# Factors Affecting Satisfaction in an Inter-Island Multiport System in Leyte-Cebu and Predicting Passenger's Port and Mode Choice using Decision Tree Algorithm

Hanzel N. MEJIA [1a], Phattarasuda WITCHAYAPHONG [2b], Felippe Niño C. LUSARES[1c], Ymber Cecilia P. MALANGUIS[1d], Glayden O. MERINO[1e]

[1] *Department of Civil Engineering, Visayas State University, Visca, Baybay City, 6521, Leyte, Philippines*
[2] *Center of Excellence in Innovation for Sustainable Infrastructure Development, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand*
[a] *E-mail: hanzel.mejia@vsu.edu.ph*
[b] *E-mail: phattarasuda@g.sut.ac.th*
[c] *E-mail: felippe.lusares@vsu.edu.ph*
[d] *E-mail: ymber.malanguis@vsu.edu.ph*
[e] *E-mail: glayden.merino@vsu.edu.ph*

**Abstract**: One of the most critical considerations in determining a country's economic development is its access to different inter-island travel. For the Philippine archipelago, inter-island travels directly affect its local economy and the development of the port's facilities and services are necessary. This study aims to determine how the travelling population from Leyte Island to Cebu City consider their travel mode and port of choice through characterizing the factors that affect their behavior and satisfaction, and to create a model that can predict the passenger's likelihood of choosing a particular port and/or travel mode. This study developed classification decision tree models using the revealed preferences of the passengers in the Island of Leyte. The modelling tool used was able to predict passenger behaviors with 90.00% fit. Significant variables are consistent with other studies using different models, particularly the more common Logit modeling.

*Keywords*: Mode Choice, Decision Tree Algorithm, Ports, Inter-Island Travel

## 1. INTRODUCTION

An advanced, safe, convenient, and sustainable public transport is one of the significant signs of urban modernization. Passengers always seek options to meet their satisfaction and comfort. Having more choices makes the passenger more satisfied. Moreover, competition of the available choices makes sure that passengers will be catered to quality of service. This study investigated the decision-making procedure and influencing factors on passengers' travel port of choice as well as their mode choice behavior to take measures to improve attractions and facilities concerning the transport and services for the passengers. Socio-demographic characteristics of the passengers were described and the factors that affect the passenger's choice pertaining to a particular port and sea transport travel modes in a multiport system were characterized. Models were created to predict the passenger's likelihood of choosing a particular port and/or travel mode. The recommendations for possible measures for facility improvements to the port management/authorities based on the gathered data were also included.

The Philippines is considered as an archipelago which ranked seventh in terms of number of islands in the world. The country has a total of 7,641 islands. Thus, air and sea travel are common and highly predominant. There are 3 major groups of islands in the country which are Luzon, Visayas, and Mindanao. Inter-island travel in the Visayas region is more common as compared to Luzon and Mindanao because it is composed of more islands. To be able to effectively transport goods and services between these islands, the government has been pursuing the development of several alternatives and facilities to satisfy and serve the people of the country. In recent years, it has been noted that these inter-island water transport systems have helped the intensification in road vehicle travel between major islands in the Philippines where construction of better facilities and production of good service is much needed. There are approximately 30 ports operating in the Visayas region and some of these ports are using different modes of water transport systems like fast craft, pump boat, and slow ferries to transfer passengers, goods, and services.

Leyte is a province in the Philippines located in the Eastern Visayas region, which has a population of 1,724,679 based on the latest census. It has a total of nine seaports located in Ormoc City, Tacloban City, Baybay City, Hilongos, Palompon, San Isidro, Bato, Liloan, and Maasin City. The ferry route between the Province of Cebu and the Province of Leyte is one of the busiest in the Philippines and among the major contributors of inter-island travel traffic. Ports of Ormoc, Baybay, and Hilongos are considered the top operating ports, located in Western Leyte, where most of the passengers travelling to Cebu City and other nearby islands embark. These ports cater to the travelling population of the Leyte Province, Southern Leyte, and some from the provinces of Samar and Biliran. According to Philippine Ports Authority (2019) in its report Philippine Ports Passenger Traffic, Ports of Ormoc, Baybay, and Hilongos has total inbound passengers of 665,547, 100,929, and 427,674, respectively.

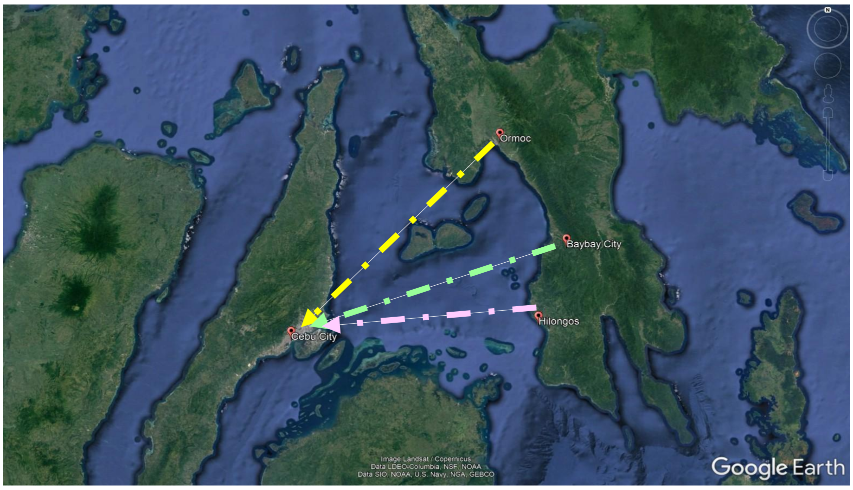

Figure 1. Major Leyte-Cebu Travel Routes (Map Source: Google Earth)

These major ports of Leyte are in the heart of each city which are approximately 45 kilometers from each other. Among these three mentioned ports, Ormoc and Hilongos can provide more alternative modes than the port of Baybay. Baybay has only one travel mode which is the slow ferry. These ports offer ferry and cargo services for Cebu city and vice versa. Due to its proximity to the City of Cebu, these ports cater numerous day and night trips to and from Cebu. Aside from that, Cebu City has big universities, recreational and tourist sites, offers more job and business opportunities, among others, which are the reasons for the increasing inter-regional travel demand.

Investigations regarding port and mode choice of passengers in an inter-island/ region sea travel have been looked at in the Philippines as seaport activity is prevalent in the country. To efficiently manage a transport network, it is important to understand how the travelling population makes their travel mode choices, just as much as the operating characteristics of the variable of interest itself.

The idea of this study is to facilitate future decision-making procedures of travelers from Leyte and nearby provinces and to suggest measures to improve choices' desirability. Findings could affect decisions by port managers as well as carriers or shippers. This also could suggest changes to ports or to use present facilities more efficiently.

