Statresearch, 2020).

The survey information bulletin was distributed throughout campus, and the data collection period lasted a month. Responders read the QR code on the notification form with the QR code reading capability of a free app for instant communication on their mobile phones to visit the web link of an online form.

In this way, the data of students' demographics, transportation preferences, and overall happiness with their academic lives were gathered through the survey questions. A total of 250 students replied to the poll, accounting for 14% of the student body.

Low response rates and nonresponse bias are usually caused for concern in the survey's studies. The tendency on a sample statistic to systematically over or under-estimate a population parameter is referred to as bias in survey sampling. In theory, comparing estimates to actual population values is the best technique to detect bias in a sample of responders' estimations; but, population numbers are not always available (Bose, 2001).

Furthermore, as an essential indicator, a survey's response rate represents the quality and dependability of the data collected. Because there is no agreed-upon minimum acceptable response rate, the creation, distribution, and management of surveys determines the quality and dependability (Keller, 2014).

A study done in Japan discovered that response rate bias was in univariate distributions of demographic data, activities, and attitudes. Despite this, data from a multivariate study assessing connections between variables while adjusting for numerous background variables do not show bias from low response rates for most dependent variables (Rindfuss, R. et al., 2015).

Furthermore, the survey environment's difficulties in measurements are defined, how questions are asked, and the responders' condition. When looking at data from another survey study concentrating on the health-promoting lifestyle profile of Japanese university students, for instance, it can be seen that the response rate drops as the student year progresses (Wei, C. et al., 2011).

However, this experiment can only achieve absolute reliability by administering the identical questionnaire to fresh university students each year and comparing the findings.

In the meantime, the distribution of sex and origin of responders generated by this survey data fits the characteristics of the current student population.

Despite the relatively low response rate from the student survey, research was nevertheless able to show the differences and similarities in the student population of a university town within a reasonable limit.

4.2.2 Methodology

Statistical formulas are available for determining the sample size (James, B. et al., 2001). The two critical indicators for these formulas are the margin of error (MOE)[28] and the accurate confidence level[29] of the results (Taherdoost, 2016).

Furthermore, using a two-step verification probe in the survey design is under the conditions mentioned before. The long-lasting neglected status of PBS in rural Japan made reinforced the belief that there will be no change in the situation; psychologically, people have accepted this as the "*status quo*."

Therefore, the first question (first inquiry - FI) was whether a student would use PBS more in the current situation with a binary response (*yes/no*).

The survey then asked a series of questions about everyday activities of life that need mobility, such as joining a part-time work, supper options, demand for nearby restaurants, and supermarket shopping.

[28] In this research, a 5% MOE is accepted.
[29] The confidence level is used as 95%.

Then the responders' public transportation habits were asked about how often they use PBS, what days and times they prefer to go out, and what mode of transportation they choose during these activities.

These questions were designed to highlight to responders the importance of PBS in everyday life. This approach also can be paired with the training step of the modeling in the machine learning research domain because, without a doubt, humans can also learn.

The survey's second question (second inquiry - SI) inquired if a student would like to use PBS in a different setting with a binary response (*yes/no*). An alternative PBS can meet the mobility needs of any person in regular life for various reasons, such as going shopping, dining in a restaurant, sightseeing, or even commuting to part-time employment locations.

In the first inquiry (FI), the difference in responders' decision distribution between the two groups (*yes/no*) revealed no evident variety; nevertheless, the second inquiry (SI) showed a significant difference. This result demonstrated that employing the two-stage verification probe to investigate this community was the correct decision.

Furthermore, the variable associated with the second inquiry (SI) was used as the target value for machine learning-based predictions across the entire survey sample. Finally, the responders' overall satisfaction with their academic lives has been examined concerning the numerous items asked.

By the end of the data collection, the responders had filled in 18 distinct categorical variables. These 18 category variables were converted to continuous variables, yielding 47 different types of constant labels.

When the dataset was created and readied for analysis, chi-square statistical tests were used to select the optimum subset collection, as the dataset attributes were categorical. The chi-square is a nonparametric statistical test that examines the relationship between two variables (Ugoni, A. and B. Walker., 1995). However, it is not applicable to analyze parametric or continuous data types (Rana, R. and R. Singhal., 2015). In addition, while the dataset attributes were of the continuous value-form, used the p-value was. Therefore,

the possibility that the observed difference between the two groups is attributable to chance is statistically significant.

Any practical difference is assumed to be explained by sampling variability if the p-value is more significant than the statistical significance level *(α)*. However, simply stating the significant p-value for analysis is insufficient to comprehend effect sizes adequately (Sullivan, M. and R, Feinn., 2012) (Ioannidis, 2018).

In summary, two different criteria were used in this experiment to evaluate the participants' behavior and better simplify the models' predictive weight from the presented data due to variances in the perception of statistical inference (Liu, S. et al., 2020). These criteria were chi-square and p-value.

4.2.3 Case study and results

This section delves into the information offered in the following tables, which are divided into three parts. The first part is about the students' demographics, the second part is about their travel habits, and the third part is about their overall satisfaction with the education experience.

Additionally, Table A. 1, in the Appendix section, provides explanations for the abbreviations used in the tables.

The size of the collected responses from a survey and the total student population calculates the MOE as ± 5.76% with a 95% confidence level. A 95% confidence level means that 95 out of 100 samples will have the actual population value within the specified MOE of ± 5.76%.

The Chi^2 column can be interpreted as follows: Categorical characteristics with the higher chi-squared statistics have more relevance and importance in predicting students' PBS demand.

The p-value, on the other hand, is evidence against the null hypothesis after transforming data to a continuous type. The evidence is statistically significant if the p-value is tiny.

- **Demographic profile of students:**

The responders were requested to enter their demographic information in this section. Table 4.1 summarizes the findings:

Table 4. 1 Descriptive distinctions of students' demographic background

No	Attributes	No	Label Name	Relative (%)	Chi^2	*p*-Value
1	**Sex**				0.18	
	Female	1	SexF	14.8		0.27
	Male	2	SexM	85.2		0.27
2	**Origin**				0.11	
	Japan, other than Hokkaido	3	OJ	58.8		0.98
	Hokkaido	4	OH	33.2		0.42
	International	5	OI	8		0.15
3	**Accommodation**				0.65	
	Apartment	6	Lapart	67.2		0.39
	Boarding house	7	Lblodge	17.2		0.35
	University dormitory	8	Ldorm	8.8		0.78
	Parent's home	9	Lhome	6.8		0.60
4	**Preferences of Dinner**				0.58	
	Cooking	10	Dcook	54.4		0.63
	University cafeteria	11	Dcafe	20		0.13
	Meal served by a boarding house	12	Dbmeal	18.4		0.38
	Lunch box	13	Dbento	5.2		0.26
	Restaurant	14	Drstrnt	2		0.13
5	**Participation in a part-time job**	15	PTime		2.99	0.01
	Yes			52		
	No			48		

The frequency of gender distribution was 85.2% males and 14.8% females among the 250 students who participated in the survey. In addition, 33.2% of students were from Hokkaido, 58.8% from other parts of Japan, and 8% were foreigners.

Furthermore, 67.2% of students lived in an apartment house[30], 17.2% in boarding homes with meals, 8.8% in a university dormitory, and 6.8% with their family.

The tendency for dinner, while 54.4% of the students stated that they prepare meals for themselves, 20% had their dinner at the campus cafeteria, 18.4% at the boarding house[31], only 2% used neighboring restaurants, while 5.2 % bought their meals/lunch boxes at convenience stores.

[30] Block of flats.
[31] Some boarding house near the campus offers meals even for not-tenant.

The main finding in this section is that 52% of responders who work part-time have the higher chi-squared statistics and the lowest p-value. In other words, university students in Kitami city add enormous economic value as students and as employees in the local business market.

- **Travel behavior of students:**

The responders were asked to describe their travel habits in this section. Table 4.2 summarizes the findings:

Table 4. 2 Descriptive distinctions of students' travel behavior

No	Attributes	No	Label Name	Relative (%)	Chi^2	*p*-Value
6	**Frequency of public buses usage**				4.30	
	Daily	16	PTFD	8		0.15
	Weekly	17	PTFW	30		0.00
	Monthly	18	PTFM	30.4		0.80
	Less than a Month	19	PTFL	31.6		2.41
7	**Possession of an automobile**	20	Car		0.3	0.55
	Yes			14.4		
	No			85.6		
8	**Demand to use buses more often**	21	**FI**		17.95	0.00
	Yes			59.6		
	No			40.4		
9	**Preferred evenings to go out**		(multi-selection)			
	Friday	22	Frdy	32.4	4.47	0.01
	Saturday	23	Strdy	84	0.12	0.38
	Sunday	24	Sndy	69.6	0.94	0.08
10	**Preferred periods to go out**		**(**multi-selection**)**			
	18:00 - 20:00	25	HL	76.4	0.00	0.94
	20:00 - 22:00	26	HM	36.8	4.13	0.01
	22:00 - 24:00	27	HH	26	7.41	0.00
11	**Preferred transport in that period**				1.09	
	Walk	28	PTTwalk	45.2		0.26
	Bicycle	29	PTTbike	21.2		0.62
	Bus	30	PTTbus	15.6		0.20
	Personal automobile	31	PTTcar	12.8		0.13
	Taxi	32	PTTtaxi	5.2		0.08
12	**Demand for buses in that period**	33	**SI**	**(target)**		
	Yes			70.8		
	No			29.2		

The frequency of using public transportation to destinations other than the campus resulted in 62% monthly or less than a month. The percentage of students who possess a car was 14.4%.

