PERCEPTION OF PEDESTRIANS TOWARDS THE USE OF FOOTBRIDGE

PROJECT REPORT

Submitted by

ANASWARA R

REG NO: TKM21CETE01

to

the APJ Abdul kalam Technological University

in partial fulfillment of the requirements for the award of degree

of

Master of Technology

In

Transportation Engineering

DEPARTMENT OF CIVIL ENGINEERING

TKM College of Engineering, Kollam

MAY 2023

DECLARATION

I undersigned hereby declare that the project report "Perception of Pedestrians Towards the Use of Footbridge", submitted for partial fulfillment of the requirements for the award of degree of Master of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Prof. Karthik S. This submission represents my idea in my own words and where ideas or words of others have been included; I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by institute and/or University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University

Place: Kollam ANASWARA R

Date: 10/05/2023

DEPARTMENT OF CIVIL ENGINEERING

TKM College of Engineering, Kollam

CERTIFICATE

Certified that this report entitled '**PERCEPTION OF PEDESTRIANS TOWARDS THE USE OF FOOTBRIDGE'** is the report of project presented by **ANASWARA R, Reg No: TKM21CETE01** during **2022-2023** in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Transportation Engineering of the A P J Abdul Kalam Technological University.

Guide:	Project coordinator:	Head of the Department:
Prof. Karthik S	**Dr. Adarsh S**	**Dr. Sajeeb R**
Assistant Professor	Professor	Professor
Dept. of Civil Engg.	Dept. of Civil Engg.	Dept. of Civil Engg.
TKMCE, Kollam	TKMCE, Kollam	TKMCE, Kollam

ACKNOWLEDGEMENT

First of all, I thank almighty for giving me strength and courage to do the project. I extend my sincere gratitude to my respected principal **Dr. T.A Shahul Hameed** for his support towards the successful accomplishment of this course.

I avail this opportunity to express my gratitude to my guide **Prof. Karthik S**, Assistant professor, TKM College of Engineering for his inspiring assistance, encouragement and useful guidance.

I'm grateful to **Prof. Anandu V.G**, Assistant Professor, TKM College of Engineering and **Prof. Nisha Mary Dcruz** Assistant Professor, TKM College of Engineering, for serving as my evaluators for the project.

I also thank **Dr. Sajeeb R**, Head of the Department of Civil Engineering, TKM College of Engineering, Kollam.

I also thank my coordinator **Dr. Adarsh S**, Professor for his timely guidance.

Finally, I would like to express my sincere thanks to my parents, family and friends for their good will and constructive ideas which helped me a lot in bringing out this report.

ANASWARA R

ABSTRACT

Pedestrians are the most vulnerable group of road users. In order to protect the pedestrians from fatalities while crossing through at grade crosswalks, grade separated crossing facilities are provided. But they are sparsely used by the pedestrians. Using zebra crossing has become part of their habitual action. Pedestrians are not willing to use footbridge even though they are safe. The study aims to identify factors influencing the non-usage of footbridge using perception survey. Machine learning techniques was used to analyze the data. Different algorithms such as random forest, decision tree, support vector machine, k-nearest neighbor and logistic regression were compared to get best model. Random forest outperformed all other model with an accuracy of 83%. The weightage of factors was computed using random forest. It was found that frequency of using footbridge has a high weightage among other factors. It was followed by easiness, stressful, weather, tiresome, night, bad infrastructure, heavy traffic, height fear, unfamiliar location, accessibility, hurry, good illumination, education, age, pedestrian accident history and occupation. Some counter measures such as escalator, alerting poster, fine for illegal crossing, seating arrangement, security staff are suggested in the study to identify the change in usage of footbridge among non-users. Most of the pedestrians are willing to use footbridge if escalator or elevator is provided. To promote footbridge utilization, a user-centered strategy is necessary.

Keywords: ***Machine learning techniques, Footbridge***

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

FN	False Negative
FP	False Positive
LOS	Level of Service
SPSS	Statistical Package for the Social Sciences
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

CHAPTER 1
INTRODUCTION

1.1 General

A pedestrian must cross the roadway as part of their daily activities and utilize facilities like zebra crossings, footbridges, and underpasses. If alternate crossing options are offered, pedestrians must choose which options to use based on convenience, safety, comfort, and other factors. The significance of walking has diminished over time, initially with the advent of public transportation and then with the increase in popularity of private automobiles. People who cannot drive or use public transportation frequently are the ones who have to cross the street frequently. Pedestrians are far more at risk from vehicles while using at grade crosswalks. Examples of pedestrian road crossings include, at-grade crossings mainly through zebra crossing and grade-separated crossings such as footbridge and underpass. At grade crossing facilities are responsible for a considerable amountof traffic accidents. Pedestrians are the most vulnerable group of road users. The majority of pedestrian collisions take place as the pedestrian is attempting to cross the street. Pedestrians are competing with vehicles to use the same narrow space for crossing. One of the majorfactors contributing to traffic congestion is the unchecked flow of pedestrians across the street. However, there aren't any sizable grade-separated pedestrian crossing facilities to meet demandfrom pedestrians. The grade-separated pedestrian facilities that are now available, including a footbridge and an underpass, have a constrained LOS. Due to this pedestrian are forced to use at grade crossing facility mainly illegal crossing by stopping the vehicles by hand and directly expose them into collisions. Pedestrian accidents constitute a significant share of total accident victims in India. Even though they continue to cross the road in an unsafe manner. In order to tackle down this situations grade separated crossing facilities are provided but these are sparsely used by pedestrians. The need for pedestrian crossings must be satisfied by a grade-separated crossing facility to avoid such consequences. Additionally, pedestrian overpasses and underpasses are uncomfortable for most of the users and inaccessible. The footbridge is sparsely used by pedestrians as compared to underpass because most of them feels difficulty in climbing up the stairs.

Footbridges are essential for improving pedestrian safety and easing traffic in crowded urban areas. They are particularly useful for giving pedestrians a safe and efficient way to cross busy streets. Preparations must be made for pedestrians safe and efficient movement, especially when crossing the street, in order to limit the incidence of pedestrian fatalities. Footbridges are expensive, particularly when long ramps or lifts are needed for wheelchair users. People with mobility issues won't be able to use the footbridge without lifts or ramps. Pedestrians who feel tired while ascending stairs, may preferred to cross a busy street rather than using a footbridge. For the efficient usage of footbridge, it should satisfy following criteria:

- Footbridge should be user-friendly
- Entry and exit should be accessible
- It should be free from beggars or traders
- The height of the stairway riser should be ideally proportionate
- Footbridge should be aesthetically beautiful and well maintained
- Public education and awareness campaigns for the use of should be organized
- Improvement in terms of their access, condition, and enforcement

To improve the road crossing system, it is important to understand the perception of pedestrians regarding the road crossing facility. The importance of various problems encountered in other traffic crossing systems must also be emphasized. It is necessary to assess the performance of footbridge and the factors depromoting the users.

1.2 Objectives of the Study

- To identify the factors related to the usability of pedestrian footbridge using perception survey
- To formulate a binary classification model of preferred crossing facility using machine learning methods
- To compare the performance of various machine learning algorithms in classifying the crossing preference using the performance metrics
- To identify which of the interventions will change the willingness to use footbridge to large extent

1.3 Gaps Identified

- Environmental factors such as rain, extreme heat does not consider
- Only few studies have been conducted regarding the perception of footbridge usage in India
- Only limited studies have been done regarding the footbridge usability using machine learning techniques, especially the classification model of preferred crossing facility of pedestrians

1.4 Organization of Report

The report is organized into 5 chapters. Chapter 1 gives a brief idea regarding the thesis work and the objectives of the work. Chapter 2 deals with literature review of topic. A detailed review was done on the topic to identify the gaps associated with it. Chapter 3 provide a detailed explanation of the methodology executed in the work. A brief idea of machine learning techniques was explained in this chapter. Chapter 4 deals with results and discussion of the work. Improvements were also discussed in this section. Chapter 5 summarize the work and also mentioned the future scope of the study.

CHAPTER 2

LITERATURE REVIEW

2.1 General

Most of the pedestrians are not willing to use footbridge as they thinking that it may consume more time. The present study focusses on the factors regarding the non-usage of foot bridge. This chapter deals with the literature review of topic.

2.2 Factors influencing the usage of footbridge

Wu et al., (2014) conducted a study in China to identify the factors influencing footbridge usage. Total 873 data was collected from footbridge users and 258 data from footbridge non-users. Binary logit model was used for analysis and the accuracy obtained from the model is 81.9%. The factors influencing the usage of footbridge are age, gender, career, education level, license, detour wishes, detour distance, and crossing time.

Oviedo and Scott (2017) conducted a study on a high-traffic highway in Barranquilla, Colombia, to identify the factors which are influencing the usage of footbridge. The interviewer was stand at each crossing to obtain the demographics, street crossing behavior, and perception regarding the safety and security of the crossing. The data were collected through questionnaire survey from the study location. Logistic regression model fitted through a backward step-wise method was used to find the predictors of crossing mode. This method removes the variables from the model which are not significant and thus made a best fitted model. The result from the logistic regression model indicates the following factors were significant at making a decision to choose which mode on crossing. They are frequency of footbridge usage, security, safety, proximity of the footbridge to ground crossing, and past involvement in accident.

Landa and Avila (2020) conducted a study in public university of Honduras to analyze the factors influencing the usage of pedestrian bridges. The population of study includes 2,180 students in which 330 was chosen as the sample size. Some of the attributes included in the study were using footbridge is stressful, tiresome, bad illumination, makes me feel lazy, too many stairs, going up and down the stairs tires me physically. Data collection include the perception of pedestrians towards the usage of footbridge. Data was collected through a questionnaire survey. It consists of 22 attributes with a five-point likert type response set Cronbach’s alpha was used to check the consistency of scale. Pearson’s r coefficient test was conducted to find the relationship between variables. Then logistic regression model was done to identify the variables, which can predict the usage of footbridge. The dependent variable in the regression model is dichotomous nature ie; usage or non-usage of footbridge. Five significant variables which explained the usage of pedestrian bridges were believing faster to cross the road, hurry, bad infrastructure, laziness and tiresome. The non-significant variables obtained from the model were age, using footbridge is stressful, crossing the street is stressful, thinking footbridge have good illumination.

Hasan et al., (2020) conducted a study on five major cities in Malaysia to identify the factors influencing footbridge usage. In addition, the study proposed some counter measures such as escalator, alerting poster, fine for jay walking, fence installation to analyze the change in usage of footbridge from the non-users. The data were collected through questionnaire survey from eight footbridges. They were located in busy areas and are with similar structural design. Decision tree analysis was used to find the decision whether to use or not a footbridge in machine learning. Again, a decision analysis was done in SPSS to find that whether the proposed interventions increase the footbridge usage among non-users. The study conclude that the frequency of usage and height were significantly depromoting the usage of footbridge. Also, while being in hurry the pedestrians will not choose the footbridge. It was found that among the interventions the pedestrians were ready to use the footbridge if escalators or fence were installed.