The technological development in the transport industry, including the sea transport industry, aims at reinforcing transport facilities to the growing needs and requirements. Rodrigue et al. (2016) suggested that when a transport mode becomes more advantageous than another over the same route, a modal shift is likely to take place. In the same manner, when a port cannot satisfy the passenger, there is a greater possibility of shifting to another port. A related study by Diaz & Sorupia (2001) showed how different shipping service types, such as between slow ferry services and High-Speed Crafts (HSC) were chosen by the passengers. The study suggests that shorter travel time and a higher level of comfort offered by HSCs were worth the higher fares for those who shifted from other conventional sea transport modes.

Recent studies on ports and shipping lines used ranking of service and cost criteria to model traveler's choice or preference in the selection of the port and mode. One may find later in this paper the distinguishing parameter/factor for port and mode preference of a passenger that came out significant in the analyses.

This study also uses a relatively new methodology for modeling mode choice. In the past, multinomial logit models have been very common for these types of studies. A paper by Sekhar & Madhu (2016) compared the traditional multinomial logit model with decision tree model in modeling the mode choice in Delhi. Their study concludes that a decision tree model outperforms multinomial logit model by up to 21.00%. Our study will expand on this work and will use decision tree models to predict mode choice and port choice in a multi-port island system.

Studying the characteristics of an inter-island travelling population, particularly the Leyte-Cebu passengers of Western Leyte ports draws an effective understanding on what people want and how they make their travel mode choice. This is essential in the planning and execution of transport policies, which answers the need for an efficient, affordable, reliable, sustainable, and safe transport system.

## 2. LITERATURE REVIEW

The efficiency of a transport system affects travel behavior on mode choice determinations. Each country applies different methods for the study and analysis on travel behavior according to methodology, data collection, location characteristics, attribute variables and units of analysis. Basically, mode choice analysis is the third step in the conventional four-step transportation forecasting model and the logit model is widely used for transportation forecasting in various forms. Choice models have attracted a lot of attention and work. This section presents a view of papers to credit the early studies that configured and optimized the investigation regarding choice of port and mode of transport, especially in an inter-island-travel. Fundamentally, in the transportation planning process, it is important to understand how the travelling population makes their travel mode choices, just as much as the operating characteristics of the variable of interest itself. This is done by analyzing each relevant data and information as a basis for predicting future happenings.

In a study in the Philippines, Diaz & Sorupia (2001) looked at mode choice analysis and focused on the impact of the emergence of high-speed crafts on inter-island trip-making behavior. As a result, it was noted that a higher level of comfort offered by High-Speed Craft (HSC) was worth the higher fares for those who shifted from other conventional sea transport modes. Another study in Spain applied discrete multimodal choice models in an inter-island context, Ortúzar & González (2002) examines variations in travelers' mode choice for inter-island trips when journey conditions are changed. The route studied is the route between Gran Canaria and Tenerife, where three transport modes are available: plane, jetfoil, and ferry. It was found that passengers with higher income levels attach the greatest importance to travel time over the weight to price. A study in China, Tiwari et al. (2003) studied the port choice behavior of shippers, using a discrete choice model based on shipping line and port combinations, and made decisions based on the various shipper and port characteristics. The results evidenced that distances, port congestion, and shipping line's fleet size were affecting the mode choice and played an important role. Another study in the Philippines, Diaz (2011) studied mode choice of inter-island travelers by analyzing the willingness of ferry passengers to shift to air transportation using binary choice logit models to evaluate the impact of fare differences, trip characteristics and the socio-economic attributes of respondents on their mode choice and the logit modelling results confirm that fare levels, traveler income and trip purpose are the most significant determinants of inter-island modal shares. They also recognized that further analysis of more recent passenger movement and fares data and their application to the modal shift model should be undertaken to further improve the robustness of the model. Later, a similar study was conducted in the Philippines by Roquel & Fillone (2013) using the logit choice model based on revealed preferences. The use of logistic regression was also applied for mode choice analysis of inter-island passenger travel from Iloilo to Negros Occidental and resulted that the models developed followed the expected outcomes about the signs of the coefficient of the variables, taking time and cost spent as disabilities to the individual. Roquel & Fillone (2013) also noted that the deceptive effects of income and age of the individual can be considered acceptable. However, the model was found to be approximately 50% accurate only, which could still be improved. Statistically speaking, logistic regression analysis is used to examine the association of (categorical or continuous) independent variable (s) with one dichotomous dependent variable. This contrasts with linear regression analysis in which the dependent variable is a continuous variable. A study in Columbia, Vega et al. (2019) applying multinomial logit models to evaluate the port choice decision when competition between ports exists by the official records of imports and exports. It shows that the port access cost, the frequency of maritime lines, maritime freight rates, maritime travel time, origin or destination and the type of cargo, play a key role in the port selection process. A study in Thailand, Witchayaphong et al. (2020) conducted a revealed preference survey and determined the variables influencing traveler's mode choice behavior on mass transit in Bangkok. The result found that gender, age, average income, and auto ownership significantly affected the traveler's choice of mode. Due to the longer distance of the station, total travel time in public transport was not affected by the Thai traveler's mode choice.

Hsu et al. (2005) conducted a case study that demonstrates the application of the logistic regression model in Taiwan. The study incorporates intercity traveler's mode choice as well as route choices into the framework by exploring the relationship among key demand and supply variables. The results show that travelers with higher values of time tend to choose faster modes as their intercity modes. The Multinomial Logit (MNL) model has received more attention in terms of choice modeling and other applicable methods were slow to adopt these other methods these days. Though forecasting using the MNL model is popular, other techniques like decision trees can give positive outcomes and good predictive ability. Akiyama & Okushima (2004) confirmed that the travel mode choice model can be formulated with a complex relation of factors as the fuzzy decision tree approach and the pruning would be explained to summarize the information for classification. This study will bridge the gap by providing concrete results and models using the Decision Tree Regression method.