FI - *An unconditional first confirmation of the demand for PBS resulted in a 59.6% approval among the responders, with a 6% MOE.*

Friday, Saturday, and Sunday were chosen as the most popular days to go out by 32.4%, 84%, and 69.6% of the students, respectively.

The most popular period was between 18:00 and 20:00, drawing 76.4% of the students' attention. The hours from 20:00 to 22:00 were chosen by 36.8% of the students, and from 22:00 to 24:00 were selected by 26% of the students.

During these times, the students' favorite modes of transportation were as follows: 45.2% selected to walk, 21.2% liked bicycle, 15.6% choose a public bus, and 12.8% use their car, and 5.2% afford a taxi.

SI - *A conditional second confirmation of the demand for PBS resulted in a 70.8% approval among the responders, with a 5.6% MOE.*

The high percentage of people who prefer to walk is the main finding in this section. Because the bus route was not available after 20:00 evening hours, people chose to walk. While passengers still have to walk halfway to the target destination in most bus lines, many prefer to walk instead of paying for bus fares.

With only 14% of students owning a car, it is clear that a personal automobile-dependent public transportation paradigm does not cover most students in this city. The highest chi-square statistic and p-value correlation were seen among weekly bus users who want to use buses more frequently or those who want to go out on Friday night after 20:00 and need a cab in this circumstance.

- **Overall Satisfaction of Students:**

In this section, the responders were asked to evaluate their general satisfaction with their academic lives and future expectations. Table 4.3 summarizes the findings:

In terms of the university's contribution to the city from the university students' perspective, 34.8% of the participants valued economic support, 26% stated the educational impact, 19.6% selected the academic, 16% industrial, and only 3.6% cultural contributions.

The percentage of students who rated the quality of their university as low, medium, or high was 47%, 34%, and 19%, respectively.

The inconvenient impact of winter seasons on students' daily lives was rated as high by 64%, medium by 22%, and low by only 14%.

Table 4. 3 Descriptive distinctions of students' overall satisfaction

No	Attributes	No	Label Name	Relative (%)	Chi^2	*p*-Value
13	**University's contribution to the city**				1.32	
	Economically	34	Conteco	34.8		0.10
	Educationally	35	Contedu	26		0.76
	Academically	36	Contaca	19.6		0.65
	Industrially	37	Contind	16		0.03
	Culturally	38	Contcul	3.6		0.08
14	**Evaluation of the university**				0.53	
	High	39	ValueH	19		0.54
	Medium	40	ValueM	34		0.26
	Low	41	ValueL	47		0.55
15	**Effect of winter on the daily life**				5.63	
	High	42	WinterH	64		0.01
	Medium	43	WinterM	22		0.15
	Low	44	WinterL	14		0.08
16	**Demand a nearby fast-food restaurant**	45	Mc		1.39	0.00
	Yes			83.2		
	No			16.8		
17	**Demand a nearby supermarket**	46	SpM		0.77	0.00
	Yes			90.4		
	No			9.6		
18	**Dwelling in the same city after graduation**	47	Grad		0.04	0.84
	Yes			20		
	No			80		

"*Having a chain-restaurant near the campus area would make life more comfortable*," received 83.2% support.

"*Having a supermarket near the campus area would make life more comfortable,*" received 90.4% support.

"*Would you stay in Kitami city after graduation if there were any career opportunities for you?*" received 20% support.

The students who responded positively were divided into 40% from Hokkaido, 46% from elsewhere in Japan, and 14% from abroad.

The following summarizes the main findings in this section: The institution's contribution to the city is not a point of agreement among the participants. However, the university receives an average rating from participants. In addition, participants who thought the winter season had a significant negative impact on their daily lives scored strongly in the chi-square statistic while also having the lowest p-value.

Similarly, most participants agreed that there was a need for a chain restaurant and a grocery close to the campus area. Both variants had a low p-value, which was statistically verified.

Another significant finding is that nearly four out of five participants do not intend to stay in Kitami after graduation. The remaining 20%, on the other hand, if extrapolated to the entire student population, equals around 200 people per year, representing a dynamic and robust workforce that, if given a chance, may resuscitate Kitami's dwindling population.

- **McNemar's Test:**

A hypothesis test determines if two mutually exclusive claims may be reconciled. This experiment's statistical significance threshold *(α)* was chosen before data collection and set to 5% (Leo, D. and F, Sardanelli., 2020). The following is the definition of the hypothesis for this experiment:

- H_0: $\mu_1 = \mu_2$ — states the null hypothesis that there is no significant difference concerning PBS requests between the weekdays (μ_1[32]) and weekend days (μ_2[33]) during the evening hours by students of the Kitami Institute of Technology;
- H_A: $\mu_2 > \mu_1$ — states the alternative hypothesis that there is a significant distinction.

The contingency table created from the dataset is represented in Table 4.4.

Table 4. 4 Contingency Table (2 × 2)

FI / SI	*Second Inquiry - Yes*	*Second Inquiry - No*	*Total*
First Inquiry - Yes	129	20 (B)	149
First Inquiry - No	48 (C)	53	101
Total	177	73	250

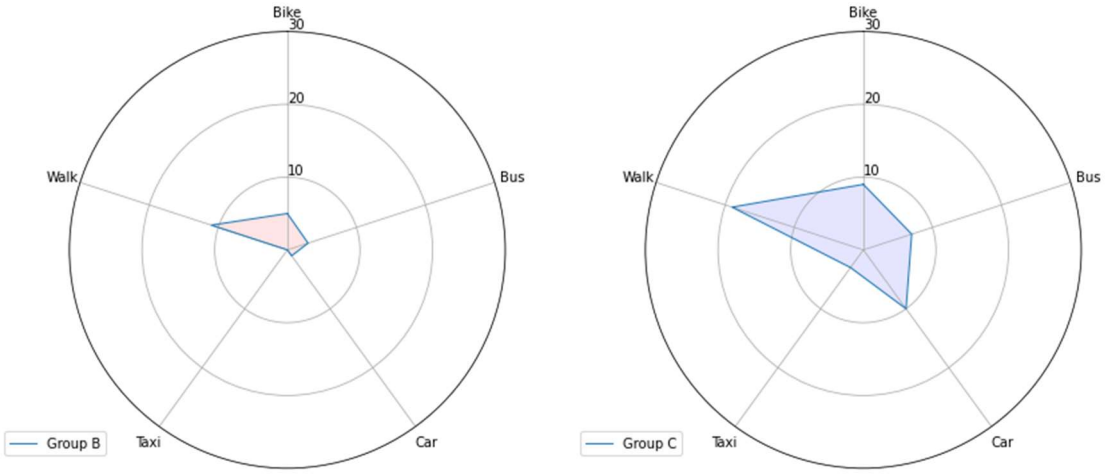
[32] Mean of FI.
[33] Mean of SI.

The number of responders willing to use public bus services at the initial survey but who changed their minds in unfavorable bus services at the second survey was represented by table-cell (B). On the other hand, the number of responders whose case is the reverse was represented by table-cell (C). Finally, the computed result of the test is shown in Table 4.5 (NIST, 2015).

Table 4. 5 The result of McNemar's Test

Set of Parameters	Value
McNemar's Chi-square (1.0)	11.5294
p-value	0.0007
Cramér's V	0.2148

In conclusion, the decision is statistically significant (since $p < \alpha$) and given 95% confidence that it could not have happened if the null hypothesis was correct. As a consequence, the null hypothesis can reject in favor of the alternative. With a value of 0.2148, Cramér's V correlation suggests a significant relationship among the variants.



Group B (8% of all responders) Group C (19% of all responders)

Figure 4. 9 The student decision change in the presence of the PBS

As a supplementary fact, the students' decision change populated by groups related to the transportation option during their preferred period in the existence of PBS is tabulated (Table 4.6) and given in a radar chart (Figure 4.9).

As seen in the table, the bulk of the student population prefers to walk; hence decision changes primarily happened among those who like to walk.

Table 4. 6 The student's decision change by groups

Cell Group	Bike	Bus	Car	Taxi	Walk	Total
A	30	24	9	9	57	129
B	5	3	1	0	11	20
C	9	7	10	3	19	48
D	9	5	12	1	26	53
Total	53	39	32	13	113	250

In high point, the data show that increasing PBS frequencies in the city for the most desirable student periods could be beneficial (e.g., Friday and Saturday evenings after 20:00 h).

- **Bernoulli trial:**

Conducting a simple Bernoulli trial from the Bayesian probabilistic point of view would be helpful to compare statistic and machine learning paradigms. A Bernoulli trial is an entry point of the probability theory and the simplest case of the Binomial distribution.

The probability mass function (PMF) of a binomial distribution is represented as below:

$$P(\theta) = \binom{n}{k} f^k (1-f)^{n-k} \qquad (6)$$

Definition 6:

Assume f is the probability of success (exactly k scores in n attempts) and θ is the probability value.

The concept behind the Bayesian trial was to calculate the probability value that a single dependent variable can produce differently than the machine learning fundamental dataset features requirement. As a result, the outcome fitted a significant variation compared to the estimation value, which the independent features can create with the classification algorithms.

The principal purpose was to take samples from given simple Bernoulli distribution and use the cumulative density function (CDF, equal to the sum of PMF) to calculate all these samples' possible outcomes. Performing a Bayesian inference, including the famous MCMC, with Metropolis-Hastings simulation, is part of the analysis.