A study was done by Ojo et al., (2022) in Ghana to investigate the usage of footbridges. The footbridges were located in heavy traffic areas such as schools, shopping malls, bus stops etc. The interviewer collected data such as gender, age, occupation, and academic qualification, frequency of use and non-use of footbridge. A questionnaire survey was conducted among pedestrians. The data were collected from morning. afternoon and evening on all seven days. Totally 1852 data were surveyed. It took 5 minutes to complete a pedestrian response in the survey. The respondents were selected through accidental sampling technique. Structural Equation Modelling in SPSS was used to analyze the data. This was used to find the relationships between dependent and independent variables. Socio-demographic characteristics and reasons for footbridge usage and non-usage data were extracted. As compared to females, males were identified as more non-users of footbridge. The result showed that those with secondary education and the pedestrians involved in accident in the past were more footbridge users. The variables influencing the usage of footbridge were age, gender, training in pedestrian safety, frequency of use, walking distance, how often one crosses the stretch road, and length of stay.

Bhatia et al., (2022) conducted a study at six skywalk locations in Mumbai to analyse pedestrian skywalks using videographic and pedestrian perception survey. Video data was used to determine skywalks capacity utilization and level of service. The perception survey was used to find the influencing factors of its usage. The perception of 1118 pedestrians were used for modelling. The modelling of skywalk utilization was done using binary logistic regression. The study found that the following parameters influence the skywalk usage such asfrequency of usage, easiness, security at night, and its length whereas familiar location and agewas found to be negatively impacts its usage. Also, pedestrians occupation has an effect on the usage that students found to use skywalk less than servicemen. It suggested that suitable interventions such as elevators or escalators, improved security at night, security guards, CCTV, proper maintenance, and accessible entry-exit points can improve skywalk facility utilization

2.3 Utilization of crossing facilities

Demiroz et al., (2015) conducted a study in Turkey to analyze the road crossing behavior of pedestrians. The pedestrians use at grade crossing facilities in order to save time while crossing. It was also found that pedestrians crossing speed decreases if the speed limit of vehicle decreases.

Ferenchak (2016) conducted a study in Bangalore, Karnataka, India to analyze the behavior of pedestrians while crossing. The data was collected through manual observation. Only behaviour of individual pedestrians was evaluated. Gender, age waiting time, utilization of crosswalks were estimated. Quadratic and logistic regressions were used in the model. It was found that with increase in age, pedestrian delay and utilization of crossings increases and thus the conflict decreases. As compared to females, males have less waiting time almost half of female waiting time and the chance of conflict is more.

Rankavat and Tiwari (2016) conducted a study in Delhi to analyze the perception ofpedestrian regarding underutilization of pedestrian crossing facilities. In the study dependent variables chosen were the use of zebra crossing, overpass and underpass and the independent variables were age, gender, safety and convenience. From seven different locations including facilities such as zebra crossing, pedestrian underpasses and overpasses, five hundred pedestrians were interviewed. An ordinal logit model was used to analyze the factors related to the use of crossing facilities. Kruskal–Wallis tests was conducted to evaluate the usage of different crossing facilities. To compare the preference of different crossing facilities Spearman's correlation was used. It was found that a positive correlation exists for the preference of zebra crossing also showed that convenience perception is statistically significant for the use of zebra crossings. It was obtained that safety perception of underpass is less for females and according to perception it was found that zebra crossing was convenient but not safe.

Sinclair and Zuidgeest (2016) conducted a study in South Africa regarding the pedestrian crossing choices. Survey was conducted for both footbridge users and at-grade users. Time and distance plays an important role in choosing crossing facility.

Anciaes and Jones (2018) conducted a study in London to analyze pedestrians perception regarding different crossing facilities. It was found that they are influenced by safety, convenience, crossing time, accessibility, and personal security. The footbridge and underpasses were rated below at grade crosswalks. The analysis was done using mixed logit model. The result showed that pedestrian is willing to walk extra 2.4 minute to avoid footbridge and 5.3 min to avoid underpasses in order to use signalized crossing.

Prajapati et al., (2018) conducted a study at six locations in Delhi to identify the optimum crossing facility. The average delay caused due to pedestrian and vehicles were estimated. The data collection was done through videographic survey. Delay occurring in grade separated and at- grade crosswalks were evaluated. From the selected area, at four locations at grade crossing was suggested with optimum signal time and minimum delay. As compared to footbridge condition, at grade facility is preferred.

A study was conducted by Al bargi and Daniel (2020) in Malaysia to analyze the utilization of zebra crossings among pedestrians. The data was collected from 9 study locations through videographic method. The pedestrians crossing through zebra crossing were directly obtained from the videographic survey. The data was analyzed using binary logit model. This was used to find the probability of pedestrians choosing zebra crossing. The following variables such as speed of vehicle, gender, age and presence of baggage were found to be an effect on the pedestrians decision regarding the utilization of zebra crossing.

Patra and Perumal (2020) conducted a study in Mumbai city, India to analyze the perception of pedestrians between at grade signalized crosswalk and footbridge. The data was collected through videographic method. Total 1134 pedestrian data was collected. Binary logit model was used to analyze the significant variables influencing the pedestrian decision regarding the choice of different facility. The dependent variable taken for the study was crossing time. The independent variables were age, gender, luggage size, accompanied by a child, mobile phone user. It was found that crossing time ratio influence the choice of crossing facility and it should be within 0.75 for more than 50% footbridge usage.

Chandrappa et al., (2021) conducted a study in Kolkata, India to identify the optimum crossing facility. Ranking method was used to find the preference of different crossing facility. Ordinal logistic regression method was used to analyze the ranked data and found that signal time with 2 minute waiting time and footbridge were preferred. Cost benefit analysis was done to find the economic alternative. The result showed that footbridge was economical.

2.4 Machine Learning Techniques

Banerjee et al., (2020) conducted a study in India regarding the overpass utilization using machine learning techniques. The aim of the study was to identify the factors depending upon the usage of foot over bridge under safety and security, mobility friction, vertical connectivity and horizontal connectivity. The gradient boosting machine outperformed other models such as random forest and generalized linear model.

Jamal et al., (2021) conducted a study on prediction of injury severity of crashes using machine learning algorithms. The different algorithms used in the study are extreme gradient boosting, random forest, decision tree and logistic regression model. The extreme gradient boosting model outperformed other model with a high accuracy in predicting the injury severity.

2.5 Summary

From the literature review, machine learning techniques were chosen for the classification modelling of preferred crossing facility. The study included the factors like environmental conditions, physically disabled pedestrians. These factors were not considered in the above studies. In India, only limited studies had been done regarding the perception of pedestrians towards the usage of footbridge using machine learning techniques.

CHAPTER 3

METHODOLOGY

3.1 General

The methodology conducted in the study is shown in the form of a flowchart in Figure 3.1.

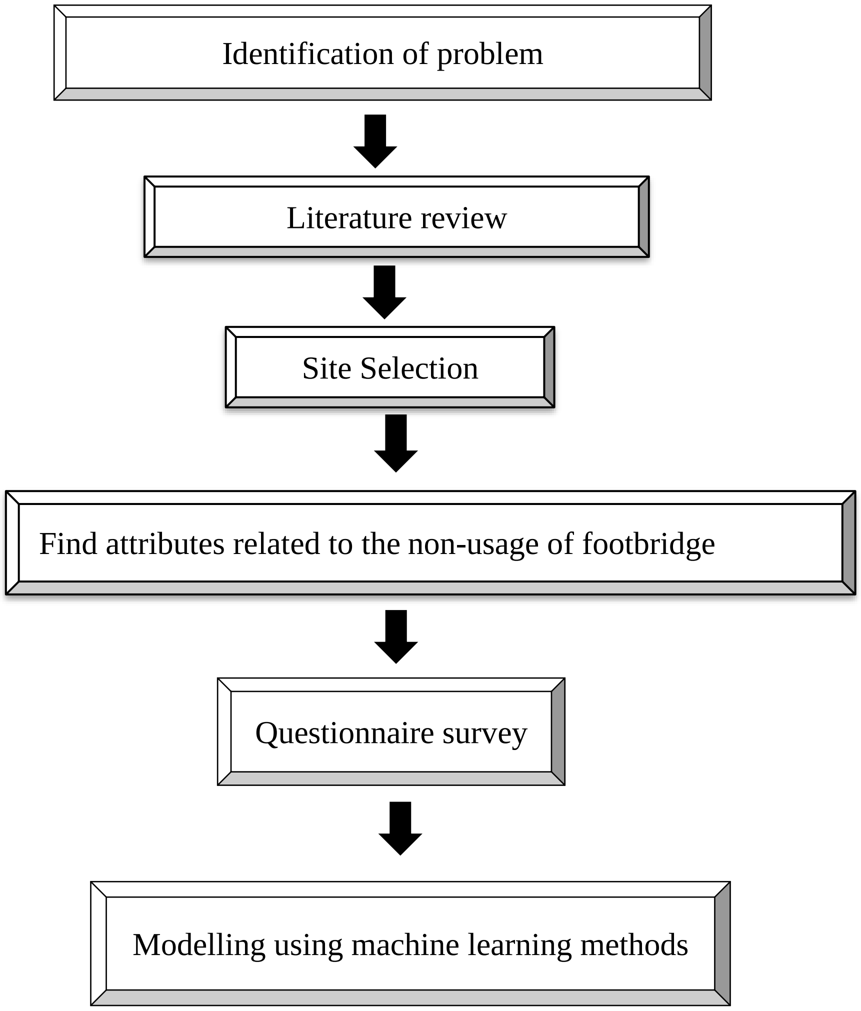

Figure 3.1 Methodology

3.2 Site Selection

The selected three footbridge was located in Kollam city at High school junction, Convent junction and Chemmanmukk. The footbridge in High school junction was situated near signalized intersection while in Chemmanmukk it was situated near unsignalized intersection. The footbridge in convent junction was situated near midblock. The three site locations are shown in Figure 3.2, 3.3 and 3.4. The three footbridges were similar in structural design and located near schools to attract school students and for their safety.

Figure 3.2 Footbridge in High School Junction

Figure 3.3 Footbridge in Convent Junction

Figure 3.4 Footbridge in Chemmanmukk

3.3 Identification of Attributes

Footbridges are provided in many areas for the safety of pedestrians as well as for the easy mobility of vehicles. Even though this facility is sparsely used by the pedestrians compared to at- grade crosswalks. It is important to identify the factors which depromoting the usage of footbridge. The following methods are used to identify the factors relating to the usage of footbridge.

3.3.1 Literature Reviews

One of the main objectives in this study is finding attributes regarding the non-usage of footbridge. Some of the attributes were identified from the literature are tiredness while climbing the stairs, perceived easiness, height fear, safety, illumination, being in hurry, bad infrastructure, involvement in accident while crossing the road, frequency of using footbridge. The literature study helps in identifying the gaps involved in it.

3.3.2 Pilot Study

The pilot study is a small-scale study conducted to identify the factors influencing the non-usage of footbridge. The study was conducted at High school junction, Convent junction and Chemmanmukk in Kollam. The responds were collected offline from the above-mentioned locations during peak as well as off-peak hours. Total 44 data was collected including 26 females and 18 males. The socio-demographic characteristics such as age, gender, education, occupation and perception of pedestrians towards the use of footbridge were collected. Also, some counter measures were asked to the non-users to know their willingness to use footbridge if these facilities are provided. The average duration for collecting the responds from a respondent is approximately 5 minutes. The pilot study helped to add some variables. The study identify that most of the pedestrians prefer zebra crossing while crossing the road.

3.3.3 Discussion with Subject Experts

Discussion with subject experts helped to fix the factors influencing the non-usage of footbridge. The final questionnaire is finalized using above mentioned methods.