Generally, decision trees have shown to be effective in forecasting freight mode choice decisions. Samimi et al. (2012) applied a decision tree approach to analyze freight mode choice decisions by conducting surveys on domestic shipments in the USA. It was found that shipment weight turned is the most influential variable on freight mode choice decisions. Another study in the USA, Tang et al. (2015) applied the decision tree (DT) to explore the underlying rules of travelers' switching decisions between two modes under a proposed framework of dynamic mode searching and switching. This study also compared and evaluated the performance between DT and Logit models and the result showed that DT models can outperform logit models in many cases in both individual prediction levels and aggregate prediction levels. This was supported by a study in India, Sekhar & Madhu (2016) conducted a household interview survey and modelled the mode choice behavior of commuters by considering Random Forrest (RF) Decision Tree method. This study also did a comparative evaluation between the traditional MNL model and the Decision tree model to demonstrate the suitableness of RF models in mode choice modeling. The result supported that the model developed by the Random Forrest-based DT model is the superior one with higher prediction accuracy. Zhan et al., (2016) applied the hierarchical tree-based regression (HTBR) model to explore university student travel frequency and mode choice patterns in China, using the data collected by a web-based travel survey. It was found that student grade, school location, public transit station coverage ratio (PTSCR), and family income affects student travel frequency. Travel distance, bike ownership, school location, PTSCR, and gender were significantly correlated to student mode choice. A study in Indonesia, Safitri & Surjandari (2017) utilized a decision tree to predict travel mode switching by collecting a survey on demographic data of Transjakarta passengers and non-passengers. It was found that the new transportation alternative mode affects travel mode choice behavior of Jakarta Greater Area people. A study in Brazil, Pitombo et al. (2017) compared decision tree algorithms to estimate intercity trip distribution. The analysis was based on a dataset from the 2012 Origin-Destination Survey carried out in Bahia. The population of the origin city, GDP of the origin city and travel distances at an aggregated level, as well as the variables: age, occupation, level of education, income (monthly), number of cars per household and gender, are the variables that affect the destination choices. The key finding was that decision tree algorithms can be applied to improve traditional trip distribution techniques by assimilating the influence of disaggregated factors into distribution models. A study in Poland, Kotowska et al. (2020) developed a database containing the factors that are decisive for selecting a transport mode by cargo shippers, and the decision tree methodology was used in the analysis. It was shown that the major attributes in selecting transport modes by cargo shippers, considering access to three modes of transport to the seaport's hinterland, are consignment size and time pressure, then owning or having access to barge terminals by cargo shippers, and the annual volume of cargoes generated by them. According to Song & Ying (2015), the decision tree method classifies a population into branch-like segments that construct an inverted tree with a root node, internal nodes, and leaf nodes. The algorithm is non-parametric and can efficiently deal with large, complicated datasets without imposing a complicated parametric structure. When the sample size is large enough, study data can be divided into training and validation datasets. Loh (2014) states that regression trees can fit almost every kind of traditional statistical model, including least-squares, quantile, logistic, Poisson, and proportional hazards models, as well as models for longitudinal and multi-response data. Also, Strobl (2014) stated that decision tree analysis can provide variable importance measures that can be used to identify the most important predictor variables. Therefore, the references mentioned in this section supports the use of Decision Tree Algorithms in the analysis of traveler's choice for this study.

## 3. METHODOLOGY

The study aims to identify passenger characteristic that affects the decision making to use a particular travel mode or port of choice, this study also gathered information from different shipping offices and agencies such as ticket price, ticketing office schedule, travel time, ship capacity, ship departure time, the average number of trips, and average number of passengers. This information was essential for the design and collection of survey data. A revealed survey used in this study. The general procedure of the research process is presented in Figure 2.

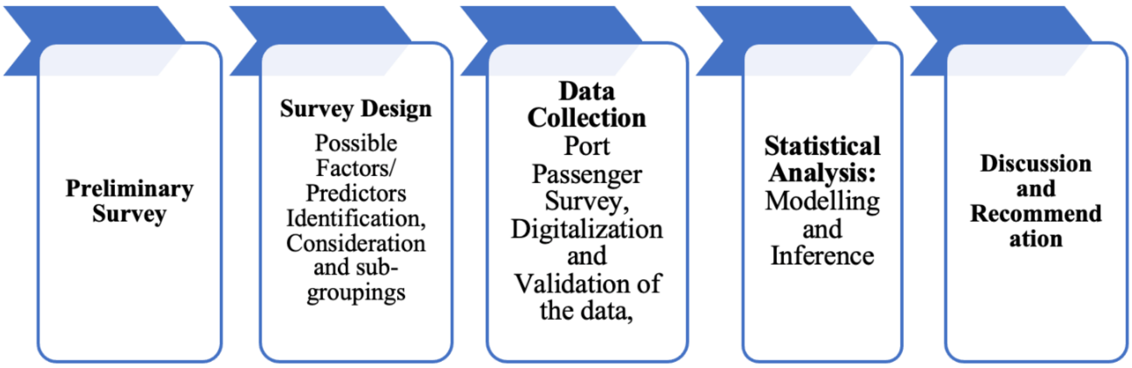


Figure 2. General procedure of the research process

A total of nine shipping agencies exists in the study area. Five of which are based in Port of Ormoc, namely the Lite Shipping Corporation, Supercat Fast Ferry Corporation, Ocean Fast Ferries Inc., Roble Shipping Inc. and the Asian Marine transport Corporation. On the other hand, Port of Hilongos has Roble Shipping Incorporated and Gabisan Shipping Lines, Inc. while Port of Baybay has Lapu Lapu Shipping Lines Incorporated and Roble Shipping Incorporated which only caters one mode (slow ferry).

### 3.1 Sampling Design

The Stratified Random Sampling (STRS) with stratifying variables per level was the sampling technique used. It is a probability sampling method which requires the division of the population $N$ into non-overlapping subpopulations or strata $Nh$, wherein each element of the population will belong in exactly one stratum. Then, a sample $nh$ is selected from each stratum using systematic random sampling, making sure that the selection of the samples from the different strata are independent of each other. In this case, the population is represented by the port passengers of Leyte Island travelling to Cebu City: taking ports of Ormoc, Baybay, and Hilongos as the substrata. Table 1 displays the data from Philippine Ports Authority (2019) pertaining the volume of passengers travelling from Ormoc, Baybay and Hilongos Ports to Cebu.

Table 1. Volume of Passengers Travelling through Western Leyte Ports

| Port of Access | Total number of Embarked Passenger | Daily Average Number of Embarked Passengers |
|---|---|---|
| Port of Ormoc | 665,547 | 1,824 |
| Port of Baybay | 100,929 | 277 |
| Port of Hilongos | 427,674 | 1,172 |
| Total | 1,194,150 | 3,271 |

Source: Passenger Statistics Summary 2019 Passenger Traffic Philippine Ports Authority

The overall sample $n$ consists of all the samples in the different strata (ports) which represents the population. Hence, a sample of size $n$ = 480 which was divided into three strata with sizes $n1$ = 160, $n2$ = 160, and $n3$ = 160 using the equal allocation method represented the $N$ = 3,271 daily average number of passengers of Western Leyte ports. Considering the data from Passenger Statistics Summary 2019, the total daily average number of passengers was used as the sampling frame. Moreover, Systematic Random Sampling (SRR) was used to determine the samples in each stratum. Systematic sampling is an extended implementation of probability sampling in which each member of the group is selected at regular periods to form a sample. In this case, a sampling interval of 5 was used repeatedly to choose subsequent samples.

### 3.2 Data collection

A questionnaire was used as a tool to collect data from randomly selected passengers. The design of the questionnaire was based on data that was acquired from the shipping offices, passengers' basic socio-demographic information and their travelling options and preferences, and related studies. The questionnaire included questions on trip characteristics such as trip purpose, trip origin and destination, transport mode, and access distance to and from ports. It also included travel cost and travel time that the traveler spent to complete the trip. The researchers also considered service characteristics such as time and cost differential and level of comfort and safety that can also be a factor or a variable that affects the passengers in their decision making. The approach used to gather the perception of the travelling population of Leyte Island is revealed preference survey. This approach is advantageous because it provides the real and actual choices made by travelling respondents in a determined context of constraints. It is used as a replication of the actual market share condition, which in this case is the share of ports and available modes, given that the data is collected on a representative sample.