From the probabilistic point of view, the result must be the best estimation of the proportion of individuals who would respond "*yes*" on the question of - "*If you have more public bus options during the evening and night, would you use them frequently?*"

The variable represented by the "*SI*" label name in the dataset consisted of a 1 or 0. The traditional statistics calculated from the data was given:

The sample mean is around 0.708, the sample standard deviation is 0.455, and the MOE is 0.056. The question here is:

"*What is the probability of, given that the number of observed yes answers was 177 from 250 participants, parameter θ?*"

To answer this question, using Bayesian inference on Binomial distribution, the likelihood of θ can get given the data follows a Beta distribution, *Beta* (α, β). Then, the parameter α equals the number of "*yes*" answers, and parameter β equals the number of "*no*" responses.

Assume to set θ of a Binomial distribution could take any value ranging from 0 to 1 equally likely, uniform distribution prior, Beta(*1,1*) could use.

The beta distribution is also defined by a conjugate prior probability distribution for the Bernoulli trial from the Bayesian point of view (Steorts, 2016). The Bayesian formula was represented as below:

$$P(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \qquad (7)$$

Definition 7:

To perform a Bayesian estimation, constructing a posterior $p(\theta|Y)$ is needed with given a likelihood $p(Y|\theta)$, a prior $p(\theta)$, and the marginal likelihood $p(Y)$.

With the combination of the beliefs that there is a requirement for improvement on the public bus services in the city[34] and the survey study results[35], the probabilistic posterior could compute.

[34] Prior beliefs.
[35] The observations.

Furthermore, summarizing two theorems indicated that T is a sample of trials from the random variable Y. Therefore, the joint likelihood of observing any specific sample Y can be presented by:

$$P(Y|\theta) = \theta^s(1-\theta)^{T-s} \quad (8)$$

Definition 8:

The *(s)* denotes the number of observed "*yes*" and *(T - s)* denotes the number of perceived "*no*."

The equation is also compatible with the binomial description given above. The marginal likelihood *p(Y)* is a constant concerning θ. To achieve the posterior probability, define the functional forms of the likelihood, and the prior was required.

Computation was carried out with a simple MCMC simulation model. MCMC has used a class for sampling from a sample distribution probability and estimating the distribution of parameters given a set of observations. Metropolis-Hastings is a specific implementation of MCMC. The trial was simulated 1.000.000 times, and the first 25% of the tests (burn-in) were neglected. With the given information, the result is shown in Figure.4.10.

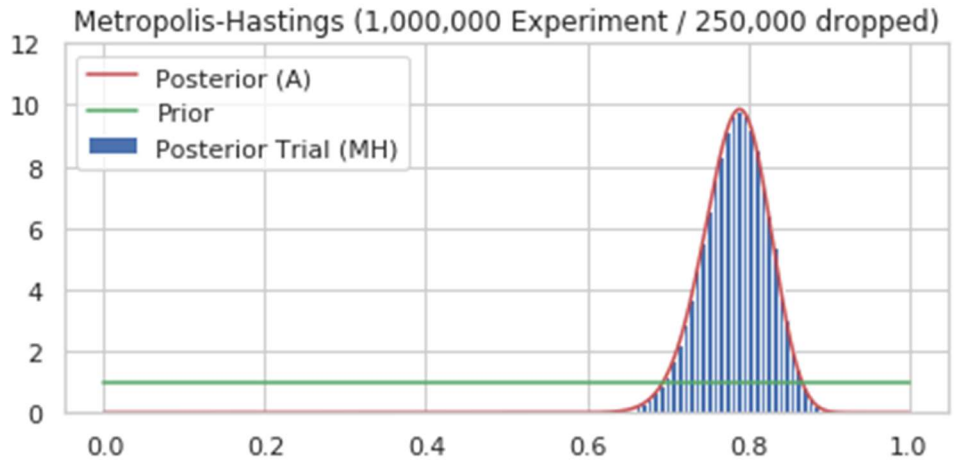


Figure 4. 10 The calculated posterior mean (0.784)

In conclusion, the posterior mean calculated round 0.784 (78.4%) was the forecast of students' bus preference possibility without the weight of independent features.

- **Classification Algorithms:**

This section discussed four machine learning classifiers' results[36] to predict the binary demand variable value based on the dataset features.

First, performance metrics were based on a 20% random sampling of test data in every trial. Then 10-fold cross-validation was performed to fine-tune the model parameters using the remaining 80% of the data. The factors in question were the tree's depth in random forests, the number of iterations in support vector machines, and the hidden layer size in neural networks.

Second, a pragmatic experiment was conducted with the selected subset in the student's dataset before moving to the broader examination. In the machine learning domain, the feature selection strategy reduces the number of available features to only those necessary to contribute the most to the prediction output (Jain, D. and V. Singh., 2018). One popular feature selection method is SelectKBest (Alaskar, L. et al., 2019).

SelectKBest is a feature selection approach for high-dimensional datasets to improve prediction accuracy or performance (Blesson, 2017). As the name implies, this strategy is best for large datasets, and SelectKBest removes all but the K highest scoring features (Palmer, C. et al., 2013). Table 4.7 shows the best predictors found by SelectKBest.

Table 4. 7 Selected top 10 features of SelectKBest techniques

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Label	FI	PTFL	PTFW	HH	Frdy	HM	Contind	Contcul	PTime	PTTtaxi
Chi^2	17.940	11.884	9.133	4.131	4.470	4.131	3.901	3.023	2.987	2.908

Additionally, Table A. 1, in the Appendix section, provides explanations for the abbreviations used in the tables.

However, with a 0.74 accuracy rating being even smaller than the Bernoulli trial result due to the small size of experimental data, the effect of the substantial number of features identified by SelectKBest on the classification model's accuracy was not as high as projected. Nonetheless, these are the factors at the forefront of situations where students feel the need for PBS.

[36] Logistic regression, support vector machines, random forests, and artificial neural networks.

Therefore, as explained above, due to the relatively small volume of students' data density, to use all dataset features and considered their effects on the estimation depending on each label's separate correlation over the target variable was preferred.

Using dummy variables to cast categorical variables into continuous applications in classification processes, the form of the data has changed. The dataset was made up of 250 records and 46 prediction features. The data was then normalized before being fitted to models. In conclusion, Table 4.8 lists the evaluation metrics for normalized data prediction.

Table 4. 8 Prediction metrics for classifiers with the normalized data

Classifier	F-Measure	Accuracy	AUC	Cohen's Kappa	Cross-Validation
LR	0.88	80%	0.81	0.39	0.82
Linear SVM	0.80	70%	0.71	0.25	0.76
RF	0.91	84%	0.61	0.25	0.81
MLP	0.67	56%	0.68	0.17	0.76

Again, the compared results were among models (AUC between 0.81 and 0.68). In the case of LR, the variations in Cohen's kappa metrics with significant gains were crucial. For instance, a higher accuracy value could indicate overfitting with the training set. As a result, because the test set accuracy is not skewed, it is more meaningful for evaluating the performance on unseen data. Furthermore, the log-loss (cross-entropy loss) characterizes the performance of an LR classifier, with the optimal value near 0. For instance, it was 0.48 in this model.

The RF algorithm, on the other hand, produced the best F-measure prediction results. Unfortunately, because of the tiny data quantity, the MLP classifier performs the poorest, as previously stated.

The LR and RF algorithms are preferred for practicality and reliability; moreover, they are frequently utilized with data mining growth in the information systems domain (Aulck, S. et al., 2017) (Alzen, L. et al., 2018).

Cohen's kappa results were one of the most influential metrics here. Furthermore, cross-validation may result in better average performance than using a single classification technique and reduce the probability of poor performance in practice (Schaffer, 1993).