3.3.4 List of Attributes

- ➢ Socio Economic Characteristics
- Age
- Gender
- Education
- Occupation
- ➢ Frequency of footbridge usage
- ➢ Pedestrian accident history
- ➢ Preference of using footbridge on following situations
- Hurry
- Night
- Weather
- Unfamiliar Location
- Heavy Traffic
- ➢ Perception of pedestrian towards footbridge
- Easiness
- Safety
- Tiresome
- Infrastructure
- Illumination
- Stressful
- Height fear
- Obstructions like seller
- Stray dogs
- Entry and exit accessible
- Street crossing stressful
- Street crossing faster
- ➢ Counter Measures
- Escalators / Elevators
- Alerting poster
- Fence installation
- Fine for illegal crossing

- CCTV
- Good illumination
- Free cooler water
- Mandatory national policy
- Seating arrangement
- Security staff

These attributes were finalized for the questionnaire survey. The counter measures were asked to know about the change in usage of footbridge if these facilities are provided.

3.4 Data Collection

The data collection for the study was done through field questionnaire survey. The questionnaire was divided into three sections. The first section includes Socio- Economic characteristics such as age, gender, education, occupation. The second section includes the perception of pedestrians towards the usage of footbridge. The third section consist of some of the interventions aimed at increasing footbridge usage. Lastly suggestions of the pedestrian were also included. The questions were translated to respondents native language such that it helps them easier to understand and by helps to reduce errors while filling. The participants were randomly selected. A total 500 data was collected from the above-mentioned locations. The data was collected during peak as well non-peak hours. The duration of data collection was around 3 weeks and it took approximately 5 minutes to complete the data from a respondent. The collected data was used to find the factors responsible for the non-usage of footbridge.

3.5 Machine Learning Techniques

Different machine learning models are being investigated for simulating the potentially nonlinear correlations of independent variables with the target variables and to overcome the limitations of statistical approaches. This method is widely used because of their high accuracy than manual computations. It is used to create algorithms that can take input data and do statistical analysis in order to predict the output. These machine learning algorithms can be divided into three categories: supervised, unsupervised, and reinforcement learning. These are shown in Figure 3.5.

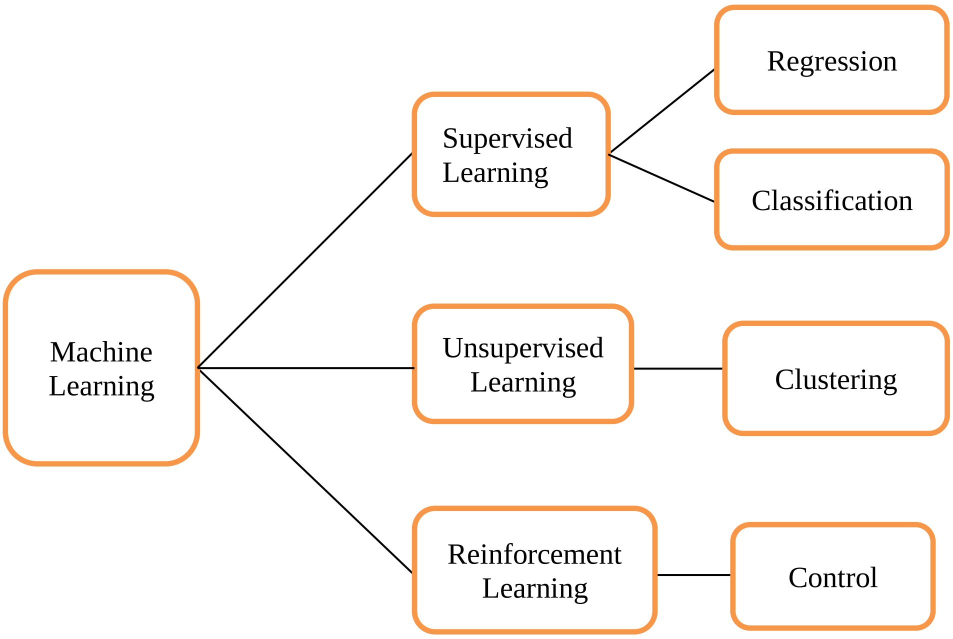

Figure 3.5 Machine Learning Techniques

Supervised machine learning learns from the data to make predictions. Unsupervised learning analyzes the data which is not labelled. It learns the data without any human guidance. Reinforcement learning uses prediction based on feedback. The feedback is either from a human or any algorithm.

Machine learning classifiers are supervised training algorithms that are used to categorize datasets. They have a multi-dimensional data processing power, flexibility in implementation, and improved predicting skills, all of which can lead to promising outcomes. Classification problems predicts the categorical output variable from the input variables. The different algorithms that come under the classification study are as follows:

- Random Forest
- Decision Tree
- Support Vector Machine
- K-Nearest Neighbor
- Logistic Regression

3.5.1 Random Forest

Random forest involves the processing of constructing numerous distinct decision trees using varied subsets of predictor information there by reducing the overfitting issue. It consists of root node, internal node and leaf node. In a decision tree, the top node is referred to as the root node.

It means to divide according to the attribute's value. A decision node signifies a characteristic or attribute, a branch denotes a decision rule, and each leaf node denotes the outcome in a decision tree. Each tree forecasts a value for the target variables. The data set is divided into number of trees as shown in the figure. As the number of trees increases the prediction accuracy also increases. The final result is selected by majority voting or averaging the result of each trees. A simple random forest model is shown in the Figure 3.6.

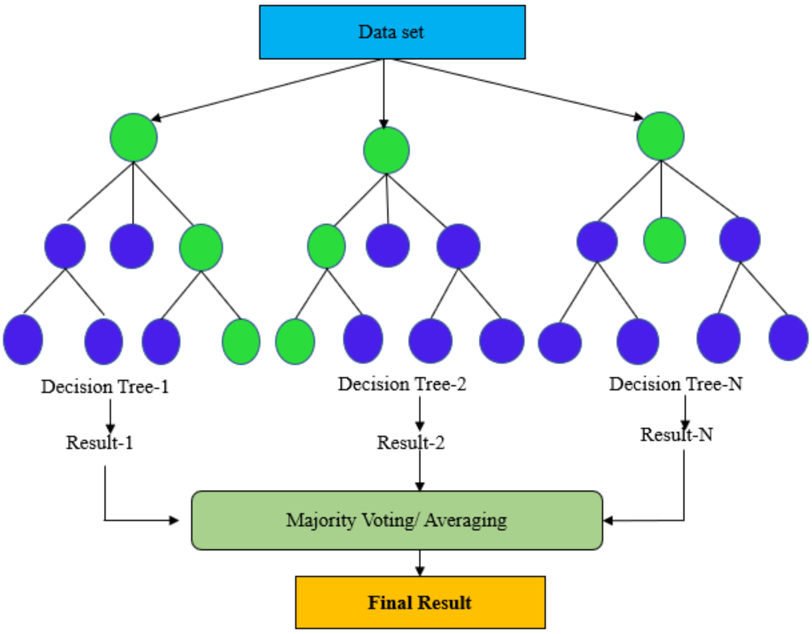

Figure 3.6. Random Forest

Advantages

- Accuracy of model is generally very high
- It reduces overfitting
- It provides weightage of variables
- It is efficient for large data set

Disadvantages

- It is complex
- It takes more time for training the data

3.5.2 Decision Tree

This algorithm learns decision rules from data. It divided the data set based on certain features using root node, decision nodes and leaf nodes. This approach is referred to as a decision tree since a tree can symbolize the collection of splitting rules used to divide the dataset. It chooses the optimum attribute to split the data. A simple decision tree model is shown in the Figure 3.7.

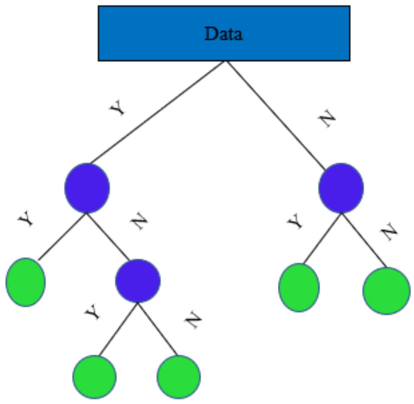

Figure 3.7 Decision Tree

Advantages

- East to interpret and execute
- It can handle non- linear data effectively
- It provides weightage of variables

Disadvantages

- It may overfit the model
- Some forms of data might not work well with them.
- They may create complex trees
- Prediction accuracy is low

3.5.3 Support Vector Machine

When classifying data, an SVM algorithm selects the optimum border to divide the data into distinct classes. The margin, or the distance between the boundary and the nearest data points from each class, is maximized by selecting a boundary that does so. It creates a hyperplane between the margins. The extreme points are known as support vectors. A gap as big as feasible divides the training data into categories and represents them as points in space. The next step is to add new points to the grid by predicting the category and space they belong in. A simple support vector machine model is shown in the Figure 3.8.

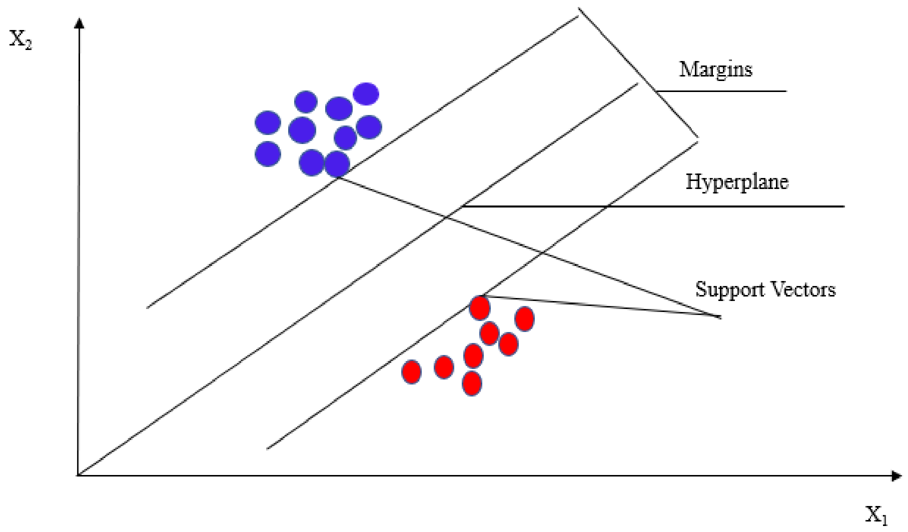

Figure 3.8 Support Vector Machine

Advantages

- When the data is high-dimensional, it works well
- It is efficient
- Less prone to overfitting

Disadvantages

- It is not able to handle missing values
- Not good for large data sets
- It takes more time for training the data

3.5.4 K-Nearest Neighbor

This algorithm finds the K-nearest neighbors to an observation location. The input data should be taken into consideration while determining the value of k. A greater value of k would be preferable if the input data contained more noise or outliers. To prevent classification ties, picking an odd value for k is advised. The best k value for the given dataset can be chosen with the use of cross-validation techniques. The target variable with the highest ratio is then predicted after evaluating the proportions of each type of target variable using the K points. A simple k-nearest neighbor model is shown in the Figure 3.9.