### 3.3 Methods of Data Analysis

The survey collected was a wide range of day-by-day travel patterns, socioeconomic data, attitudes, and preferences from the travelling population. The modeling method was based on Decision Tree Algorithm, specifically the Classification Tree (CT) procedure. Regression trees are for dependent variables that take continuous or ordered discrete values, with prediction error typically measured by the squared difference between the observed and predicted values. Such a tree is constructed through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then splitting it up further on each of the branches. Figure 3 presents the general workflow of a Decision Tree used for classification purposes.

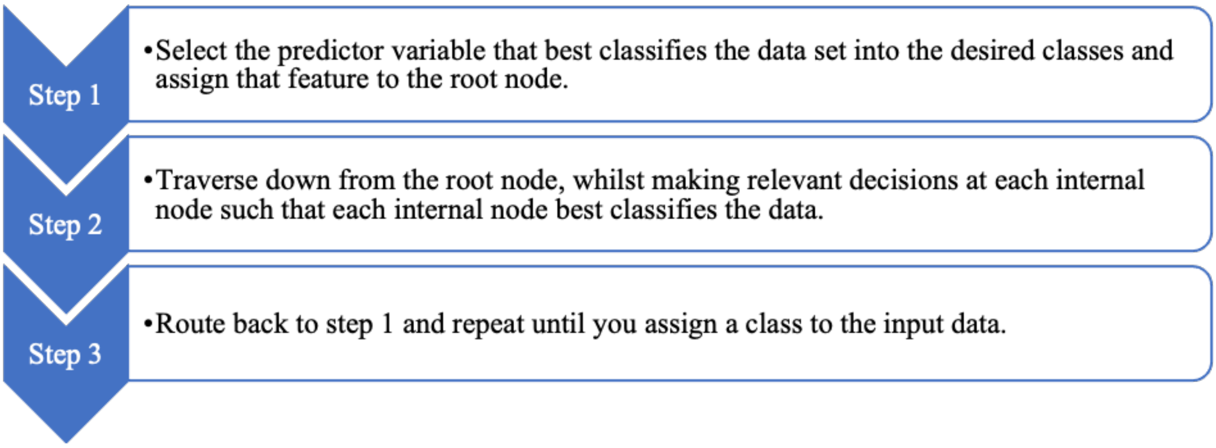


Figure 3. General Flow of a Decision Tree

In broad-spectrum, Decision Trees, specifically Classification Trees are used to predict membership of cases or objects into classes of a categorical dependent variable from their measurements on one or more predictor variables. Decision trees offer straightforward interpretation not depending so much on model adequacy and assumption.

## 4. DATA ANALYSIS

### 4.1 Descriptive Statistics

This section thoroughly describes the socio-demographic profile of the passengers of the Western Leyte ports, specifically the Ports of Ormoc City, Baybay City, and Hilongos, Leyte. The socio-demographic profile of the respondents is essential and necessary to be considered, identified, and described in this study for this will most likely affect the respondents' decision making, and influence their travel options. In total, there were 480 samples gathered for the study. 160 respondents were equally allocated in each port, however, only 159 respondents were considered in Baybay Port due to the incomplete response of a respondent passenger. Descriptive statistics of the passengers are summarized in Table 2.

Table 2. Descriptive Statistics

| | | Port (Percent) | | |
|---|---|---|---|---|
| | | Ormoc | Baybay | Hilongos |
| Gender | Male | 35.00% | 33.21% | 31.79% |
| | Female | 30.81% | 34.59% | 34.59% |
| Civil Status | Single | 35.00% | 33.21% | 31.79% |
| | Married | 30.81% | 34.59% | 34.59% |
| | Widowed | 35.71% | 14.29% | 50.00% |
| Income | Below PHP 10,000 | 25.08% | 28.57% | 46.35% |
| | PHP 10,000 - 19,999 | 51.65% | 32.97% | 15.38% |
| | PHP 20,000 - 29,999 | 75.86% | 24.14% | 0.00% |
| | PHP 30,000 - 39,999 | 75.00% | 25.00% | 0.00% |
| | above PHP 40,000 | 16.67% | 83.33% | 0.00% |

Other meaningful information about the trip were also included such as the purpose of the trip, land travel time, trip fare and schedule, and whether they travel alone. This information is presented in Figure 4.