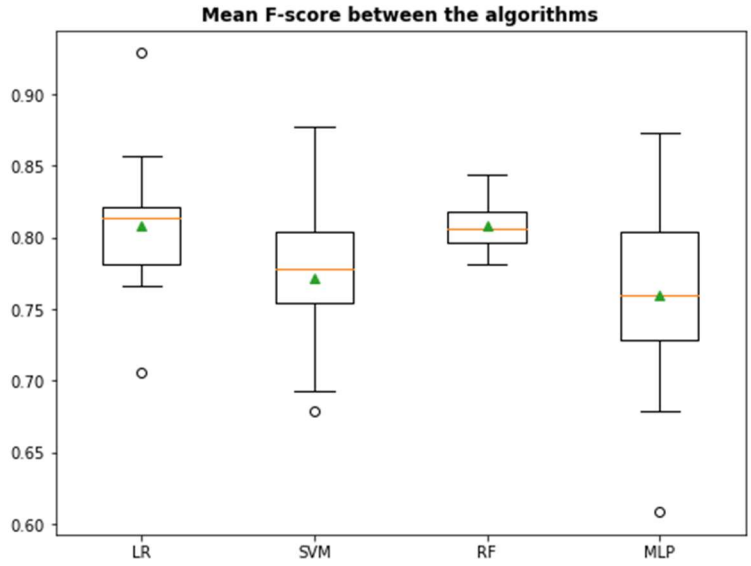


Figure 4. 11 Mean F1-score between the four algorithms

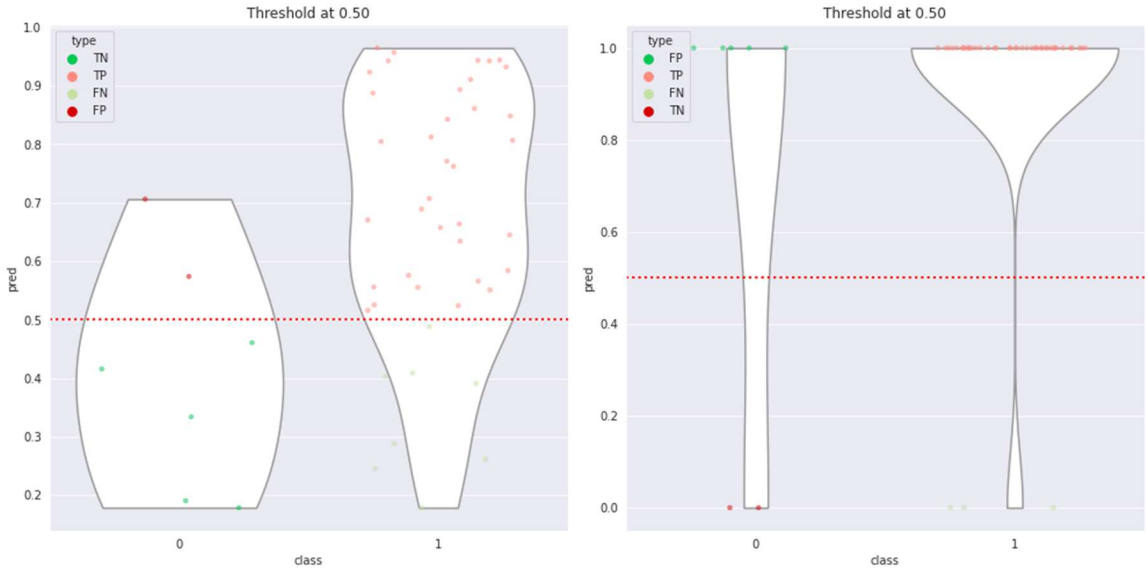
Figure 4.11 shows a graph of mean F-scores between the algorithms used to compare the models' performance. In the figure, LR and RF classifiers outperform the other algorithms. Between these two classifiers, a null hypothesis states that both methods perform equally well on the dataset. To verify this one, for instance, the calculated 5x2cv paired t-test results were as follows:

- The p-value = 0.52,
- The t-statistics = (- 0.67).

As a result, since $p > \alpha$, the null hypothesis cannot be rejected and infer that the performance of the two algorithms is not that significantly different when the chosen significance threshold level of α was 0.05 (Dietterich, 1998).

Finally, this experiment defined 82% of a reliable variation on-demand mobility selection with optimized PBS using all survey dataset attributes. Furthermore, classification is a practically applicable method in the real world's matters. Therefore, policymakers can use this strategy to forecast demand each year and adjust the PBS accordingly for on-demand mobility.

Additionally, a graph for class distribution differences of LR and RF classifiers for display purposes is given in Figure 4.12.



Logistic regression Random forest

Figure 4. 12 Class distribution plots of two classifiers

4.2.4 Conclusion

The community-specific social problems due to the nature of life may not prevent immediately anywhere in the world. However, the decline in life quality is preventable, even reversible, as long as the concepts and benefits are analyzed correctly.

This experiment compared simple statistical modeling with machine learning to implementing a real-life example to gain expertise in solving similar problems in new settings. The progress of artificial intelligence-related studies in social system-oriented issues affecting people's lives would provide a more visionary approach to evaluate the situation from a scientific standpoint.

On the other hand, the obtained result of 82% of variation may not be a steady estimator for the consideration of modernizing city bus schedules from a rigorous economic aspect of Japanese public transportation policy. However, suppose the bus services are improved; in that case, the student's tendency will increase towards using the buses, which is statistically significant and proved by a hypothesis test in this experiment.

Carrying out the necessary daily life activities in a city with comfort mainly depends on public transportation services efficiencies. Furthermore, revitalizing the local

areas with tourism initiatives also depends on enriching transportation possibilities to provide more tourists.

Finally, the areas desired by students to enrich around campus by authorities brought up to the light to a substantial extent:

- A supermarket was rated by 90.4% of the student's approval
- A chain restaurant (i.e., McDonald's) was rated by 83.2% of the student's approval

The concept is already one of the United Nations development programs definitions at goal 11: Sustainable cities and communities (United Nations, 2020).

"Truth is singular. Lies are words, words, and words. You met the Doctor, didn't you? "[37]

[37] Doctor Who. The Snowmen (2012). https://www.imdb.com/title/tt2380513/quotes/qt1902609

CHAPTER FIVE

KEY FINDINGS, CONCLUSIONS, AND IMPLICATIONS

5.1 Research Findings and Limitations Discussion

There are limited researches on public bus services (PBS) in rural areas or small towns in Japan, although it is growing. However, the peculiarities of rural areas or small cities have encouraged private cars, posing a financial barrier to the continuation of PBS.

In addition, there were three main reasons for people who were uninterested in using the bus: The first is the restricted number of buses available, and the second is the duration of the services (Santoso, S. et al., 2012).

Another significant issue come forward is the accessibility of bus stops. For instance, a survey measuring people's bus usage intention in Japan's rural city discovered a remarkable result. Almost half of the responders live within five minutes, and approximately 80% reside within ten minutes walking distance from a bus stop (Tran, Y. et al., 2020). Furthermore, the studies all agree that if demand is adequately met, people's willingness to use buses rather than own automobiles grows.

In adherence with previous research, this research emphasized "*why*" and "*how*" the transportation quality of rural city university students needs to be enhanced based on demand. Otherwise, from a rigorous business perspective, it would not be sensible to discuss other details in an environment where the market could not meet the prime demand. Therefore the critical findings of this study can be summarized:

- First, a robust administrative and social culture regarding practical usage of buses inside the public transportation systems has not evolved well in Japan, and the situation has somehow escalated over time.
- Second, existed bus services are somehow considered an extension of train lines due to a perception that has existed since the past. On this account, their accessibility is not flexible. This situation pushes people to provide flexibility in urban transportation, especially in rural areas, by personal vehicles or taxis.

- Third, although the presented set of data is small and its applicability is considered insufficient to generalize with very high critical perceptions, it still provided extensive evidence for the relevance of the mentioned issues with detailed data analyses to regional development decision-makers.

Therefore, while transportation is one of the basic requirements of daily mobility in urban life, the priority of this thesis has been to achieve a social reconciliation by drawing attention to the importance of this public service with the findings it has obtained. Furthermore, the motivating factors come from a purely humanitarian perspective to assist in transforming local cities of Japan, which are on the verge of pressure from globalization policies.

In this way, this thesis, in the end, provided an entry-level micro-analysis result, especially for urban mobility studies where secondary data were unavailable or limited. However, if the suggested procedures of this study need to be placed in a more generic context, then it must respond to the following requirements:

First, a population's sample size is critical to avoid any scientific suspicions about the relevance of the statistic's accuracy in the research. Thus, the sample size of 250 students from an 1800-student group is noteworthy. Still, it falls short of the exact desired confidence level[38].

The low response rate of Japanese is a challenge for this assignment. Social and economic factors can influence the response rate; for instance, voter turnout, a measure of citizens' participation in the political process, was 53% in recent elections, lower than the OECD average of 68% (OECD Better Life Index, 2021).

Second, based on a year-ago student opinion poll, the proposed sequences of techniques provide a short foresight into next year's PBS demand prediction. That implies ensuring a data flow by acquiring new survey data from newly enrolled students each year is critical to the method's success. This way, the low response rate in terms of reliability, non-random response model, and possible selection bias can be examined.

[38] The minor requirement of a fair 5% MOE for a survey population of 1800 people is 317 responses when population proportion value is considered equivalent.

On the other hand, based on each year's new conditions, the survey question design and order may need to be updated, enabling the city's transportation bureaus to manage their decisions more interactively to match passenger demand with their services.

Third, the critical topic in this case study scenario was PBS demand analysis. Therefore, at least one additional test with a different case scenario and a different topic is required on the appropriate quality of the procedure.

It is also worth noting that the impact of unexpected occurrences in the historical progression of humanity, such as social or cultural upheaval, and like an environmental or viral disease, should not be overlooked in research that tries to make predictions about the coming years. Nevertheless, unfortunately, this very research also has experienced a double whammy of unexpected events during the development phases[39].

Against all odds, this study discovered that expanding university students' mobility makes it easier for them to attend more social activities after school hours, enhancing their social skills, which is a bit problem in Japan. The summary of the findings for the three research objectives was listed as follows:

- Initially, the current demand rate of 60%, as assessed by the two-step verification probe, was perfectly compatible with the real-life scenario in the Japanese countryside. PBS has a comparatively low demand. The secondary probing, on the other hand, increased the rate to 71%;
- The first and secondary probing findings were subjected to a hypothesis test to resolve two mutually exclusive statements. Based on the obtained student data sample, the final choice was statistically confirmed and highly supported the premise that the PBS has to be improved to meet students' mobility requirements;
- Following the implementation of numerous machine learning-based prediction methods to determine if an increase in demand will occur, the best and most reliable demand prediction was set at 82%. Additionally, a variation of 20% was presented to raise the need for PBS using a simple sequence of questions.

[39] A personal one and a more global COVID-19 outbreak.

Additionally, Table B. 1, in the Appendix section, provides explanations for the survey questions.

As a result, this thesis collected and evaluated the data systematically, unbiased, and solution-oriented to assist policymakers in making practical judgments because a crucial element of decision-making is providing long-term and reliable data (Goeldner, C. and J, Ritchie., 2011).

5.2 Current Policies and Implications

In general, statistics derived from various sources, including survey information, are always needed to comprehend a region's actual transportation and traffic constraints fully. Based on this statistical data, the following is a comprehensive appraisal of present and proposed solution-oriented strategies for boosting PBS over private vehicles:

- In an uncertain demand level, on-demand transportation services can run more economically, such as driverless shuttles;
- E-bikes for students and tourists and e-wheelchairs for the elderly are examples of micro-mobility solutions that suit public transportation needs.

A demand-based and situation-specific approach may be more practical for rural communities with varying geographical, cultural, and population characteristics. For instance, hosting a university is beneficial to a community's economy, as this study shows, but not every rural city has this opportunity.