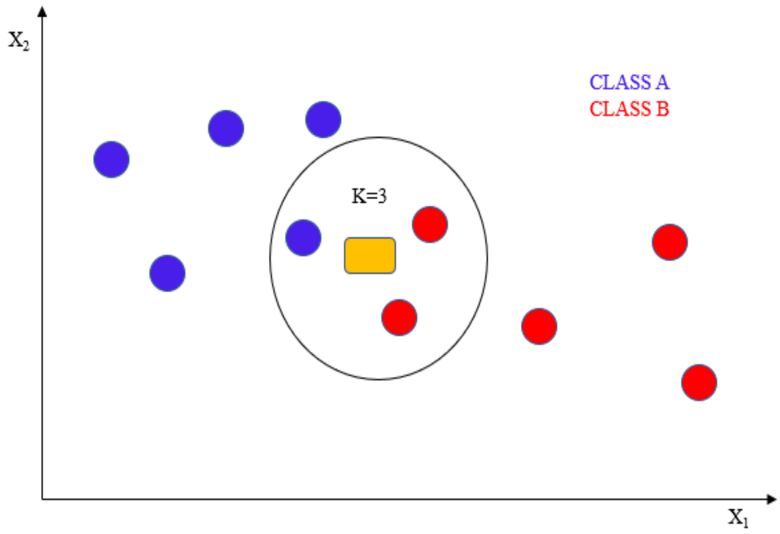

Figure 3.9 K-Nearest Neighbor

Advantages

- It is easy to interpret
- It is simple

Disadvantages

- Easily affected by outliers
- It favors a class that appears more frequently in the dataset.
- It is not able to handle missing values
- It does not work well when the data is high-dimensional
- It is prone to overfitting

3.5.5 Logistic Regression

Logistic regression is used in classification problems, where the objective is to estimate the likelihood that a given instance belongs to a particular class. It's a type of statistical technique that examines the correlation between a group of independent variables and a set of binary dependent variables. It employs a sigmoid function to estimate the probability for the specified class using the output of the linear regression function as input. It is an effective method for making decisions. This model is shown in Figure 3.10.

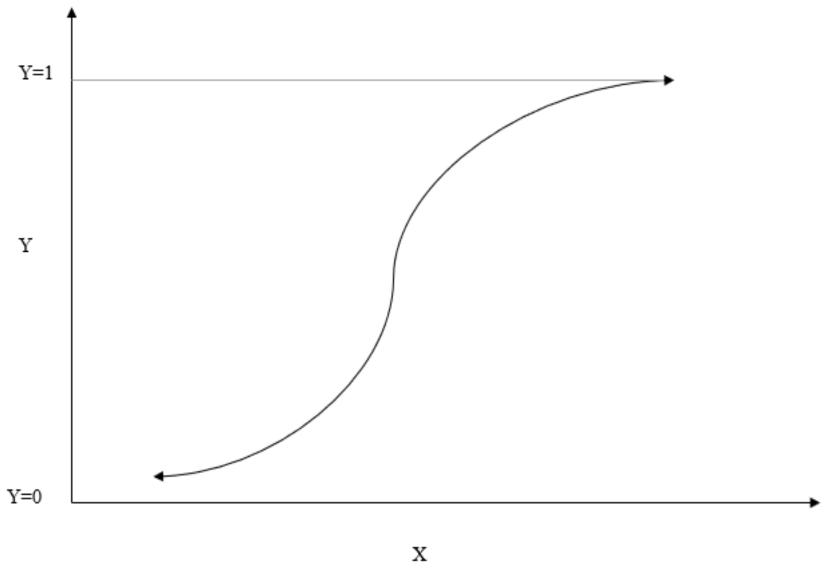

Figure 3.10 Logistic Regression

Advantages

- It is one of the simplest methods
- The training time is very less
- It can be implemented quickly and simply

Disadvantages

- It has difficulty in capturing complicated relationships.
- It is sensitive to outliers in this algorithm.
- It requires large data as well as enough training instances for each category it needs to recognize.

3.6 Steps in Classification Model

The different steps in classification model using machine learning algorithms are data labelling, feature selection, data cleaning, model selection and model training and evaluation. The steps are shown in the form of a flow chart in Figure 3.11.

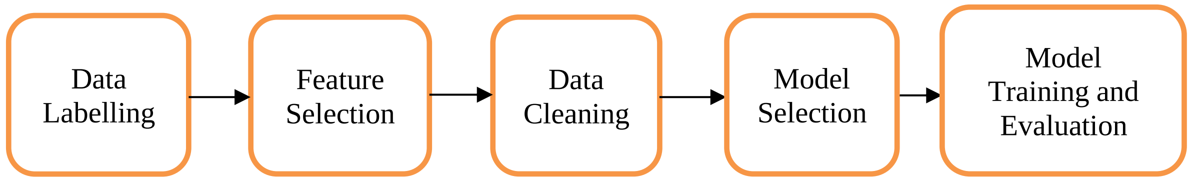

Figure 3.11 Steps in Classification Model

3.6.1 Data Labelling

Dummy variable encoding and ordinal encoding were used for encoding the variables. Dummy variable encoding is used when there is no link between categories in categorical data such as gender, occupation, crossing facility, frequency of footbridge usage, pedestrian accident history. It always uses C-1 binary variables to represent C categories. Ordinal encoding is used when there is a relationship between categorical data. The relationships between the integer numbers are naturally organized and machine learning algorithms may be able to recognize and take advantage of this relationship. Ordinal encoding was assigned to age, education, hurry, night, weather, unfamiliar location, heavy traffic and for likert type questions such as ease of use, safety, tiresome, infrastructure, good illumination, footbridge stressful, height fear, disability comfort, obstruction free, dog free, accessibility, street crossing stressful, street crossing faster.

The data of variables was renamed as follows.

CF_FB - Crossing preference Footbridge

UL – Unfamiliar Location

HT – Heavy Traffic

FBE – Footbridge ease of use

FBS – Footbridge safety

FBT – Footbridge Tiresome

FBBI – Footbridge Bad Infrastructure

FBGI – Footbridge Good Illumination

FBST – Footbridge Stressful

FBHF – Footbridge Height Fear

FBCDP – Footbridge comfort for disabled pedestrians

FBOF – Footbridge obstruction-free

FBA – Footbridge accessible

FBSD – Footbridge free from stray dogs

ZCST- Zebra Crossing Stressful

ZCF – Using Zebra Crossing is Faster

OCW – Occupation Company Worker

OGW- Occupation Government Worker

OS – Occupation Student

OU – Occupation Unemployed

FBUN – Footbridge usage never

FBUS – Footbridge usage sometimes

PAH – Pedestrian accident history

3.6.2 Feature Selection

Feature engineering is usually the initial step in creating a model. This was executed using mutual information. This method is suitable for both regression and classification problem. A data set's relevant variables can be focused on with the aid of feature selection, which can also be used to get rid of collinear variables. It aids in lowering the amount of noise in the data set. This method was used to reduce the dimensionality of a data set and to increase the performance of machine learning models. Categorical features in a dataset are tested using this method. The mutual information between each feature with the targeted variable is determined in order to choose the features with the highest value.

The range of the mutual information score is from 0 to infinity. The closer the relationship between this feature and the target, the higher the value; hence, we should include this feature in the training dataset. The low score indicates a tenuous connection between this feature and the target variable. This method decreases the input feature set by retaining the class discriminating information for classification problems. Feature reduction can decrease a problem's complexity or computing time and even increase the model's accuracy.

3.6.3 Data Cleaning

The quality and performance of a model can be greatly affected by data cleaning. Errors and inconsistencies in the data are removed during data cleaning. After feature selection data cleaning was done to remove unwanted variables.

3.6.4 Model Selection

A mathematical model containing several parameters that must be learned from the data is referred to as a machine learning model. Different models were incorporated using machine learning methods in order to identify the best model. The different models adopted were random forest, decision tree, support vector machine, K nearest neighbour and logistic regression. The analysis was done using google colab. The model which outperformed well is taken as the best model.

Hyperparameters, on the other hand, are a different class of parameter that cannot be directly learned through ordinary training. Before the training itself starts, they are typically rectified. These parameters convey crucial model characteristics including complexity and learning rate. Models may have a large number of hyperparameters, and determining the ideal set of parameters can be approached as a search problem. The methods adopted for hyperparameter tuning are GridSearchCV and RandomizedSearchCV.

GridSearchCV function is used for hyperparameter tuning using cross validation. The machine learning model is assessed for a variety of hyperparameter values using this technique. This method is known as GridSearchCV because it analyses a grid of hyperparameter values to find the optimum set of hyperparameters. It is used to examine a variety of hyperparameters for a certain machine learning algorithm, estimating the effectiveness of each set of hyperparameters using cross-validation.

Cross-validation is done in GridSearchCV in addition to Grid Search. The model is trained using cross-validation. The data is split into two categories train data and test data before the model is trained using the data. The procedure of cross-validation further splits the train data into the train data and the validation data. K-fold Cross-Validation is the most widely used kind of cross-validation. The procedure of dividing the train data into k parts is iterative. One division is kept for testing purposes while the remaining k-1 partitions are used to train the model during each iteration. In the following iteration, the following partition will be designated as test data, the remaining k-1 as train data, and so on. The model's performance will be recorded at each iteration, and the average of all the performances will be provided. Thus, evaluating the optimum hyperparameters using Grid Search and cross-validation requires a significant amount of work over time. The model which outperformed other models is chosen as the best model.

3.6.5 Model Training and Evaluation

Train test split function is used for splitting the data for training and testing. For modelling, only 20% data was used and the remaining 80% data was used for training. So out of 500 data collected 100 data were used for modelling. The model performance was evaluated using the confusion matrix. It demonstrates the correctly classified and misclassified data. The performance metrics such as accuracy, precision and recall of each algorithm was compared to find the best model.

3.7 Confusion Matrix

The effectiveness of a machine learning model on a set of test data is summarized by a confusion matrix. It is used to access the effectiveness of the model using performance metrics such as precision, recall, accuracy and F1 score. The confusion matrix for classification problem consists of four potential outcomes, namely the true positive rate (TP), true negative rate (TN), false positive rate (FP), and false negatives rate (FN). A figure of confusion matrix is shown in Figure 3.12.

Confusion matrix was used to identify the correctly classified and misclassified data.