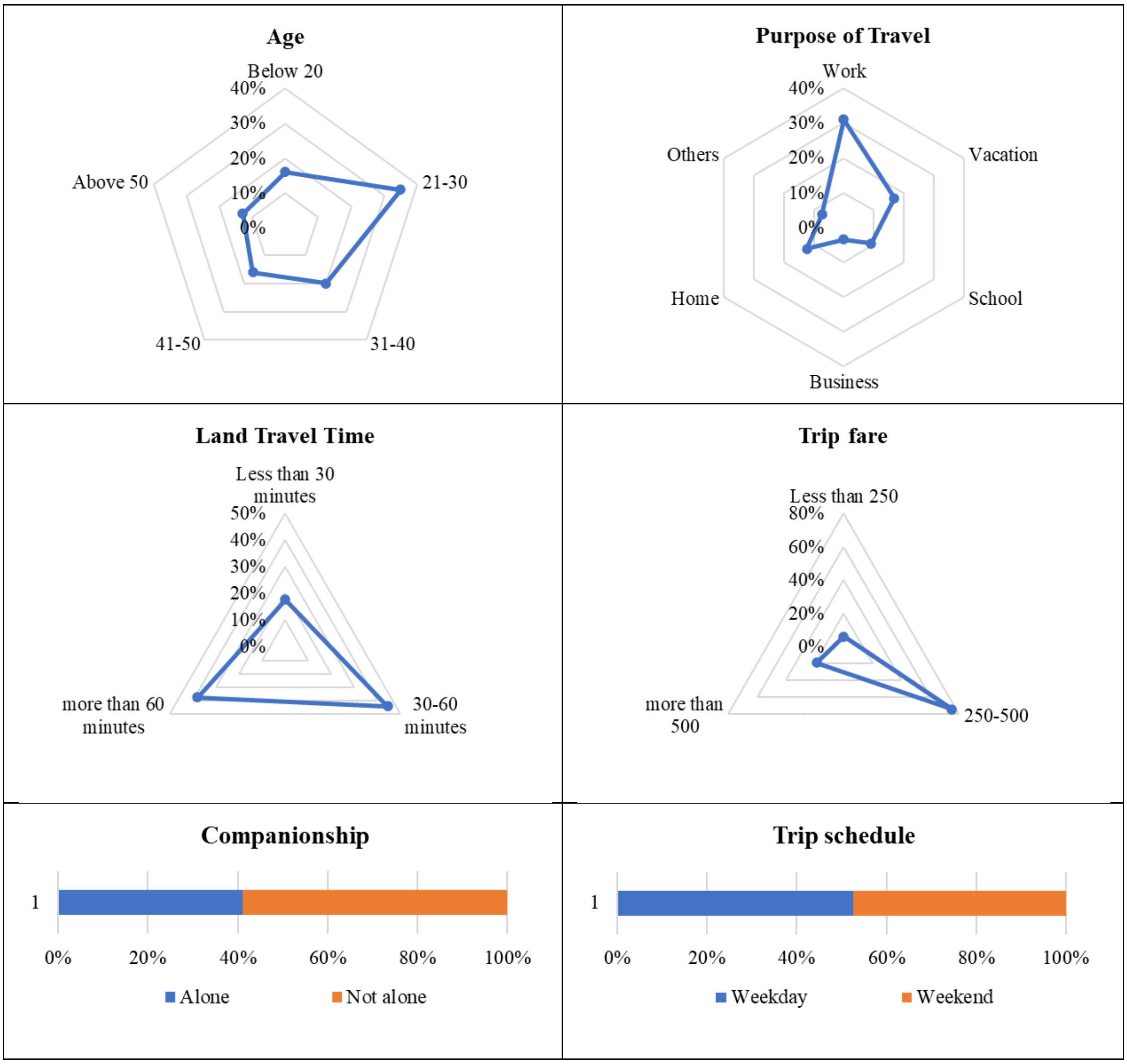


Figure 4. Sample Characteristics

### 4.2 Factors Affecting Satisfaction on Ports

The factors influencing the passengers' port of choice and transport mode were indicative. These influencing factors were based on the assessment of the passenger's satisfaction level of the port's amenities and services. These determining factors were revealed as the port's cleanliness, security, reliability, customer service, and accommodation. And since the study was able to identify these factors, articulated and suggested recommendations for the improvement of the management of the port can be done to be able to satisfy the needs of every passenger in each port. In Figure 5, most of the passengers were satisfied with Ormoc ports' quality of services, having at least 93.00% of the passengers satisfied in 5 parameters. It also shows that several passengers in Baybay port were dissatisfied in terms of accommodation, making up to 36.48% of the sample. This suggests that the Port of Baybay needs to provide better accommodation services. The port of Hilongos will also benefit by improving cleanliness and security, which have a high dissatisfaction rate of about 24.00% and 69.00%, respectively.

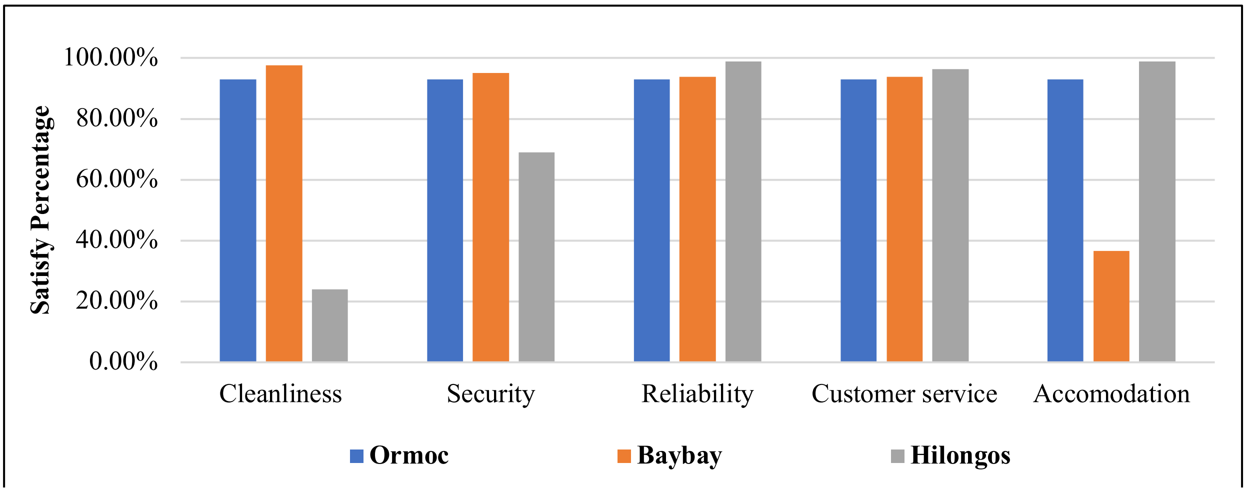

Figure 5. Passengers' satisfaction assessment

### 4.3 Classification Decision Tree Model

The Tree Analysis is the schematic representation of several decisions followed by different chances of occurrence. The nodes in a decision tree graph represent an event while the edges of the graph represent the decision rules or conditions in that graph. Classification Decision Tree provides a measure of confidence that the classification is correct. Since the responses (port of choice and transport mode) are categorical and opting for a model with straightforward interpretation and not depending so much on model adequacy, the classification decision tree can be used in the analysis approach for this study.

Table 3 shows the significant variables that are relevant in constructing the Decision Tree Analysis which are as follows: departure, trip, purpose, alone, fare, age, and mode. The second column defines each variable stated. The third column is the value to be used as codes in the analyses. An example is for departure schedule, the value is 0 if the boat's schedule for departure is in the daytime while the value is 1 if it travels at nighttime.

The predictors or variables that resulted significantly in the analysis were recognized and characterized as factors that affect the passenger's choice pertaining to what port and sea transport travel modes will be used in a multiport system; and these factors relate to time, cost, comfort, and purpose of the trip. According to Roquel & Fillone (2013), time-related factors may be contributed by the discomfort experienced during the travel or when travelling on land, e.g., traffic delays, noise, and pollution. They also stated that cost-related factors can be explained depending on the individual perception of his/her money's worth and with respect to time. Older people are less likely to be in a hurry, so time is understandably not as much a major factor as cost. The study also showed the significance of age in the mode selection of passengers. For the trip, this is much expected as one of the significant variables in this study as it can be seen in the distribution of respondents. Likewise, Diaz (2011) confirmed on their inter-island travel study that fare level and purpose of the trip are the most significant determinants of inter-island modal shares.