Promoting more public transportation instruments while still operational, rather than personal vehicles, is an excellent strategy to circumvent the current issue gradually. Simultaneously, as passenger demand for public transportation grows, so does the frequency and quality of services.

Research into the future applicability of mobility as a service in Japan is still ongoing. However, it stressed the importance of providing nationwide transportation data by standardizing and opening it to the researchers was (MLIT, 2017).

Improved self-driving cars providing a promising solution to passenger transportation issues in rural locations can radically alter the transportation landscape. However, there will be times when autonomous vehicles fail to meet transportation needs. For instance, autonomous cars struggle to operate in bad weather like fog or snow, and reflecting road surfaces from rain and ice pose further hurdles for sensors and driving operations (Fagnant, J. and K, Kockelman., 2015).

Higashihiroshima City is implementing an experimental project in partnership with private efforts to tackle Hiroshima University's traffic problems, provide a loop bus service, and collect data for a future automated driving society as part of an ongoing investigation (Hiroshima University, 2019). A fixed-route loop bus in communities with no regular access to current public transportation, like in this case, can be an example of best practices.

Finally, this research concluded that public transportation is necessary for a community to engage in mobility-based everyday activities. It will grow more comfortable if students and other inhabitants such as foreign laborers, travelers, domestic tourists, and the elderly, who make up a significant portion of the city population, choose it more frequently. In equilibrium, the quantity of a good supplied by producers equals the amount reversibly sought by consumers (Britannica, Encyclopedia., 2019).

Last but not least, this study will perhaps be repeated in a different country with a new participant group and a very different scenario in the coming years. Therefore, it will be necessary to increase the performance of the design and the number of participants to prove its efficiency.

APPENDICES

Table A. 1 The used abbreviations in tables

No	Acronyms	Explanation
1	SexF	Label of female responder
2	SexM	Label of male responder
3	OJ	Label of a Japanese student who is origin different than Hokkaido
4	OH	Label of a Japanese student who is origin Hokkaido
5	OI	Label of an international student
6	Lapart	Label of a student who is residing in an apartment flat
7	Lblodge	Label of a student who is residing in a boarding house with a meal
8	Ldorm	Label of a student who is residing university dormitory
9	Lhome	Label of a student who is residing with their parent's home
10	Dcook	Label of a student who is cooking dinner for themselves
11	Dcafe	Label of a student who is dining at the university canteen
12	Dbmeal	Label of a student who is dining at the boarding house
13	Dbento	Label of a student who purchases a lunch box (bento) for dinner
14	Drstrnt	Label of a student who prefers a restaurant for dinner
15	PTime	Label of a student who is doing a part-time job
16	PTFD	Label of a student whose public transport frequency is daily
17	PTFW	Label of a student whose public transport frequency is weakly
18	PTFM	Label of a student whose public transport frequency is monthly
19	PTFL	Label of a student whose public transport frequency is less than a month
20	Car	Label of a student who has a car
21	FI	First inquiry: label of a student who wants to use more public buses initially
22	Frdy	Friday: the preferred day of the week to go out for entertainment
23	Strdy	Saturday: the preferred day of the week to go out for entertainment
24	Sndy	Sunday: the preferred day of the week to go out for entertainment
25	HL	Hours between 18:00 and 20:00 segment
26	HM	Hours between 20:00 and 22:00 segment
27	HH	Hours between 22:00 and late segment
28	PTTwalk	Label of a student whose preferred transport type is walk
29	PTTbike	Label of a student whose preferred transport type is a bicycle
30	PTTbus	Label of a student whose preferred transport type in a bus
31	PTTcar	Label of a student whose preferred transport type is car
32	PTTtaxi	Label of a student whose preferred transport type is a taxi
33	**SI**	Secondary inquiry: label of a student who wants to use public buses finally
34	Conteco	The university has a significant contribution to the city's economy
35	Contedu	The university has a significant contribution to the city's education system
36	Contaca	The university has a significant contribution to the academy
37	Contind	The university has a significant contribution to the city's industry
38	Contcul	The university has a significant contribution to the city's culture
39	ValueH	Evaluation of the university by a student (high)
40	ValueM	Evaluation of the university by a student (medium)
41	ValueL	Evaluation of the university by a student (low)
42	WinterH	The negative effect of the winter season on student life (high)
43	WinterM	The negative effect of the winter season on student life (medium)
44	WinterL	The negative effect of the winter season on student life (low)
45	Mc	Label of a student who prefers a chain restaurant near campus
46	SpM	Label of a student who prefers a supermarket near campus
47	Grad	Label of a student who remains if a career opportunity exists after graduation

Table B. 1 The used questions in the survey

No	Question
1	What is your sex?
2	Where are you from?
3	What kind of residence do you live in?
4	Where do you usually have dinner?
5	Do you have a part-time job?
6	How often do you use the public bus services of Kitami city?
7	Do you have a car?
8	Would you like to use the public bus services of Kitami city more often?
9	Which days of the week would you like to go out the most?
10	During which periods would you like to go out?
11	Which transportation option are you using during those periods?
12	If public bus services would be available until late at night, would you use them often?
13	In what way is the institute of technology contributing to Kitami city the most?
14	How would you evaluate the importance of the institute of technology for Kitami city?
15	Do you think the winter season makes your life challenging in Kitami city?
16	Do you think it would be helpful to have a restaurant like McDonald's around the campus?
17	Do you think it would be helpful to have a supermarket around the campus?
18	Would you stay in Kitami city after graduation if there are any career opportunities for you?

REFERENCES

AIRIA. (2021). *Automobile Inspection and Registration Information Association.* Retrieved from https://www.airia.or.jp/publish/statistics/number.html

Akira, I. (2017). A Turning Point for Tourism Informatics. *New Breeze, 29*, 1–4.

Alaskar, L. et al. (2019). Employee turnover prediction using machine learning. In A. M. Alfaries, *Advances in Data Science, Cyber Security and IT Applications* (pp. 301–316). Cham, Switzerland: Springer. doi:10.1007/978-3-030-36365-9_25

Albino, S. (2015). Tourism in University Cities. The Role of Universities in Place Branding. Exeter, UK: University of Exeter. Retrieved July 6, 2015, from https://ore.exeter.ac.uk/repository/handle/10871/21034

Almeida, A. et al. (2013). Segmentation by benefits sought: the case of rural tourism in Madeira. *Current Issues in Tourism, 17*, 813–831. doi:10.1080/13683500.2013.768605

Alzen, L. et al. (2018). A logistic regression investigation of the relationship between the Learning Assistant model and failure rates in introductory STEM courses. *International Journal of Stem Education, 5*, 1–12.

Aoki, E. (2000). *A History of Japanese Railways, 1872–1999.* Tokyo, Japan: East Japan Railway Culture Foundation. Retrieved from https://www.ejrcf.or.jp/jrtr/history/index_history.html

Archer, M. (2003). *Structure, Agency and the Internal Conversation.* Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139087315

Arlot, S. and A, Celisse. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40–79.

Ashley, C. et al. (2002). Methodology for Pro-Poor Tourism Case Studies. *10*.

Aulck, S. et al. (2017). Predicting Student Dropout in Higher Education. *arXiv:1606.06364*. Retrieved from https://arxiv.org/abs/1606.06364

Azzali, S. and E, Sabour. (2018). A framework for improving sustainable mobility in higher education campuses: the case study of Qatar University. *Case Studies on Transport Policy, 6*, 603–612. doi:10.1016/j.cstp.2018.07.010

Bach, M. et al. (2015). What Is Exadata? In *Expert Oracle Exadata* (pp. 1-20). Berkeley, CA: Apress. doi:10.1007/978-1-4302-6242-8_1

Barnard, R. (2008). 3.2.2 What is wrong with Reliability Engineering? *INCOSE International Symposium*, *18*, pp. 357-365. doi:10.1002/j.2334-5837.2008.tb00811.x

Beaulac, C. and J. Rosenthal. (2019). Predicting university students' academic success and major using random forests. *Research in Higher Education, 60*, 1048–1064. doi:10.1007/s11162-019-09546-y

Behnel, S. et al. (2011). Cython: The Best of Both Worlds. *Computing in Science & Engineering, 13*, 31-39. doi:10.1109/MCSE.2010.118

Berger, J. O. (1989). Statistical Decision Theory. In M. M. Eatwell J. (Ed.), *Game Theory* (pp. 217-224). London: The New Palgrave. Palgrave Macmillan. doi:10.1007/978-1-349-20181-5_26

Bethapudi, A. (2013). The Role of ICT in Tourism Industry. *Journal of Applied Economics and Business, 1*(4), 67–79.

Blesson, G. (2017, January- June). A Study of the Effect of Random Projection and other Dimensionality Reduction Techniques on Different Classification Methods. *J. Interdiscip. Stud. Res. Baselius Coll., XVIII No.1*, 69-75. Retrieved from https://baselius.ac.in/wp-content/uploads/2017/05/Baselius-Journal-Jan-June-2017.pdf

Bonnefon, J. et al. (2016). The social dilemma of autonomous vehicles. *Science, 352*, 1573–1576. doi:10.1126/science.aaf2654

Bose, J. (2001). Nonresponse Bias Analyses at the National Center for Education Statistics. *Proceedings of Statistics Canada Symposium*, *Issue Number: 2001001.*

Retrieved from https://www150.statcan.gc.ca/n1/en/catalogue/11-522-X20010016269

Branco, P. et al. (2015). A Survey of Predictive Modelling under Imbalanced Distributions. *ArXiv, abs/1505.01658*.