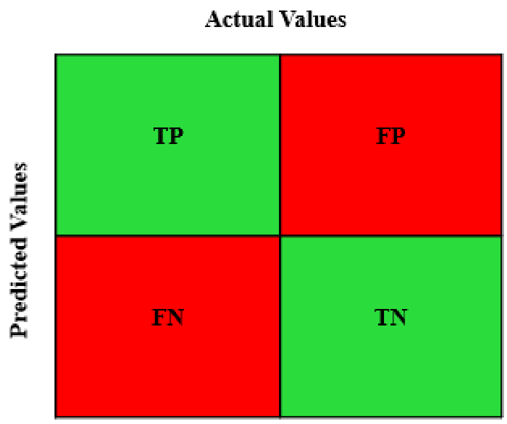

Figure 3.12 Confusion Matrix

- True positive: The predicted value of the algorithm is positive and the actual value is also positive.
- False positive: The predicted value of algorithm is positive but actually it is negative
- False negative: The predicted value of algorithm is negative but actually it is positive
- True negative: The predicted and actual value is negative
- Accuracy: It is defined as the ratio of correct predictions to the total number of predictions. It is calculated using Equation 3.1. A model is very accurate if it consistently generates correct predictions. It is used to assess the performance of classification model.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (3.1)$$

- Precision: It is defined as the ratio of true positives to all true positives and false positives. It examines the number of false positives that were included in the sample. The model has 100% precision if there were no false positives. The value of precision will decrease if there are more false positive. Precision is calculated using Equation 3.2.

$$\text{Precision}= \frac{True\ positive}{Predicted\ Results} = \frac{TP}{TP+FP} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.2)}$$

- Recall: It is defined as the ratio of true positive to sum of true positive and false negative. It examines the amount of false negative that were included in the sample. The model has 100% recall if there were no false negative. The value of recall will decrease if there are more false negative. Recall is calculated using Equation 3.3.

$$\text{Recall}= \frac{True\ positive}{Actual\ Results} = \frac{TP}{TP+FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.3)}$$

- F1 score: It is the harmonic mean for recall and precision. A classification model's overall performance is assessed using the F1-score. It is calculated using Equation 3.4.

$$\text{F1 score} = 2\times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.4)}$$

3.8 Summary

This chapter deals with the detailed methodology of the research work. The techniques used in the analysis was explained in detail. This gives a clear understanding of the machine learning techniques and the basic steps in execution of this method.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 General

The results obtained from the modelling is explained in this section. The performance metrics of different algorithm is evaluated.

4.2 Data Analysis

The data analysis was done through machine learning methods in order to classify the footbridge users and non-users and to identify the weightage of variables predicting the usage of footbridge. Data cleaning was a difficult task since the data of footbridge users and non-users were collected separately. The data was analyzed using google colab.

The data was used to develop count plot of each variable as shown in Figure 4.1. It indicates the count of each value of variables and also gives the idea of distribution pattern of data.

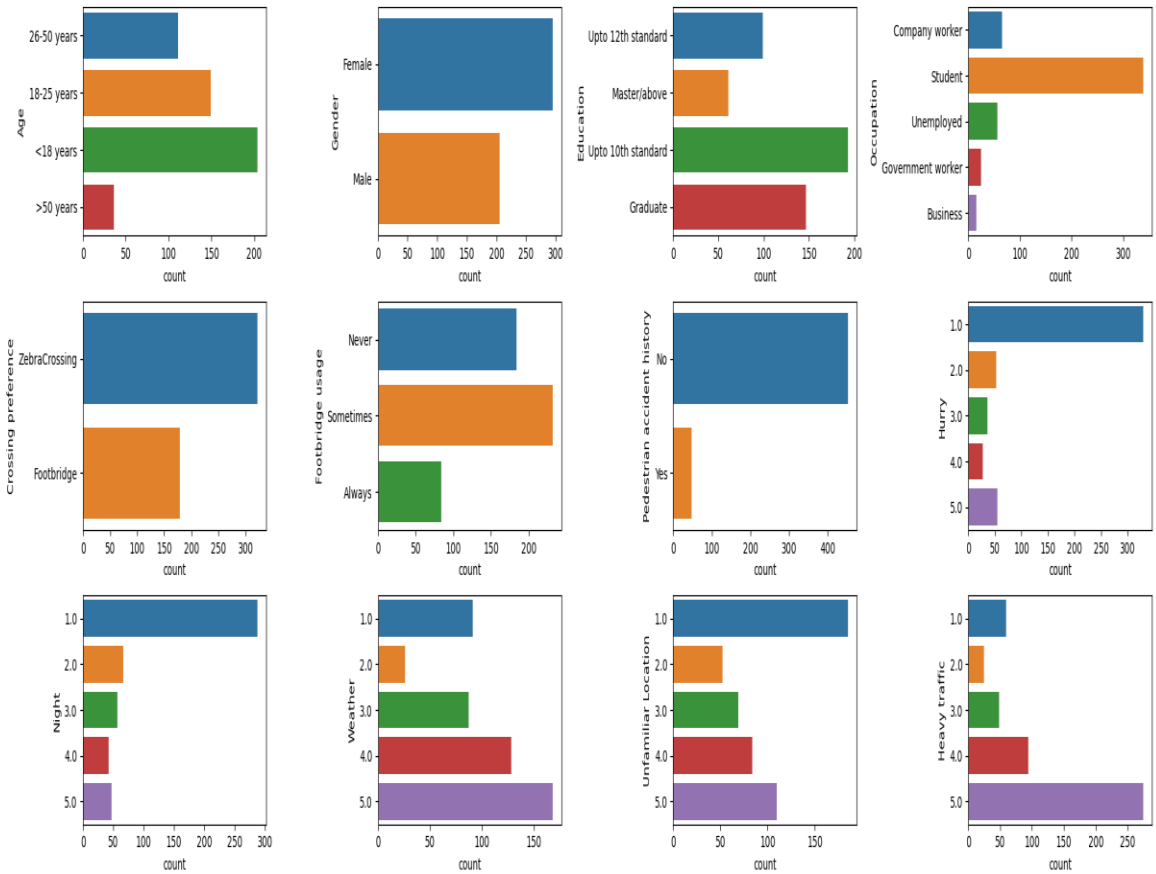

Figure 4.1 Count Plot

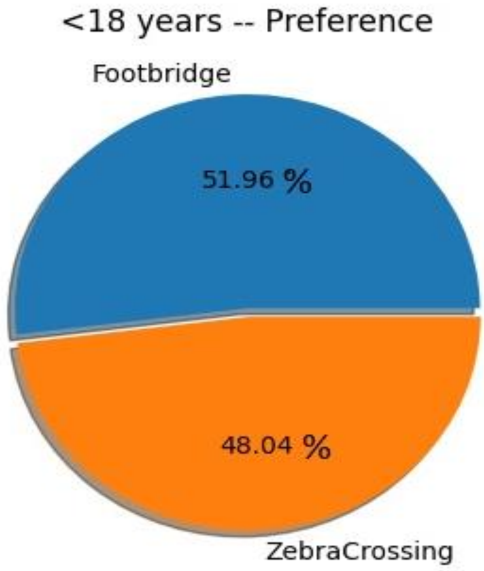

Figure 4.2 Pie chart of Age group <18years

From the count plot we can see that most of the respondents were in the age group <18 years, in which almost 52% prefer footbridge and 48% prefer zebra crossing while crossing as shown in Figure 4.2. School students use footbridge because the schools provide instructions to the students to use footbridge. The female respondents were more since most of the male use private mode. The preferred crossing facility for most of the pedestrians is zebra crossing. Frequency of footbridge usage is opted as never and sometimes by most of the respondents. Preference of footbridge usage during hurry, night, weather, unfamiliar location and heavy traffic were collected on a rating scale in which 1 indicates poor and 5 indicates good. Most of the pedestrians are not willing to use footbridge in hurry situations and during night. Also, they are willing to use footbridge on extreme heat or sunny day. Weather is an important criterion in choosing the crossing facility. In unfamiliar locations most of them are not willing to use footbridge especially females due to security perception but some are totally agreeing to use because of the unawareness of traffic in that area. Almost all are willing to use it on heavy traffic.

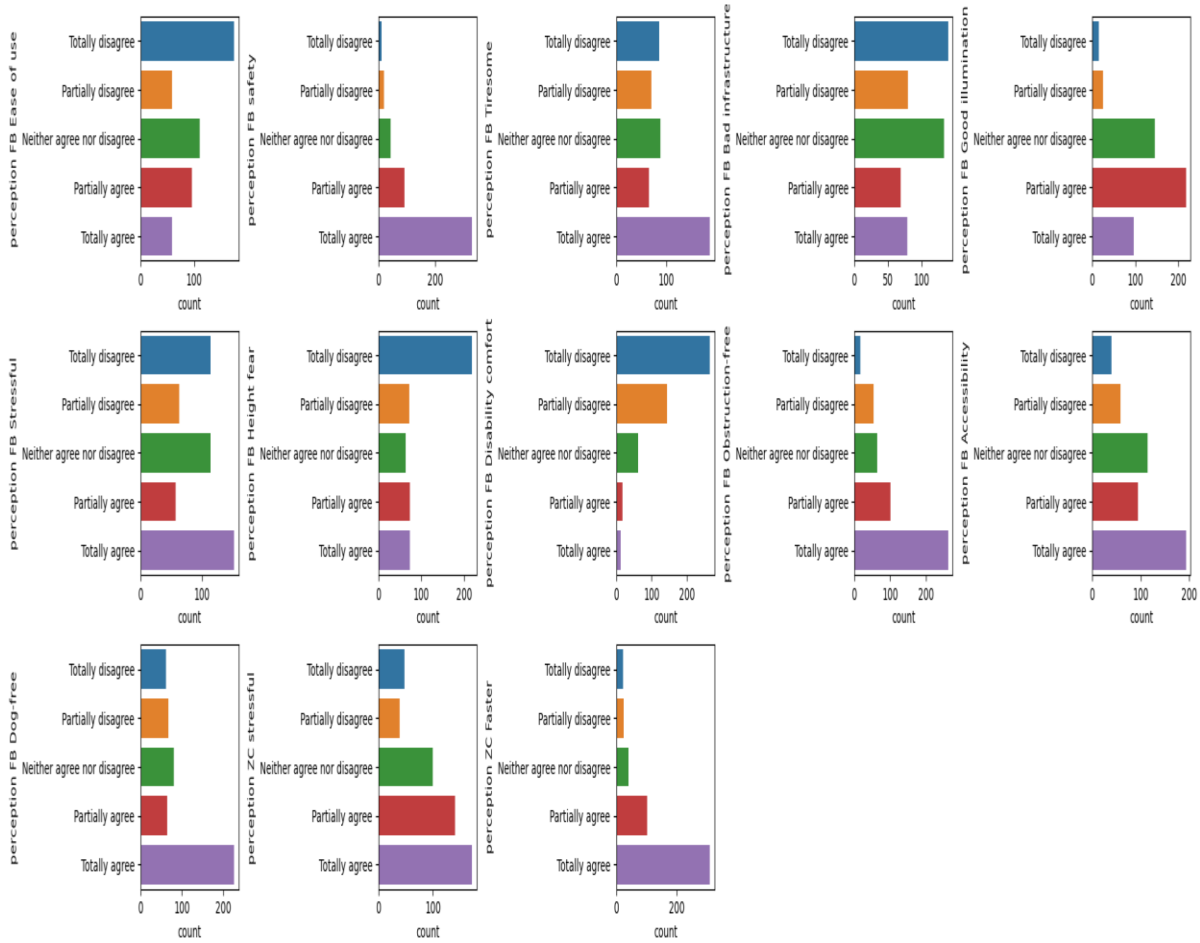

Figure 4.3 Count Plot of Perception of Footbridge

The questions regarding the perception of footbridge are provided as a likert type scale such as totally disagree, partially disagree, neither agree nor disagree, partially agree and totally agree as shown in Figure 4.3. From the count plot it is easily understand that footbridge was not easy to use. Its usage makes them feel tired and for some its usage makes them feel stressful. Almost all of them agreeing that it is safe but females are not willing to use it on unfamiliar locations during night. This is not comfort for disabled pedestrians. Almost all of them agreeing that footbridge is free from obstructions like sellers, stray dogs and easily accessible. Even though pedestrians preferring zebra crossing, they agreeing that its usage make them feel stressful especially when the vehicles do not slow down and have a fear of accidents. But they use zebra crossing because they feel that it takes only less time and is easy to cross in signalized intersection or when a group of people is there to cross.