Table 3. List of significant variables used in modelling

| Variable | Definition | Value/Category | Significance |
|---|---|---|---|
| Departure | Departure Schedule | 0 = day, 1 = night | Port Choice, Mode Choice |
| Trip | Schedule of Trip | 0 = weekend, 1 = weekday | Port Choice |
| Purpose | Purpose of Trip | 0 = work, 1 = vacation, 2 = school, 3 = business, 4 = home, 5 = others | Mode Choice |
| Alone | Solo Travel | 0 = no, 1 = yes | Mode Choice |
| Fare (PHP) | Fare for the Trip | 0 = less than 250, 1 = 250 - 500, 2 = more than 500 | Port Choice, Mode Choice |
| Age | Age of Passenger | 11 – 78 years old | Mode Choice |
| Mode | Mode of Transport | 0 = fast craft, 1 = slow ferry | Port Choice |

### 4.3.1 Model 1: Passenger's Port of Choice

Figure 6 shows the classification tree for port choice having 15 nodes connected with each other. For Node 1, the place Ormoc inside the node means that it has the highest probability from where passengers depart. The probabilities were shown just below Ormoc where the 3 values of 33.00% representing the probability for passenger's origin of travel, which is Ormoc, Baybay, and Hilongos respectively. It splits into two considering variable Departure 1 which is nighttime, having 39.00% of the passengers choosing the Port of Ormoc while the 61.00% choose another port like Baybay if they travel during daytime. As shown in Node 2 among the 39.00% of the passengers from Ormoc who wants to travel at nighttime, 69.00% of them will choose to travel from Ormoc and the remaining 31.00% to the other ports like Hilongos. Considering the variable Trip 1 which is a weekday, 27.00% among the 39.00% of the passengers will continue choosing their origin at Ormoc Port to Cebu and only 11.00% chooses to travel Hilongos - Cebu route if it's not during the weekdays. From Node 3 has 61.00% of the passengers who choose to travel any time other than nighttime, there's a probability that 11.00% would still choose to travel from Ormoc port, 54.00% will choose Baybay Port and 35.00% probability will choose Hilongos Port. Starting at Node 3, considering the travel mode, which is a fast craft, 14.00% of the passengers choose to travel by means of a fast craft while 47.00% of them choose slow ferry. At Node 6 among the 14.00% of passengers who choose to travel with fast craft, 48.00% of them will prefer the Port of Ormoc while 52.00% of the passengers who want to travel with a fast craft prefer the Port of Hilongos. At Node 7 among the 47.00% of passengers who want to travel with a slow ferry, there's the probability that 71.00% of them will choose the Port of Baybay while the remaining 29.00% will choose the Port of Hilongos. From Node 6 which has 14.00% of the passengers who choose fast craft as a mode, considering the fare of more than PHP 500, 7.00% of them will choose to travel from the Ormoc Port while the remaining half will choose to travel from Hilongos Port. The results from the Classification Tree show that 34.00% of the total passengers will choose the Port of Ormoc, 40.00% will choose the Port of Baybay and the remaining 25.00% will choose to travel from the Port of Hilongos.

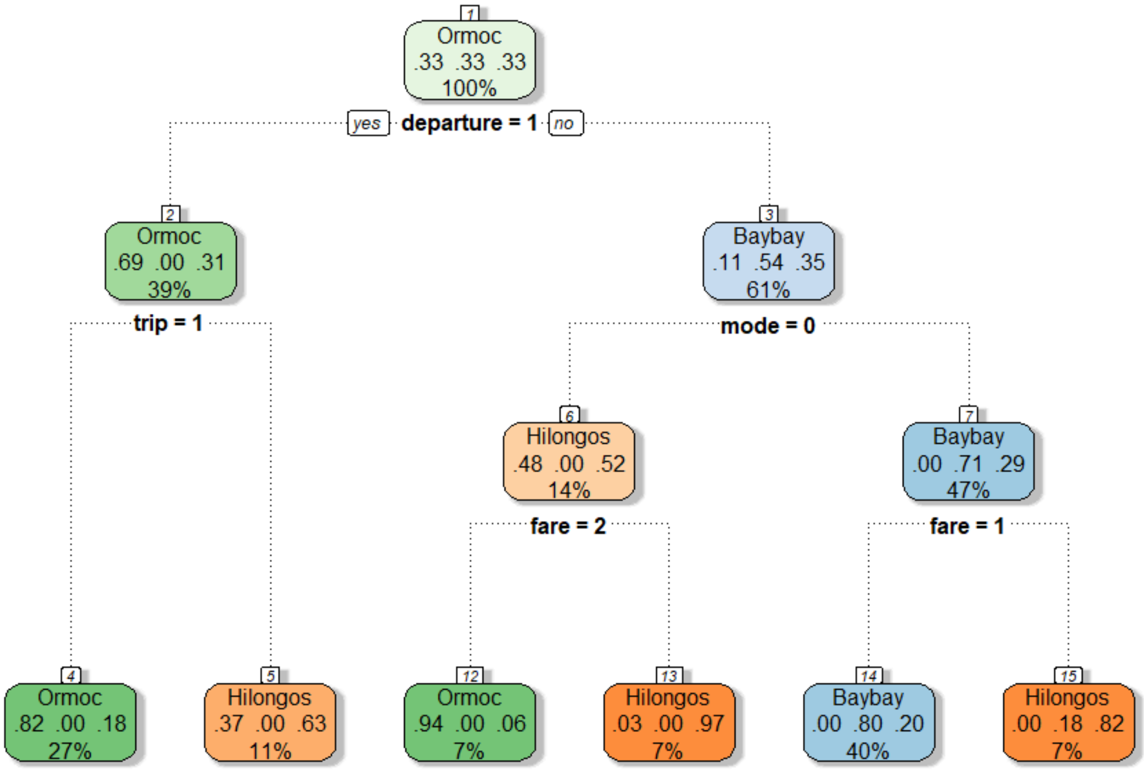


Figure 6. Classification Tree of the Passenger's Port of Choice

Table 4. Assessment of the Obtained Classification Tree in Model 1

| Split Sequence | Standard Error | Misclassification (%) | Prediction (%) |
|---|---|---|---|
| 0 | 1 | 33.19 | 66.81 |
| 1 | 0.6050 | 20.08 | 79.92 |
| 2 | 0.4922 | 16.34 | 83.66 |
| 3 | 0.3981 | 13.21 | 86.79 |
| 4 | 0.3323 | 11.03 | 88.97 |
| 5 | 0.2884 | 9.57 | 90.43 |
| Root Node Error = 0.3319 | | | |

Table 4 shows the different values of standard errors, misclassifications rate, and prediction accuracy that happened in each sequence in Model 1. The standard error is a measure of the statistical accuracy of an estimate, equal to the standard deviation of the theoretical distribution of a large population of such estimates. The misclassification rate refers to the percentage of training (recall) and testing (generalization) examples misclassified from a given data set. Prediction accuracy is a percentage to assess if results shown in the model with the given predictions are accurate enough for data analysis. The root node error is the percent of correctly sorted records at the first (root) splitting node. As seen on the table, the model has the largest standard error of 1, then it lowers down from its first to fifth split-sequence with values of 0.6050, 0.4922, 0.3981, 0.3323, and down to the lowest value of 0.2884. The only significant values for the misclassification rate and prediction accuracy lies on the last split-sequence. Therefore, the classification model has a misclassification rate of approximately 10.00%, the prediction accuracy of approximately 90.00%, and a root node error of 0.3319.