Breiman, L. (2004). Random Forests. *Machine Learning, 45*, 5–32. doi:10.1023/a:1010933404324

Brennan, J. and A, Cochrane. (2019). Universities: in, of, and beyond their cities. *Oxford Review of Education, 45*, 188–203. doi:10.1080/03054985.2018.1551198

Briedenhann, J. and E. Wickens. (2004). Tourism routes as a tool for the economic development of rural areas—vibrant hope or impossible dream? *Tourism Management, 25*, 71–79. doi:10.1016/s0261-5177(03)00063-3

Britannica, Encyclopedia. (2019). *Supply and demand.* Retrieved from Britannica, T. Editors of Encyclopaedia.: https://www.britannica.com/topic/supply-and-demand

Calimente, J. (2012). Rail integrated communities in Tokyo. *Journal of Transport and Land Use, 5*, 19–32. doi:10.5198/jtlu.v5i1.280

Camargo-Perez, J. et al. (2015). Multi-criteria approaches for urban passenger transport systems: a literature review. *Annals of Operations Research, 226*, 69–87. doi:10.1007/s10479-014-1681-8

Cao, Y. et al. (2013). An MLP classifier for prediction of HBV-induced liver cirrhosis using routinely available clinical parameters. *Disease markers, 35*(6), 653–660. doi:10.1155/2013/127962

Castillo, D. and F, Benítez. (2012). A Methodology for Modeling and Identifying Users Satisfaction Issues in Public Transport Systems Based on Users Surveys. *Procedia - Social and Behavioral Sciences, 54*, 1104–1114. doi:10.1016/j.sbspro.2012.09.825

Castillo, M. and F, Benítez. (2013). Determining a public transport satisfaction index from user surveys. *Transportmetrica A: Transport Science, 9*, 713–741. doi:10.1080/18128602.2011.654139

Chawla, N. et al. (2002). SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res., 16*, 321–357. doi:10.1613/jair.953

Chen, B. et al. (2018). Tourism's impacts on rural livelihood in the sustainability of an aging community in Japan. *Sustainability, 10*, 2896. doi:10.3390/su10082896

Codd, E. (1970). A Relational Model of Data for Large Shared Data Banks. *13*(6), 377–387. doi:10.1145/362384.362685

Connelly, L. (2020). Logistic Regression. *MEDSURG Nursing, 29*, 353–354.

Couronne, R. et al. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics, 19*, 1–14. doi:10.1186/s12859-018-2264-5

Cramér, H. (1946). *Mathematical Methods of Statistics (PMS-9).* Princeton, NJ, USA: Princeton University Press.

Cruz-Rodríguez, J. et al. (2020). Analysis of Interurban Mobility in University Students: Motivation and Ecological Impact. *International Journal of Environmental Research and Public Health, 17*, 9348. doi:10.3390/ijerph17249348

Davidson-Pilon, C. (2015). *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference.* Addison-Wesley Professional.

Dickens, M. and M. Hughes-Cromwick. (2019). *Leveraging Big Data in the Public Transportation Industry.* New York, NY: The American Public Transportation Association (APTA). Retrieved from https://www.apta.com/wp-content/uploads/Big-Data-Policy-Brief.pdf

Dietterich, T. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation, 10*, 1895–1923. doi:10.1162/089976698300017197

Dimitrovski, D. et al. (2012). Rural Tourism and Regional Development: Case Study of Development of Rural Tourism in the Region of Gruţa, Serbia. *Procedia environmental sciences, 14*, 288–297. doi:10.1016/j.proenv.2012.03.028

Dominik, L. and M, Avary. (2020). *White Paper on Transforming Rural Mobility in Japan and the World.* Cologny, Switzerland: World Economic Forum. Retrieved from https://www.weforum.org/whitepapers/transforming-rural-mobility-in-japan-and-the-world

Eder, J. et al. (2010). Exploring factors influencing student study abroad destination choice. *Journal of Teaching in Travel & Tourism, 10*, 232–250. doi:10.1080/15313220.2010.503534

Ergun, C. (2014). Üniversite ve Kent İlişkisi Üzerine Görüşler: Mehmet Akif Ersoy Üniversitesi Örneği. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 216-237. doi:10.21764/EFD.05947

Fagnant, J. and K, Kockelman. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice, 77*, 167–181. doi:10.1016/j.tra.2015.04.003

Farkhondehzadeh, A. et al. (2013). E-Tourism: The role of ICT in tourism industry. *European Online Journal of Natural and Social Sciences, 2*, 566–573. Retrieved from https://european-science.com/eojnss/article/view/451/pdf

Fellesson, M. and M. Friman. (2011). Perceived Satisfaction with Public Transport Service in Nine European Cities. *Journal of the Transportation Research Forum, 47*, 1–12. doi:10.5399/osu/jtrf.47.3.2126

Feng, X. et al. (2012). Towards a unified architecture for in-RDBMS analytics. *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 325–336). New York, NY, USA: Association for Computing Machinery. doi:10.1145/2213836.2213874

Fotiadis, A. et al. (2014). Measuring Dimensions of Business Effectiveness in Greek Rural Tourism Areas. *Journal of Hospitality Marketing & Management, 23*, 21–48. doi:10.1080/19368623.2012.746931

Fox, A. and E. Brewer. (1999). Harvest, yield, and scalable tolerant systems. *Proceedings of the Seventh Workshop on Hot Topics in Operating Systems*, (pp. 174-178).

Gabor, M. (2015). A content analysis of rural tourism research. *Journal of Tourism, Heritage & Services Marketing, 1*(1), 25–29. Retrieved from https://mpra.ub.uni-muenchen.de/77416/1/MPRA_paper_77416.pdf

Gertsbakh, I. (1989). *Statistical Reliability Theory.* Taylor & Francis.

Goeldner, C. and J, Ritchie. (2011). *TOURISM Principles, Practices, Philosophies* (12th ed.).

Gold, N. et al. (2014). Cultural differences in responses to real-life and hypothetical trolley problems. *Judgment and Decision Making, 9*, 65–76. Retrieved from http://journal.sjdm.org/12/121101/jdm121101.pdf

Golob, T. and M, Makarovič. (2018). Student Mobility and Transnational Social Ties as Factors of Reflexivity. *The Social Sciences, 7*, 46. doi:10.3390/socsci7030046

Gustafsson, B. et al. (2017). China's emerging global middle class. *Centre for Human Capital and Productivity (CHCP)*. doi:10.13140/RG.2.2.21467.67364

H. Kharas and D. Meagan. (2020). China's Influence on the Global Middle Class. *Global Governance and Norms*. Retrieved 2020, from https://www.brookings.edu/wp-content/uploads/2020/10/FP_20201012_china_middle_class_kharas_dooley.pdf

Hammond, S. et al. (1998). The hidden traps in decision making. *Harvard business review, 76 5*, 47-8, 50, 52 passim. Retrieved from https://hbr.org/1998/09/the-hidden-traps-in-decision-making-2

Harris, H. and B. Nicol. (2013). SQL/DS: IBM's First RDBMS. *IEEE Annals of the History of Computing, 35*(2), 69-71. doi:10.1109/MAHC.2013.28

Hiroshima University. (2019). *Hiroshima University Loop Bus.* Hiroshima: Hiroshima University. Retrieved from https://www.hiroshima-u.ac.jp/system/files/139225/About_Hiroshima_University_Loop_Bus.pdf

Hoerler, R. et al. (2020). What are the factors and needs promoting mobility-as-a-service? Findings from the Swiss Household Energy Demand Survey (SHEDS). *European Transport Research Review, 12*, 1–16.

Hoffman, M. and A. Gelman. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15*, 1593-1623.

IBM Skills Network Labs. (2018). *IBM Developer Skills Network* . Retrieved from https://labs.cognitiveclass.ai/

Ibrahim, H. et al. (2020). Understanding Users' Intention to Use Park-and-Ride Facilities in Malaysia: The Role of Trust as a Novel Construct in the Theory of Planned Behaviour. *Sustainability, 12*, 2484. doi:10.3390/su12062484

Imashiro, M. (1996). Maintaining Public Transport in Japan's Countryside - Burden Sharing and Subsidies. *Japan Railway & Transport Review, 9*, 2–8. Retrieved from https://www.ejrcf.or.jp/jrtr/jrtr09/pdf/f02_ima.pdf

Ioannidis, J. (2018). The Proposal to Lower P Value Thresholds to .005. *JAMA, 319*, 1429–1430. doi:10.1001/jama.2018.1536

Jain, D. and V. Singh. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal, 19*, 179–189. doi:10.1016/j.eij.2018.03.002

James, B. et al. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal, 19*, 43–50.

James, P. (2014). *Urban Sustainability in Theory and Practice: Circles of sustainability* (1st ed.). London: Routledge. doi:10.4324/9781315765747

JAOS. (2019). *Statistical Report on Japanese Studying Abroad.* Tokyo, Japan: Japan Association of Overseas Studies. Retrieved from https://www.jaos.or.jp/wp-content/uploads/2020/03/JAOS-Survey-2019_-JapaneseStudent191219.pdf

JASSO. (2020). *Result of an Annual Survey of International Students in Japan.* Tokyo, Japan: Japan Student Services Organization. Retrieved from https://www.studyinjapan.go.jp/en/_mt/2021/04/date2020z_e.pdf

JNTO. (2021). *Japan National Tourism Organization, Japan Weather Forecast, Kitami (Hokkaido).* Retrieved from https://www.jnto.go.jp/weather/eng/area_detail.php?area_id=1720

JTA. (2016). *Statistical Informations.* Tokyo: Japan Tourism Agency. Retrieved from https://www.mlit.go.jp/kankocho/en/siryou/toukei/index.html

JTA. (2020). *White Papers on Tourism.* Tokyo, Japan: Japan Tourism Agency. Retrieved from https://www.mlit.go.jp/kankocho/en/siryou/whitepaper.html

Kakumoto, R. (1997). Transportation Investment and Japan's Experience. *Japan Railway & Transport Review, 11*, 4–12. Retrieved from https://www.ejrcf.or.jp/jrtr/jrtr11/pdf/f04_kak.pdf

Kato, H. (2014). Declining population and the revitalization of local regions in Japan. *Meiji J. Political Sci. Econ., 3*, 25–35.