4.3 Feature Selection

Feature selection was done to exclude the variables that does not make much relation to the target variable. Feature selection was done using mutual information. The low score indicates a tenuous connection between this feature and the target variable. Only top feature are selected for modelling.

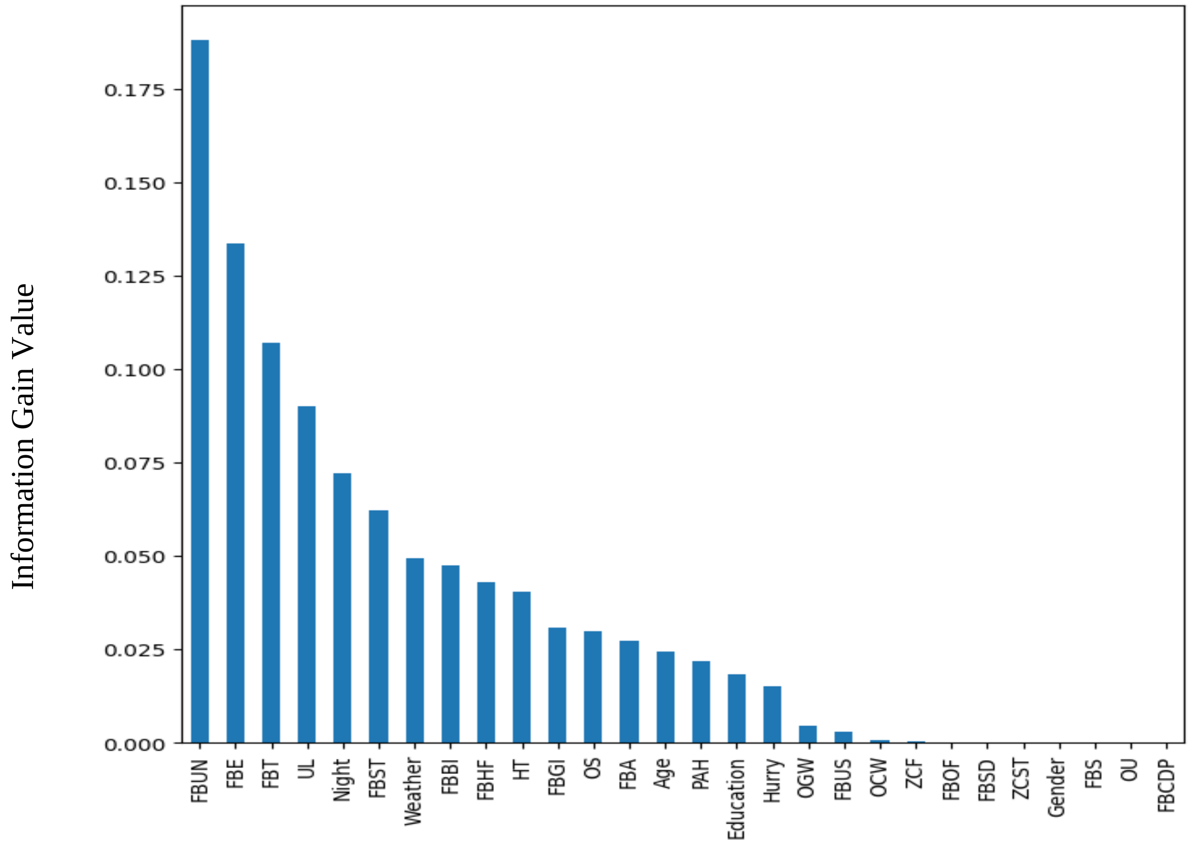

Figure 4.4 Feature Selection using Mutual Information

From the Figure 4.4 the top features such as FBUN, FBE. FBT, UL, Night, FBST, Weather, FBBI, FBHF, HT, FBGI, OS, FBA, Age, PAH, Education, Hurry were selected for modelling. Other features were neglected for modelling to increase the accuracy of the model. This method was used to reduce the dimensionality of a data set and to increase the performance of machine learning models.

Table 4.1 Mutual Information Gain Value

Features	**Mutual Information Gain Value**
FBUN	0.19
FBE	0.13
FBT	0.1
UL	0.09
Night	0.07
FBST	0.06
Weather	0.05
FBBI	0.05
FGHF	0.04
HT	0.04
FBGI	0.03
OS	0.03
FBA	0.03
Age	0.03
PAH	0.02
Education	0.02
Hurry	0.02

The Table 4.1 shows the mutual information gain value of variables in descending order. The frequency of footbridge usage shows a high gain value and least value is obtained for pedestrian accident history, education and hurry.

Table 4.2 Data for Classification Model

Age	FBE	FBT	FBBI	FBGI	FBST	FBHF	FBUN	Weather	UL	HT
2	0	4	2	2	4	0	1	3	1	2
1	4	0	3	2	1	0	0	2	1	1
0	1	0	0	3	0	0	0	4	1	4
1	2	2	3	2	4	0	1	4	3	3
1	0	4	0	4	4	2	1	1	2	1
2	2	0	0	4	0	0	1	1	1	1
1	2	4	0	4	4	0	0	4	1	4
2	0	4	2	4	4	0	0	4	1	4
0	4	0	4	1	4	4	0	3	2	4
1	2	4	2	3	0	0	0	4	1	4
1	0	4	1	3	4	0	0	4	3	4
1	4	4	1	2	3	3	1	1	4	3

The variables were encoded using dummy variables and ordinal encoding depending upon the type of data. A sample of data is shown in Table 4.2. Dummy variable encoding was assigned to variables like gender, occupation, crossing facility, frequency of footbridge usage, pedestrian accident history. Ordinal encoding was assigned to age, education, hurry, night, weather, unfamiliar location, heavy traffic and for likert type questions such as ease of use, safety, tiresome, infrastructure, good illumination, footbridge stressful, height fear, disability comfort, obstruction free, dog free, accessibility, street crossing stressful, street crossing faster.

4.4 Heat Map

A heatmap is a graphical display of data in which values are represented by colours. This was developed to know about the correlation of each variable with respect to the target variable and the relation of independent variables with each other. The target variable taken for the study is crossing preference that is Footbridge.

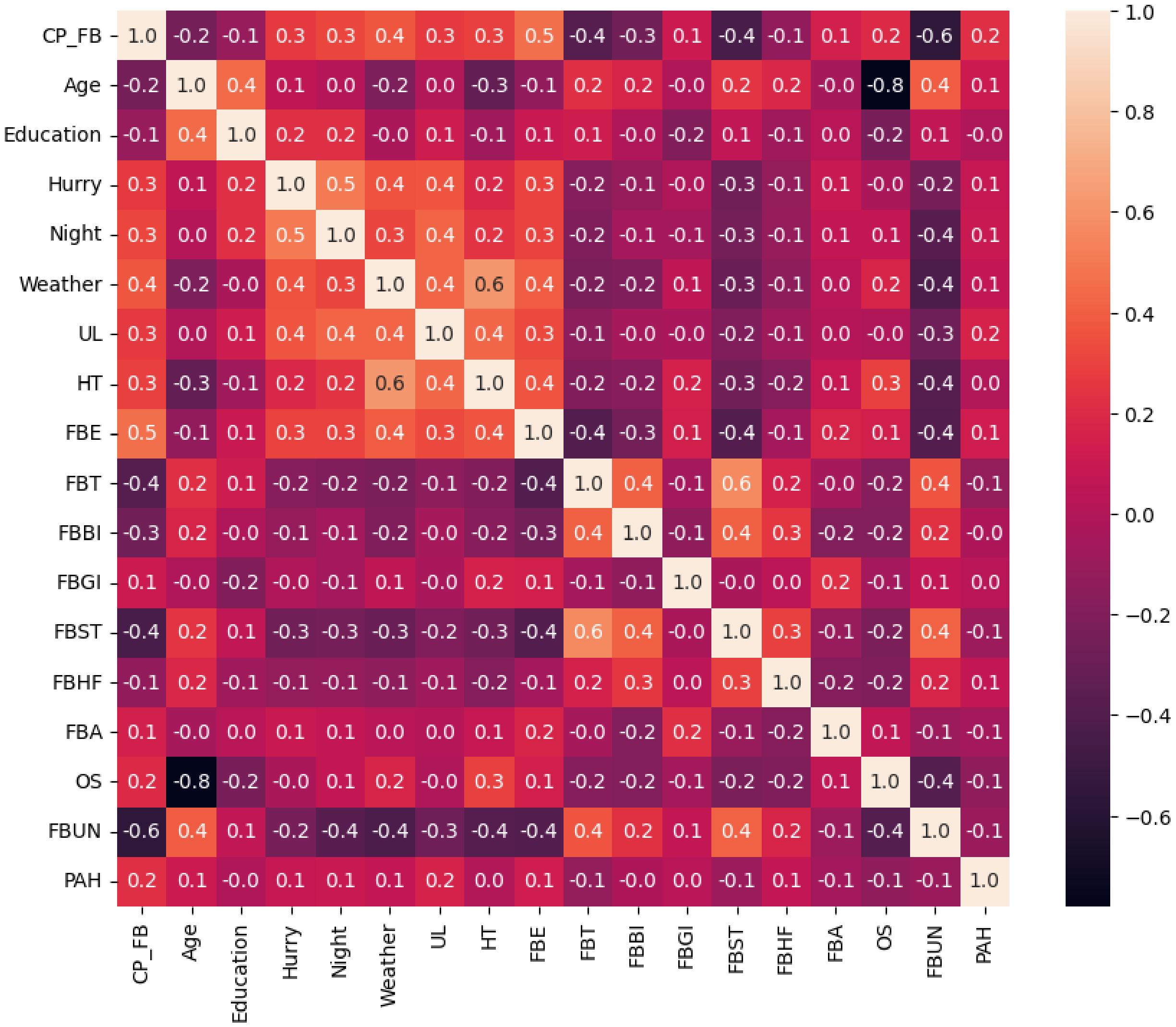

Figure 4.5 Heat Map after Feature Selection

Figure 4.5 shows the heat map of variables after feature selection. This is a binary classification problem. The dependent variable is crossing preference that is footbridge which is encoded as 1. The independent variable selected after feature selection were Age, Education, Hurry, Night, Weather, UL, HT, FBE, FBT, FBBI, FBGI, FBST, FBHF, FBA, OS, FBUN, PAH. As age increases the preference of using footbridge decreases. Also, more the people educated, the preference of using footbridge decreases. From the count plot itself it was observed that footbridge users were mostly the school students. Footbridge is preferred during hurry, night, weather that is extreme rainy or sunny day, unfamiliar location, heavy traffic. If footbridge is easy to use and having good illumination then its preference increases. If using footbridge makes them feel tired, stressful and having height fear then its preference decreases. Also, if it is bad infrastructure then its usage decreases. The responds indicating the frequency of footbridge usage stated as never increases then its preference decreases. It is because for most of the pedestrians using at grade cross walks for road crossing becomes a part of their habitual action. They even didn’t try to use it once especially the aged pedestrians. The pedestrians who had past involvement in accident uses the footbridge.

4.5 Classification Model

The classification model of footbridge users was developed by different algorithms using machine learning methods. The model with high accuracy was selected as the best model. The algorithms were developed by tuning the hyperparameters using grid search function in order to get the best output of the algorithm. The different algorithms used are random forest, decision tree, support vector machine, k-nearest neighbor and logistic regression.