### 4.3.2 Model 2: Passenger's Transport Mode in Ormoc City Port

Figure 7 shows the classification tree of the passenger's transport mode in Ormoc City port with 7 nodes connected with each other having 3 different significant variables each to be considered in every split-sequence. Each percentage split can be seen at each node the same as the previous model. The final results show that 79.00% of the total passengers will choose slow ferry and the remaining 21.00% for fast craft as their mode of transport in the Port of Ormoc City.

Table 5 shows the different values of standard errors, misclassifications rate, and prediction accuracy that happened in each sequence in Model 2. As seen in the table, the model has the largest standard error of 1, then it lowers down from its first to third split-sequence with values of 0.6522, 0.4988, and down to its lowest value of 0.4613. The only significant values for the misclassification rate and prediction accuracy lies on the last split-sequence. Therefore, the classification model has a misclassification rate of approximately 10.00%, prediction accuracy of approximately 90.00%, and a root node error of 0.2125.

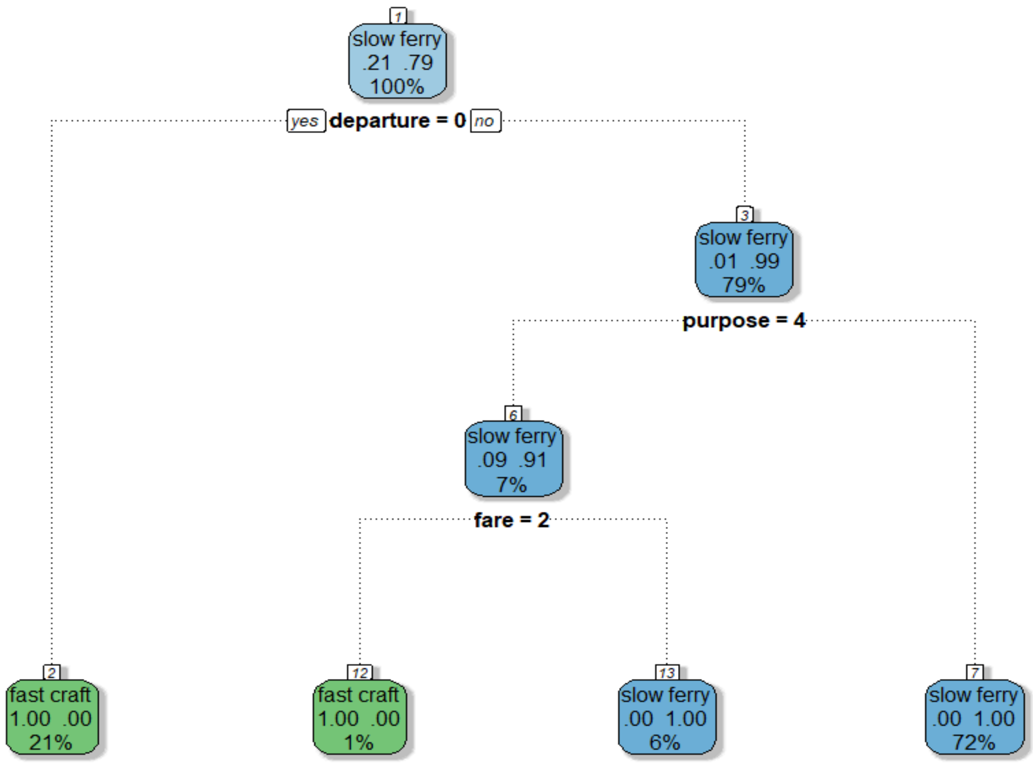


Figure 7. Classification Tree of the Passenger's Transport Mode in Port of Ormoc

Table 5. Assessment of the Obtained Classification Tree in Model 2

| Split Sequence | Standard Error | Misclassification (%) | Prediction (%) |
|---|---|---|---|
| 0 | 1.0000 | 21.25 | 78.75 |
| 1 | 0.6522 | 13.86 | 86.14 |
| 2 | 0.4988 | 10.60 | 89.40 |
| 3 | 0.4613 | 9.80 | 90.20 |
| Root Node Error = 0.2125 | | | |

### 4.3.3 Model 3: Passenger's Transport Mode in Hilongos Port

Figure 8 shows the model for classification tree for transportation mode in Hilongos port having 11 nodes with 5 significant variables or conditions each one needed to be considered for deciding a split sequence. From Node 1, among the 100.00% passengers in Hilongos port, only 23.00% of them prefer to travel using a fast craft while 77.00% of them choose to travel by means of a slow ferry. From Node 2, it shows that among the 55.00% passengers, a probability of 39.00% will want to travel with a fast craft while the 61.00% will continue choosing for the slow ferry. In summary, it can be seen in the model that only 22.00% of the passengers in Hilongos Port will choose to travel in a fast craft while 78.00% of the passengers will mostly prefer travelling in a slow ferry.

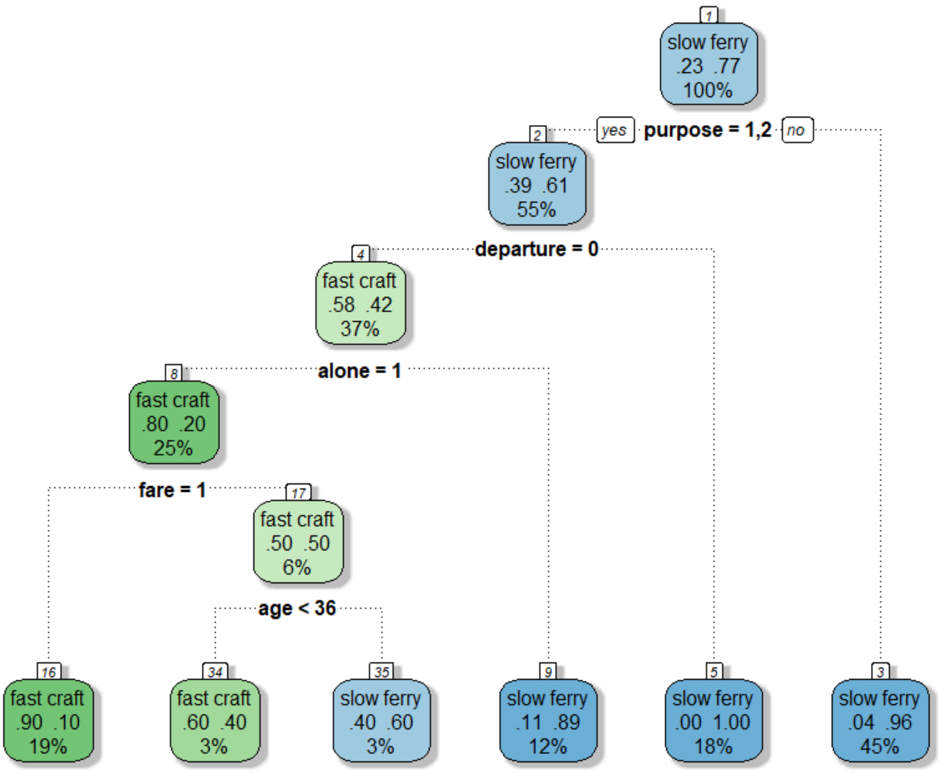


Figure 8. Classification Tree of the Passenger's Transport Mode in Port of Hilongos