Kawasaki, S. (2015). The challenges of transportation/traffic statistics in Japan and directions for the future. *IATSS Research., 39*, 1–8. doi:10.1016/j.iatssr.2015.06.002

Kawazaki, M. and Luczak, D. (2020). *A roadmap for repairing rural mobility in Japan - and beyond.* Cologny, Switzerland: World Economic Forum. Retrieved from https://www.weforum.org/agenda/2020/01/japans-much-admired-public-transit-system-is-leaving-its-rural-areas-behind/

Keller, A. (2014). *What is an acceptable survey response rate?* Retrieved from National Social Norms Institute at Michigan State University (NSNC): http://socialnorms.org/what-is-an-acceptable-survey-response-rate/

Kingma, P. and J. Ba. (2015). Adam: A Method for Stochastic Optimization. *CoRR, abs/1412.6980*(arXiv:1412.6980).

KIT. (2019). *Kitami Institute of Technology.* Kitami. Retrieved from https://www.kitami-it.ac.jp/about/students/

KitaBus. (2021). *Kitami Bus time-charts.* Retrieved from https://www.h-kitamibus.co.jp/

Kleinberg, J. et al. (2017). Human Decisions and Machine Predictions. *The quarterly journal of economics, 133,1*, 237–293. doi:10.1093/qje/qjx032

Kozak, M. et al. (2007). The impact of the perception of risk on international travellers. *International Journal of Tourism Research, 9*, 233–242. doi:10.1002/jtr.607

Kroese, P. et al. (2014). Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics.*, 386-392. doi:10.1002/wics.1314

Law, R. et al. (2014). Progress on information and communication technologies in hospitality and tourism. *International Journal of Contemporary Hospitality Management, 26*, 727–750. doi:10.1108/IJCHM-08-2013-0367

Leo, D. and F, Sardanelli. (2020). Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *European Radiology Experimental, 4*, 1–8. doi:10.1186/s41747-020-0145-y

Liu, S. et al. (2020). P-Value as the strength of evidence measured by confidence distribution. *arXiv: Methodology*(arXiv:2001.11945). Retrieved from https://arxiv.org/abs/2001.11945

Llewellyn-Smith, C. and V. McCabe. (2008). What is the attraction for exchange students: the host destination or host university? Empirical evidence from a study of an Australian university. *International Journal of Tourism Research, 10*, 593–607. doi:10.1002/jtr.692

Lohmann, G. and A, Netto. (2017). *Tourism Theory: Concepts, Models and Systems.* CABI Publishing. doi:10.1079/9781780647159.0000

Long, P. and Lane, B. (2000). Rural tourism development. In W. C. Gartner (Ed.), *Trends in outdoor recreation, leisure and tourism* (pp. 299–308). CABI Publishing. doi:10.1079/9780851994031.0299

Lu, Y. et al. (2016). A new revised version of McNemar's test for paired binary data. *Communications in Statistics - Theory and Methods, 46*, 10010–10024. doi:10.1080/03610926.2016.1228962

Luminousmen. (2020, April 30). *Demystifying hypothesis testing*. Retrieved from Luminousmen: https://luminousmen.com/post/demystifying-hypothesis-testing

Lynch, K. (2008). What Is the Form of a City, and How Is It Made? *Urban Ecology*. doi:10.1007/978-0-387-73412-5_44

Macioszek, E. et al. (2020). The bike-sharing system as an element of enhancing sustainable mobility — A case study based on a City in Poland. *Sustainability, 12*, 3285. doi:10.3390/su12083285

Manyika, J. et al. (2011). *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute.

McHugh, M. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica, 22*, 276–282. doi:10.11613/bm.2012.031

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*, 153–157. doi:10.1007/bf02295996

Mietke, F. et al. (2006). Analysis of the Memory Registration Process in the Mellanox InfiniBand Software Stack. In W. W. Nagel W.E. (Ed.), *Euro-Par 2006. Lecture Notes in Computer Science. 4128*, pp. 124-133. Berlin, Heidelberg: Springer. doi:10.1007/11823285_13

Mitchell, G. (2020). Sustainability. (T. Brock, Ed.) *Sustainable Investing*. Retrieved from https://www.investopedia.com/terms/s/sustainability.asp

Miyake, A. et al. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology, 41*, 49-100. doi:10.1006/cogp.1999.0734

MLIT. (2008). *Promotion of Urban Renovation and Compact Cities.* Tokyo, Japan: Ministry of Land, Infrastructure, Transport and Tourism. Retrieved from https://www.mlit.go.jp/common/000996976.pdf

MLIT. (2017). *Proposal by the Round-table Conference of Considering Future Vision of Rejuvenation and Revitalization of Local Public Transport.* Tokyo, Japan: Ministry of Land, Infrastructure, Transport and Tourism.

Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective.* United Kingdom: MIT Press.

Nakaniwa, C. (2006). *Guideline for Information and Communication Technology (ICT) Eco-Efficiency Evaluation.* Tokyo, Japan: The Japan Forum on Eco-efficiency. Retrieved March 2006, from https://lca-forum.org/english/eco/pdf/05.pdf

Nandimath, J. et al. (2013). Big data analysis using Apache Hadoop. *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)* (pp. 700-703). San Francisco, CA, USA: IEEE. doi:10.1109/IRI.2013.6642536

Nghiêm-Phú, B. (2015). Japanese university students' life satisfaction and their intentions to travel/study abroad. *EJCJS, 15*(2).

Nikiforiadis, A. et al. (2020). Assessing the Impact of COVID-19 on Bike-Sharing Usage: The Case of Thessaloniki, Greece. *Sustainability, 12*, 8215. doi:10.3390/su12198215

NIST. (2015). *McNemar test.* Retrieved from The National Institute of Standards and Technology: https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/mcnemar.htm

Obolensky, N. (2007). Chaos Leadership and Polyarchy - countering leadership stress? (e. University of Exeter, Ed.) *Centre for Leadership Studies*.

OECD. (2016). "Japan". In *OECD Regional Outlook 2016: Productive Regions for Inclusive Societies.* Paris: OECD Publishing. doi:10.1787/9789264260245-32-en

OECD. (2016). Japan: Boosting Growth and Well-being in an Ageing Society. In *Better Policies.* Paris: OECD Publishing. doi:10.1787/9789264256507-en

OECD Better Life Index. (2021). *"Japan"*. OECD Publishing. Retrieved from http://www.oecdbetterlifeindex.org/countries/japan/

Okano, Y. (1994). The Backdrop to Privatisation in Japan. *Japan Railway & Transport Review, 2*, 2–9. Retrieved from https://www.ejrcf.or.jp/jrtr/jrtr02/pdf/f02_oka.pdf

Oracle. (2021). *Exadata Database Machine*. Retrieved from Oracle: https://www.oracle.com/database/technologies/exadata.html

Palmer, C. et al. (2013). Faster tissue interface analysis from Raman microscopy images using compressed factorization. *In Proceedings of the European Conferences on Biomedical Optics.* Munich, Germany. doi:10.1117/12.2032817

Pamplona, A. and A, Oliveira. (2016). Economic Indicators For The Public Transportation Aggregate Demand Estimation In São Paulo. *Journal of Urban and Environmental Engineering, 10*(2), 169–176. doi:10.4090/JUEE.2016.V10N2.169176

Piri, S. et al. (2018). A synthetic informative minority over-sampling (SIMO) algorithm leveraging support vector machine to enhance learning from imbalanced datasets. *Decis. Support Syst., 106*, 15–29. doi:10.1016/j.dss.2017.11.006

Plonsky, O. et al. (2019). Predicting human decisions with behavioral theories and machine learning. arXiv:1904.06866. Retrieved from https://arxiv.org/abs/1904.06866

Politis, I. et al. (2020). Shifting to Shared Wheels: Factors Affecting Dockless Bike-Sharing Choice for Short and Long Trips. *Sustainability, 12*, 8205. doi:10.3390/su12198205

Preger, R. (2012). The Oracle Story, Part 1: 1977-1986. *IEEE Annals of the History of Computing, 34*(4), 51-57. doi:10.1109/MAHC.2012.54

Public Relations Office. (2015). *Overview of the Public Opinion Survey on the Life of the People.* Cabinet Office. Tokyo: The Government of Japan. Retrieved from https://www.gov-online.go.jp/eng/pdf/summaryk15.pdf

Public Relations Office. (2018). *Overview of the Public Opinion Survey on the Life of the People.* Cabinet Office. Tokyo: The Government of Japan. Retrieved from https://www.gov-online.go.jp/eng/pdf/summaryk18.pdf

Puchongkawarin, C. and K. Ransikarbum. (2020). An integrative decision support system for improving tourism logistics and public transportation in Thailand. *Tourism planning and development*, 1–16. doi:10.1080/21568316.2020.1837229

Rana, R. and R. Singhal. (2015). Chi-square test and its application in hypothesis testing. *Journal of the Practice of Cardiovascular Sciences, 1*, 69. doi:10.4103/2395-5414.157577

Reid, D. (2006). Consumer change in Japan: A longitudinal study. *Thunderbird International Business Review, 49*, 77–101. doi:10.1002/tie.20132

Reid, S. (2007). Identifying social consequences of rural events. *Event Management, 11*, 89–98. doi:10.3727/152599508783943192

ReliaSoft. (2001, September). The Reliability Function. *The eMagazine for the Reliability Professional*(7). Retrieved from https://weibull.com/hotwire/issue7/relbasics7.htm

ReliaSoft. (2004, May). Foundations of Reliability. *The eMagazine for the Reliability Professional*(39). Retrieved from https://www.weibull.com/hotwire/issue39/relbasics39.htm

Research and Markets. (2021). *Big Data Market by Leading Companies, Solutions, Use Cases, Infrastructure, Data Integration, IoT Support, Deployment Model and Services in Industry Verticals 2021 - 2026.* Global: Mind Commerce.