4.5.1 Random Forest

Using random forest algorithm, the accuracy obtained from training data is 84% and for testing data it is 83%. Here the accuracy of training and testing data is similar which indicates the model is fit.

Table 4.3 Model Summary of Random Forest

Crossing Preference	Precision	Recall	F1 Score	Support
0	0.87	0.88	0.87	66
1	0.76	0.74	0.75	34

The Table 4.3 shows the performance metrics of model using random forest algorithm. In crossing preference, 0 denotes the preference of zebra crossing and 1 denotes footbridge. Support indicates the number of data of true samples that falls into each class of a target variable. Here 66 pedestrians use zebra crossing and 34 pedestrans use footbridge.

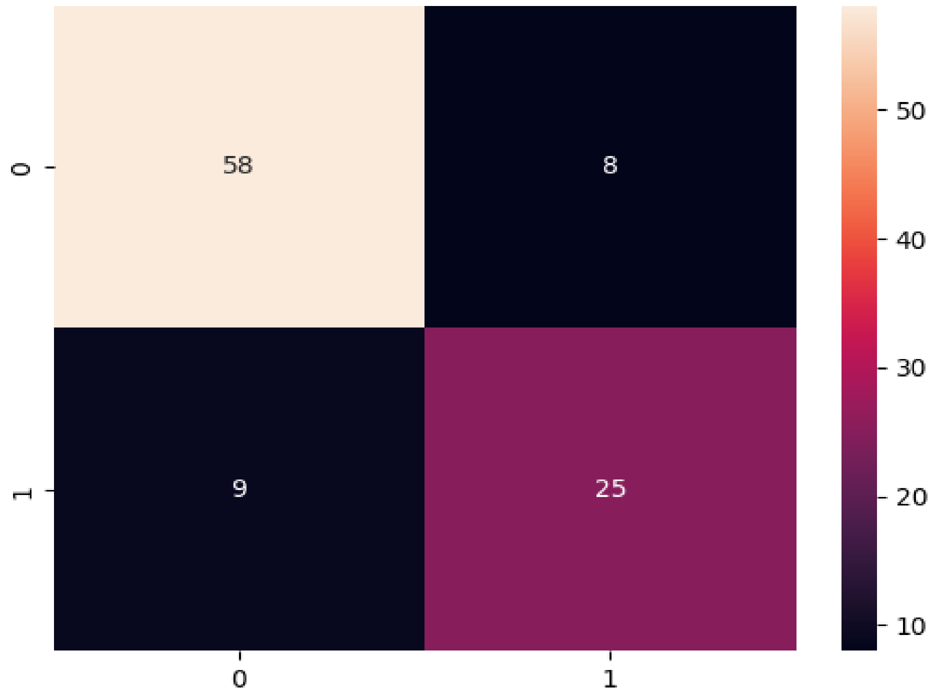

Figure 4.6 Confusion Matrix of Random Forest

Out of 66 zebra crossing users, the model correctly predicted 58 zebra crossing users and wrongly predicted 8 data as footbridge users. From 34 footbridge users the model correctly predicted 25 footbridge users and misclassified 9 data. It is shown in Figure 4.6.

4.5.2 Decision Tree

Using decision tree algorithm, the accuracy obtained from training data is 73% and for testing data it is 78%.

Table 4.4 Model Summary of Decision Tree

Crossing Preference	Precision	Recall	F1 Score	Support
0	0.81	0.84	0.82	61
1	0.73	0.69	0.71	39

The Table 4.4 shows the performance metrics of model using decision tree algorithm. Here 61 pedestrians use zebra crossing and 39 pedestrans use footbridge.

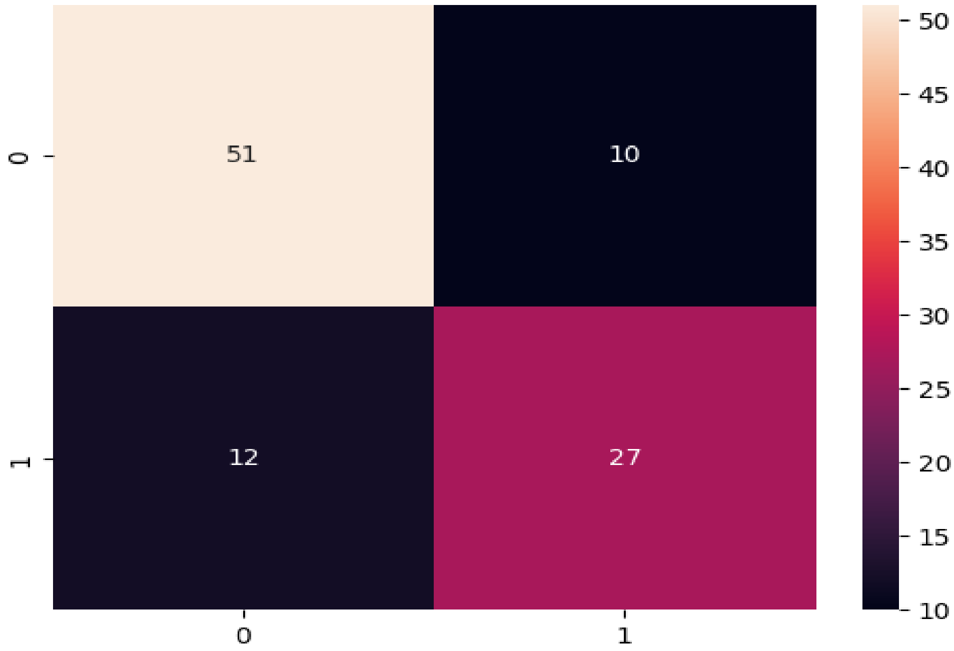

Figure 4.7 Confusion Matrix of Decision Tree

From the Figure 4.7, out of 61 zebra crossing users, the model correctly predicted 51 zebra crossing users and wrongly predicted 10 data. From 39 footbridge users the model correctly predicted 27 footbridge users and misclassified 12 data.

4.5.3 Support Vector Machine

Using support vector machine algorithm, the accuracy obtained from training data is 68% and for testing data it is 79%.

Table 4.5 Model Summary of Support Vector Machine

Crossing Preference	Precision	Recall	F1 Score	Support
0	0.84	0.82	0.83	62
1	0.72	0.74	0.73	38

The Table 4.5 shows the performance metrics of model using support vector machine algorithm. Here 62 pedestrians use zebra crossing and 38 pedestrans use footbridge.

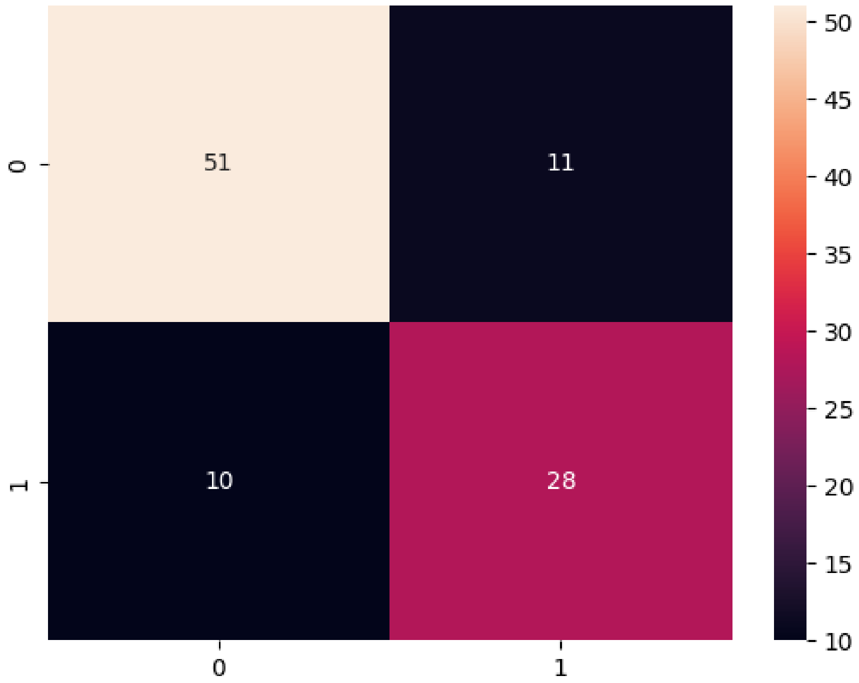

Figure 4.8 Confusion Matrix of Support Vector Machine

From Figure 4.8, out of 62 zebra crossing users, the model correctly predicted 51 people zebra crossing users and wrongly predicted 11 data. From 38 footbridge users the model correctly predicted 28 footbridge users and misclassified 10 data.

4.5.4 K-Nearest Neighbor

Using k-nearest algorithm, the accuracy obtained from training data is 98% and for testing data it is 81%.

Table 4.6 Model Summary of K-Nearest Neighbor

Crossing Preference	**Precision**	**Recall**	**F1 Score**	**Support**
0	0.85	0.84	0.85	63
1	0.74	0.76	0.75	37

The Table 4.6 shows the performance metrics of model using k-nearest neighbor algorithm. Here 63 pedestrians use zebra crossing and 37 pedestrans use footbridge.

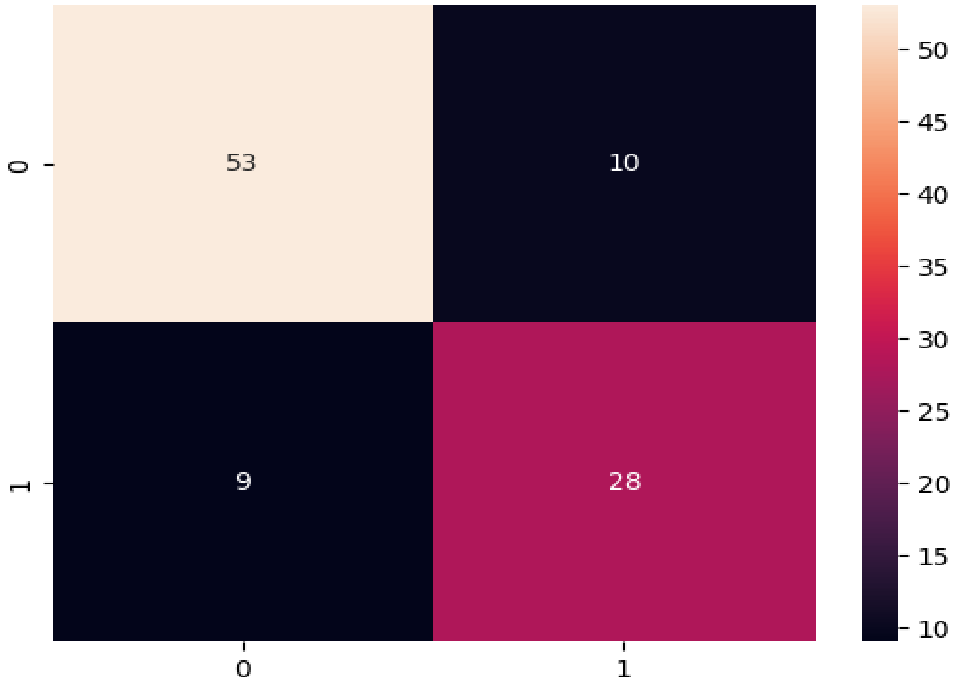

Figure 4.9 Confusion Matrix of K-Nearest Neighbor

Out of 63 zebra crossing users, the model correctly predicted 53 zebra crossing users and wrongly predicted 10 data. From 37 footbridge users the model correctly predicted 28 footbridge users and misclassified 9 data. It is shown in Figure 4.9.