Table 6. Assessment of the Obtained Classification Tree in Model 3

| Split Sequence | Standard Error | Misclassification (%) | Prediction (%) |
|---|---|---|---|
| 0 | 1.0000 | 23.27 | 76.73 |
| 1 | 0.7002 | 16.29 | 83.71 |
| 2 | 0.6659 | 15.50 | 84.50 |
| 3 | 0.5921 | 13.78 | 86.22 |
| 4 | 0.4834 | 11.25 | 88.75 |
| 5 | 0.4443 | 10.34 | 89.66 |
| Root Node Error = 0.2327 | | | |

Table 6 shows the different values of standard errors, misclassifications rate, and prediction accuracy that happened in each sequence in Model 3. As seen in Table 6, the model has the largest standard error of 1, then it lowers down from its first to fifth split-sequence with values of 0.7002, 0.6659, 0.5921, 0.4834, and down to its lowest value of 0.4443. The only significant values for the misclassification rate and prediction accuracy lies on the last split-sequence. Therefore, the classification model has a misclassification rate of approximately 11.00%, the prediction accuracy of approximately 89.00%, and a root node error of 0.2327.

## 5. CONCLUSION

This study presented passenger socio-economic data and assessment of ports based on satisfaction of passengers which can be helpful in other modeling studies, port planning and improvement, and decision making by the proper authorities. Traveler's choice was modeled particularly their mode and port choice in a multi-port inter-island travel using decision tree algorithm. The significant variables the passengers take into consideration for their port choice during travel are trip (schedule of the trip) and mode (mode of transport). For their mode choice, factors such as purpose (purpose of the trip), alone (solo travel), and age (age of passenger) were considered. The remaining factors such as departure (departure schedule) and fare (fare for the trip) were taken into considerations for both port and mode choice. These variables were used for the classification decision tree analysis. Said variables also shows consistency with previous studies particularly on inter-island travel.

The models developed provided good fit results when the different factors were considered simultaneously in the passenger's mode choice. For model 1, which is the passenger's port choice, the results show that most of the passengers will prefer Baybay Port, next is Ormoc Port and their last choice would be Hilongos Port. Model 2 and 3 consider the passenger's mode choice for Ormoc Port and Hilongos Port. Most of the passengers have chosen slow ferries for their travel while the remaining passengers prefer to travel with a fast craft. The three models in the classification decision tree show a consistent correctly classified instance of about 90.00%. This shows that decision tree models can yield a good fit for both mode and port choice. Further study must be done to test these models and validate it with another set of data to show its prediction strength.

## REFERENCES


1. Akiyama, T., & Okushima, M. (2004). Commuter modal choice model with using fuzzy decision tree. Proceedings of the SCIS & ISIS,
2. Diaz, C., & Sorupia, E. (2001). A study on inter-island passenger transport mode switching in the Philippines. J East Asia Soc Transp Stud, 4(1), 291-301.
3. Diaz, C. E. D. (2011). Mode choice of inter-island travellers: Analyzing the willingness of ferry passengers to shift to air transportation. Journal of the Eastern Asia Society for Transportation Studies, 9, 2058-2073.
4. Hsu, C.-I., Chen, Y.-C., & Li, H.-C. (2005). A model on market share distribution between air transportation, high-speed rail, and automobiles. Journal of the Eastern Asia Society for Transportation Studies, 6, 2003-2018.
5. Kotowska, I., Mankowska, M., & Plucinski, M. (2020). The decision tree approach for the choice of freight transport mode: the shippers' perspective in terms of seaport hinterland connections.
6. Loh, W. Y. (2014). Fifty years of classification and regression trees. International Statistical Review, 82(3), 329-348.
7. Ortúzar, J. D. D., & González, R. M. (2002). Inter-Island Travel Demand Response with Discrete Choice Models: Functional Form, Forecasts, and Elasticities. Journal of Transport Economics and Policy (JTEP), 36(1), 115-138.
8. Philippine Ports Authority. (2019). Summary Port Statistics Philippine Ports Authority 2019. R. o. t. Philipines. https://www.ppa.com.ph/?q=content/statistics-1
9. Pitombo, C. S., de Souza, A. D., & Lindner, A. (2017). Comparing decision tree algorithms to estimate intercity trip distribution. Transportation Research Part C: Emerging Technologies, 77, 16-32.
10. Rodrigue, J.-P., Comtois, C., & Slack, B. (2016). The geography of transport systems. Routledge.
11. Roquel, K. I. D., & Fillone, A. (2013). Mode Choice Analysis of Inter-Island Passenger Travel from Iloilo to Negros Occidental, Philippines. Journal of the Eastern Asia Society for Transportation Studies, 10, 586-599.
12. Safitri, D. M., & Surjandari, I. (2017). Travel mode switching prediction using decision tree in Jakarta greater area. 2017 International Conference on Information Technology Systems and Innovation (ICITSI)
13. Samimi, A., Razi-Ardakani, H., & Mohammadian, K. A. (2012). A Decision Tree Approach to Analyze Freight Mode Choice Decisions. Proceedings of the 1st European Symposium on Quantitative Methods in Transportation Systems. Available at: https://transp-or. epfl. ch/heart/2012/latsis2012_submission_109. pdf,
14. Sekhar, C. R., & Madhu, E. (2016). Mode choice analysis using random forrest decision trees. Transportation Research Procedia, 17, 644-652.
15. Song, Y.-Y., & Ying, L. (2015). Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), 130.
16. Strobl, C. (2014). Discussion on fifty years of classification and regression trees. International Statistical Review, 82(3), 349-352.
17. Tang, L., Xiong, C., & Zhang, L. (2015). Decision tree method for modeling travel mode switching in a dynamic behavioral process. Transportation Planning and Technology, 38(8), 833-850.
18. Tiwari, P., Itoh, H., & Doi, M. (2003). Shippers' port and carrier selection behaviour in China: a discrete choice analysis. Maritime Economics & Logistics, 5(1), 23-39.
19. Vega, L., Cantillo, V., & Arellana, J. (2019). Assessing the impact of major infrastructure projects on port choice decision: The Colombian case. Transportation Research Part A: Policy and Practice, 120, 132-148.
20. Witchayaphong, P., Pravinvongvuth, S., Kanitpong, K., Sano, K., & Horpibulsuk, S. (2020). Influential Factors Affecting Travelers' Mode Choice Behavior on Mass Transit in Bangkok, Thailand. Sustainability, 12(22), 9522.
21. Zhan, G., Yan, X., Zhu, S., & Wang, Y. (2016). Using hierarchical tree-based regression model to examine university student travel frequency and mode choice patterns in China. Transport Policy, 45, 55-65.