Rindfuss, R. et al. (2015). Do low survey response rates bias results? Evidence from Japan. *Demographic Research, 32*, 797–828. doi:10.4054/demres.2015.32.26

Ritchie, B. et al. (2005). *Tourism research methods: integrating theory with practice.* CABI Publishing. doi:10.1079/9780851999968.0000

Rosenfeld, A. et al. (2012). Combining psychological models with machine learning to better predict people's decisions. *Synthese, 189*, 81–93. doi:10.1007/s11229-012-0182-z

Runco, M. et al. (1999). *Encyclopedia of Creativity* (1st ed.). (S. Pritzker, Ed.) Academic Press.

Sakai, K. (2020). Public transport promotion and mobility-as-a-service. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci., 103-A*, 226–230. doi:10.1587/transfun.2019tsi0001

Salvatier, J. et al. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*. doi:10.7717/peerj-cs.55

Santoso, S. et al. (2012). Opportunities and strategies for increasing bus ridership in rural Japan: A case study of Hidaka City. *Transport Policy, 24*, 320–329. doi:10.1016/j.tranpol.2012.09.005

Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine Learning, 13*, 135–143. doi:10.1007/bf00993106

Schofer, J. (1983). Routes to the future of urban public transit. *Urban Affairs Review, 19*, 149–166. doi:10.1177/004208168301900202

Scott, L. et al. (2016). Bayes and big data: the consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management, 11*(2), 78-88. doi:10.1080/17509653.2016.1142191

Shepherd, J. (2004). What is the Digital Era? In G. M. Doukidis (Ed.), *Social and Economic Transformation in the Digital Era* (pp. 1-18). IGI Global. doi:10.4018/978-1-59140-158-2.ch001

Shmilovici, A. (2006). Support vector machines. In R. L. Maimon O. (Ed.), *Data Mining and Knowledge Discovery Handbook* (pp. 257–276). Boston, MA, USA: Springer. doi:10.1007/0-387-25465-X_12

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science, 25*, 289–310. doi:10.1214/10-STS330

Shoji, K. (2001). Lessons from Japanese Experiences of Roles of Public and Private Sectors in Urban Transport. *Japan Railway & Transport Review, 29*, 12–18. Retrieved from http://www.ejrcf.or.jp/jrtr/jrtr29/pdf/f12_sho.pdf

Simpson, M. (2009). An Integrated Approach to Assess the Impacts of Tourism on Community Development and Sustainable Livelihoods. *Community Development Journal, 44*, 186–208. doi:10.1093/cdj/bsm048

Smith, P. and G. Ruxton. (2020). Effective use of the McNemar test. *Behavioral Ecology and Sociobiology, 74*, 1–9. doi:10.1007/s00265-020-02916-y

Sohrabi, B. et al. (2019). A predictive analytics of physicians prescription and pharmacies sales correlation using data mining. *International Journal of Pharmaceutical and Healthcare Marketing, 13*, 346–363. doi:10.1108/ijphm-11-2017-0066

Sohrabi, B. et al. (2020). A predictive model of tourist destinations based on tourists' comments and interests using text analytics. *Tourism Management Perspectives, 35*, 100710. doi:10.1016/j.tmp.2020.100710

Statistics Bureau. (2016). *Ministry of Internal Affairs and Communications*. Retrieved from Statistics Bureau of Japan: http://www.stat.go.jp/english/data/index.html

Statistics Dashboard. (2020). *Ministry of Internal Affairs and Communications*. Retrieved from Statistics Bureau of Japan: https://dashboard.e-stat.go.jp/en/

Statistics Japan. (2021). *Automobiles Registered, Hokkaido.* Retrieved from https://stats-japan.com/t/tdfk/hokkaido

Statresearch. (2020). *Basic information of Kitami Institute of Technology*. Retrieved from http://statresearch.jp/school/university/university_0120.html

Steorts, R. C. (2016). Introduction to Bayesian Statistics, Part I. In *Bayesian Methods and Modern Statistics.* Durham, NC: Duke University. Retrieved from http://www2.stat.duke.edu/~rcs46/bayes.html

Stephenson, M. (2011). Conceiving land grant university community engagement as adaptive leadership. *Higher Education, 61*, 95–108. doi:10.1007/s10734-010-9328-4

Stewart, D. (1999). Twenty-Five Most Common Mistakes with Real-Time Software Development. *Embedded Systems Conference.* San Jose, CA.

Sullivan, M. and R, Feinn. (2012). Using effect size—Or why the P value is not enough. *Journal of graduate medical education., 4*, 279–282. doi:10.4300/jgme-d-12-00156.1

Taherdoost, H. (2016). Sampling methods in research methodology; How to choose a sampling technique for research. *SSRN Electronic Journal*. doi:10.2139/ssrn.3205035

Takamatsu, R. et al. (2021). Moralization of Japanese cultural norms among student sojourners in Japan. *International Journal of Intercultural Relations, 80*, 242–249. doi:10.1016/j.ijintrel.2020.12.001

Takumi, T. (2017). Improving sustainability in rural communities through structural transitions, including ICT initiatives. *JRI Research Journal, 1*, 44-66.

Tang, R. (2020). Does trade facilitation promote the efficiency of inbound tourism?—The empirical test based on Japan. *International Journal of Tourism Research, 23*, 39–55. doi:10.1002/jtr.2390

Toprak, M. (2018). Functional transformation of universities: an assessment on literature. *Journal of Higher Education and Science, 8*, 227. doi:10.5961/jhes.2018.265

Tran, Y. et al. (2020). Attitude toward physical activity as a determinant of bus use intention: A case study in Asuke, Japan. *IATSS Research., 44*, 293–299. doi:10.1016/j.iatssr.2020.03.002

Trommershäuser, J. et al. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12*(8), 291-297. doi:10.1016/j.tics.2008.04.010

Tsuyoshi, A. (2017). ICT-driven Regional Revitalization—seichi junrei by the numbers. *New Breeze, 29*, 9–13.

Ugoni, A. and B. Walker. (1995). The Chi square test: an introduction. *COMSIG Rev., 4*, 61–4.

Ukpabi, D. and H, Karjaluoto. (2017). Consumers' acceptance of information and communications technology in tourism: A review. *Telematics Informatics, 34*, 618–644. doi:10.1016/j.tele.2016.12.002

United Nations. (2020). *Department of Economic and Social Affairs*. Retrieved from Sustainable Development: https://sdgs.un.org/goals/goal11

UNWTO. (2016). *Tourism Highlights.* Madrid, Spain: World Tourism Organization. doi:10.18111/9789284418145

UNWTO. (2019). *International Tourism Highlights.* Madrid: World Tourism Organization. doi:10.18111/9789284421152

Walters, A. (1979). Costs and Scale of Bus Services. *World Bank Staff Working Paper*(325).

Weforum. (2015). *The Travel & Tourism Competitiveness Report 2015.* Geneva, Switzerland: World Economic Forum. Retrieved April 6, 2016, from https://www.weforum.org/reports/travel-and-tourism-competitiveness-report-2015

Weforum. (2017). *The Travel & Tourism Competitiveness Report 2017.* Geneva, Switzerland: World Economic Forum. Retrieved April 6, 2017, from https://www.weforum.org/reports/the-travel-tourism-competitiveness-report-2017

Weforum. (2019). *The Travel & Tourism Competitiveness Report 2019.* Geneva, Switzerland: World Economic Forum. Retrieved April 6, 2020, from https://www.weforum.org/reports/the-travel-tourism-competitiveness-report-2019

Wei, C. et al. (2011). Assessment of health-promoting lifestyle profile in Japanese university students. *Environmental Health and Preventive Medicine., 17*, 222–227. doi:10.1007/s12199-011-0244-8

Wilson, S. et al. (2001). Factors for Success in Rural Tourism Development. *Journal of Travel Research, 40*, 132–138. doi:10.1177/004728750104000203

World Bank. (2021). *Gross domestic product 2020, PPP.* World Bank. Retrieved from https://databank.worldbank.org/data/download/GDP_PPP.pdf

WTTC. (2015). *Travel & Tourism Economic Impact Report, Japan.* London, UK: The World Travel & Tourism Council Publishing.

WTTC. (2017). *Benchmark Report - Japan.* London, UK: The World Travel & Tourism Council Publishing.

WTTC. (2018). *Domestic Tourism Importance And Economic Impact.* London, UK: The World Travel & Tourism Council Publishing.

WTTC. (2020). *Travel & Tourism Economic Impact Report, Japan.* London, UK: The World Travel & Tourism Council Publishing.

WTTC. (2021). *Japan, Annual Research: Key Highlights.* London, UK: WTTC Publishing.

Zambon, I. (2019). Exploring Student Mobility: University Flows and the Territorial Structure in Viterbo. *Urban Science, 3*(2), 47. doi:10.3390/urbansci3020047