4.5.5 Logistic Regression

Using logistic regression algorithm, the accuracy obtained from training data is 51% and for testing data it is 78%. The accuracy of training data is very low as compared to other algorithms. The accuracy of training and testing data were not similar which means overfitting occurred in this case.

Table 4.7 Model Summary of Logistic Regression

Crossing Preference	Precision	Recall	F1 Score	Support
0	0.85	0.80	0.82	64
1	0.68	0.75	0.71	36

The Table 4.7 shows the performance metrics of model using logistic regression algorithm. Here 64 pedestrians use zebra crossing and 36 pedestrans use footbridge.

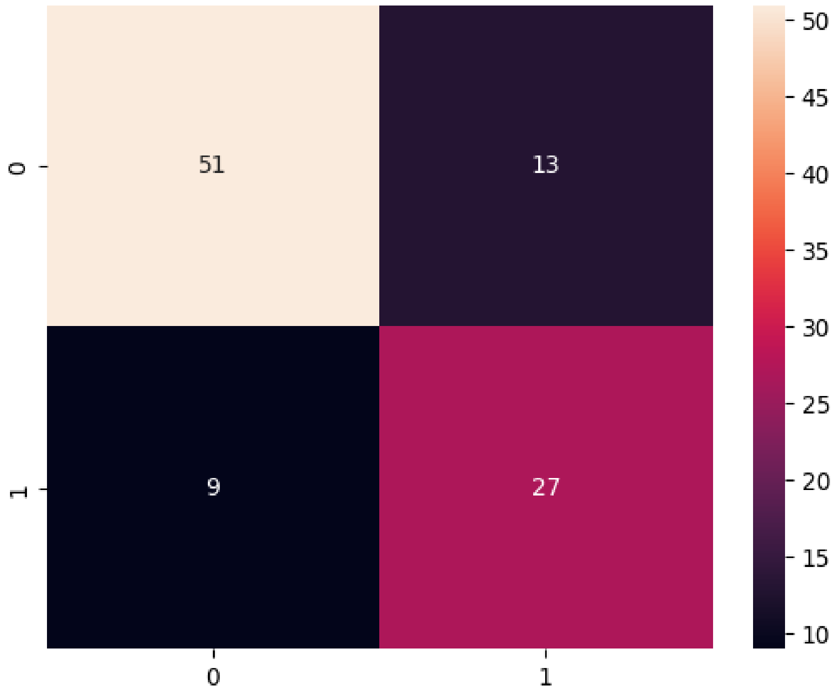

Figure 4.10 Confusion Matrix of Logistic Regression

From the Figure 4.10, out of 64 zebra crossing users, the model correctly predicted 51 zebra crossing users and wrongly predicted 13 data. From 36 footbridge users the model correctly predicted 27 footbridge users and misclassified 9 data.

4.6 Model Comparison

Different models were compared based on their accuracy, precision and recall. A graph is plotted in order to visualize the changes in the performance of different algorithm. Table 4.8 shows the comparison of the performance metrics using different algorithms. It helps to identify the best model using performance metrics.

Table 4.8 Model Comparison

Algorithms	Accuracy of training data (%)	Accuracy of testing data (%)	Precision (%)	Recall (%)
Random Forest	84	83	81	81
K-nearest Neighbour	98	81	80	80
Support Vector Machine	68	79	78	78
Decision Tree	73	78	77	76
Logistic Regression	51	75	76	77

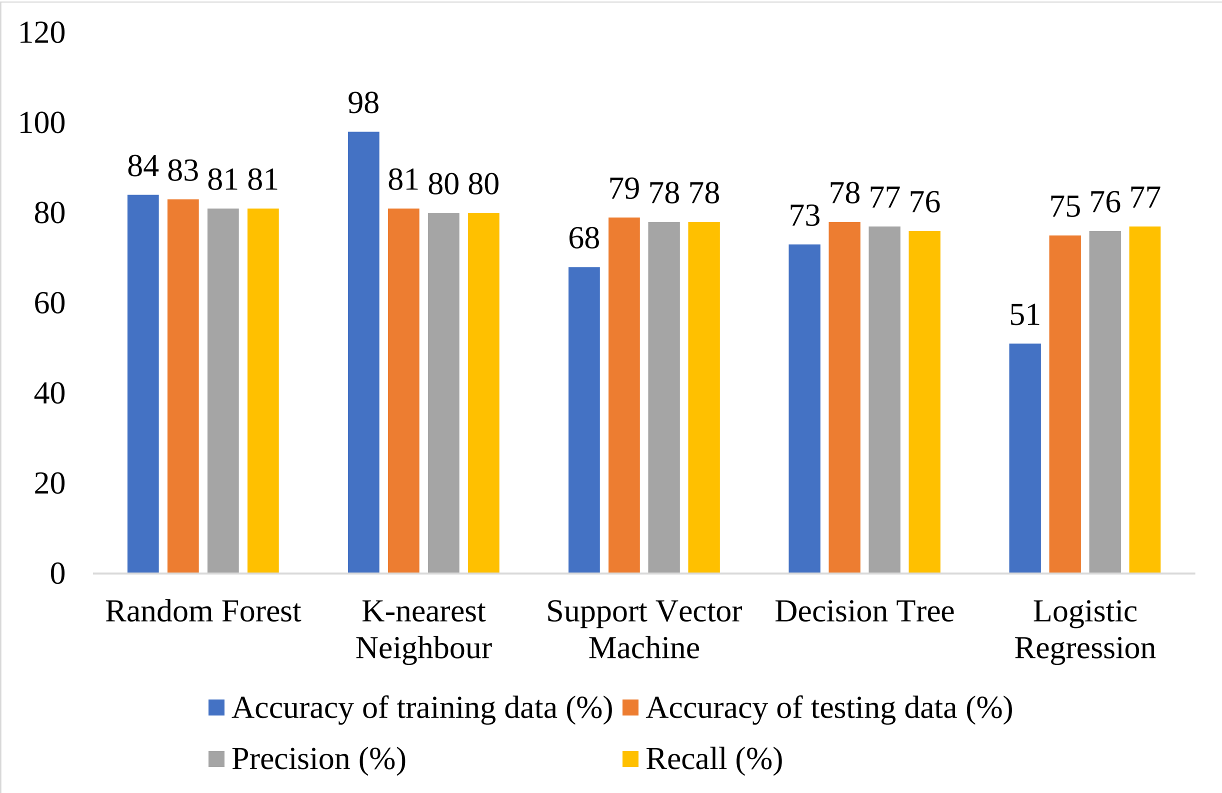

Figure 4.11 Bar chart of Model Comparison

From the above Figure 4.11, it can be concluded that random forest outperformed all other model with an accuracy of 83% followed by k-nearest neighbor, support vector machine, decision tree and logistic regression. The least accuracy is obtained from logistic regression.

The weightage of variables is calculated using random forest since it is obtained as the best model. Table 4.9 shows the weightage of factors in decreasing order.

Table 4.9 Weightage of Factors

Weightage	**Factors**
FBUN	0.22
FBE	0.17
FBST	0.09
Weather	0.07
FBT	0.06
Night	0.05
FBBI	0.05
HT	0.04
FBHF	0.04
UL	0.04
FBA	0.04
Hurry	0.03
FBGI	0.03
Education	0.03
Age	0.02
PAH	0.01
OS	0.01

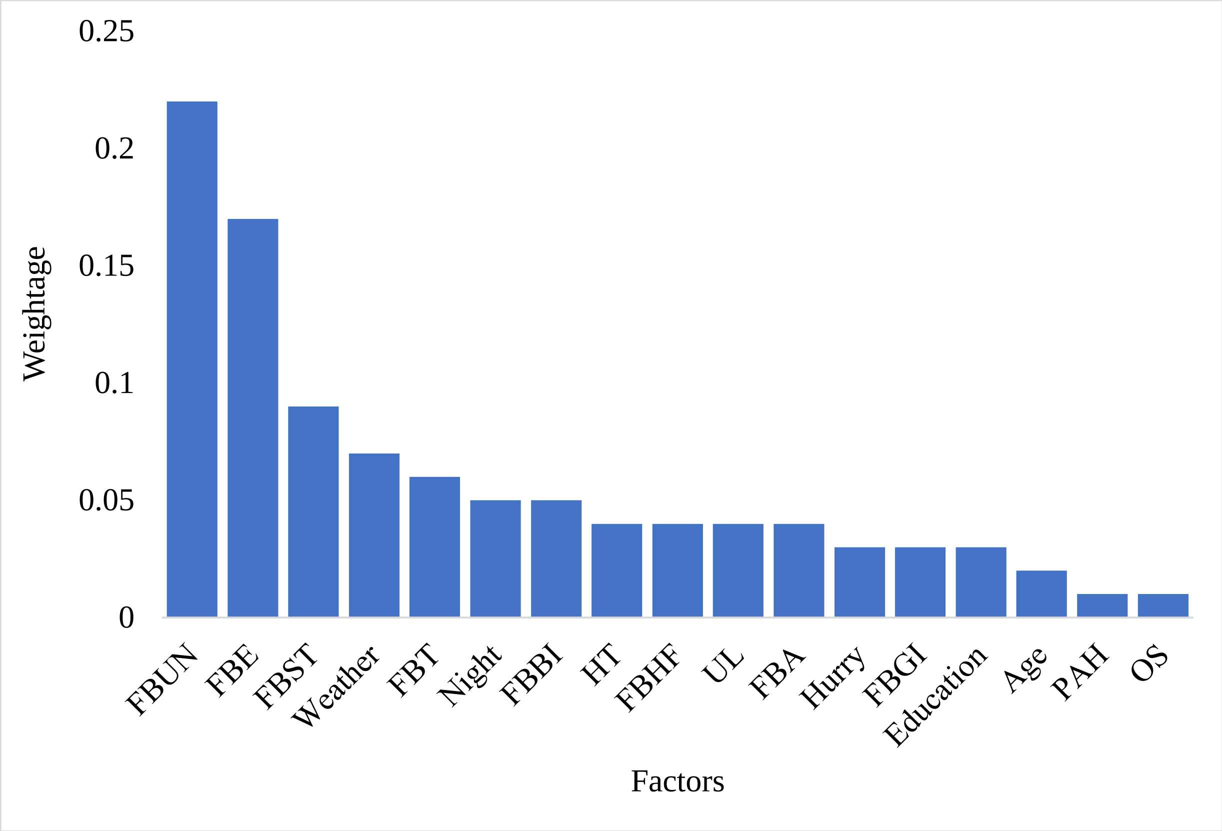

Figure 4.12 Weightage of Factors over Footbridge

The Figure 4.12 indicates the individual weightage of factors influencing the preference of footbridge. The frequency of footbridge usage stated as never shows a high weightage among other factors. It is followed by easiness, stressful, weather, tiresome, night, bad infrastructure, heavy traffic, height fear, unfamiliar location, accessibility, hurry, good illumination, education, age, pedestrian accident history, occupation as student. The sum of individual weightage is 1.

4.7 Improvements

Some counter measures were suggested to the non-users of footbridge to know their willingness to use footbridge if the below mentioned facilities are provided. Likert type scale questions were asked to the pedestrians.

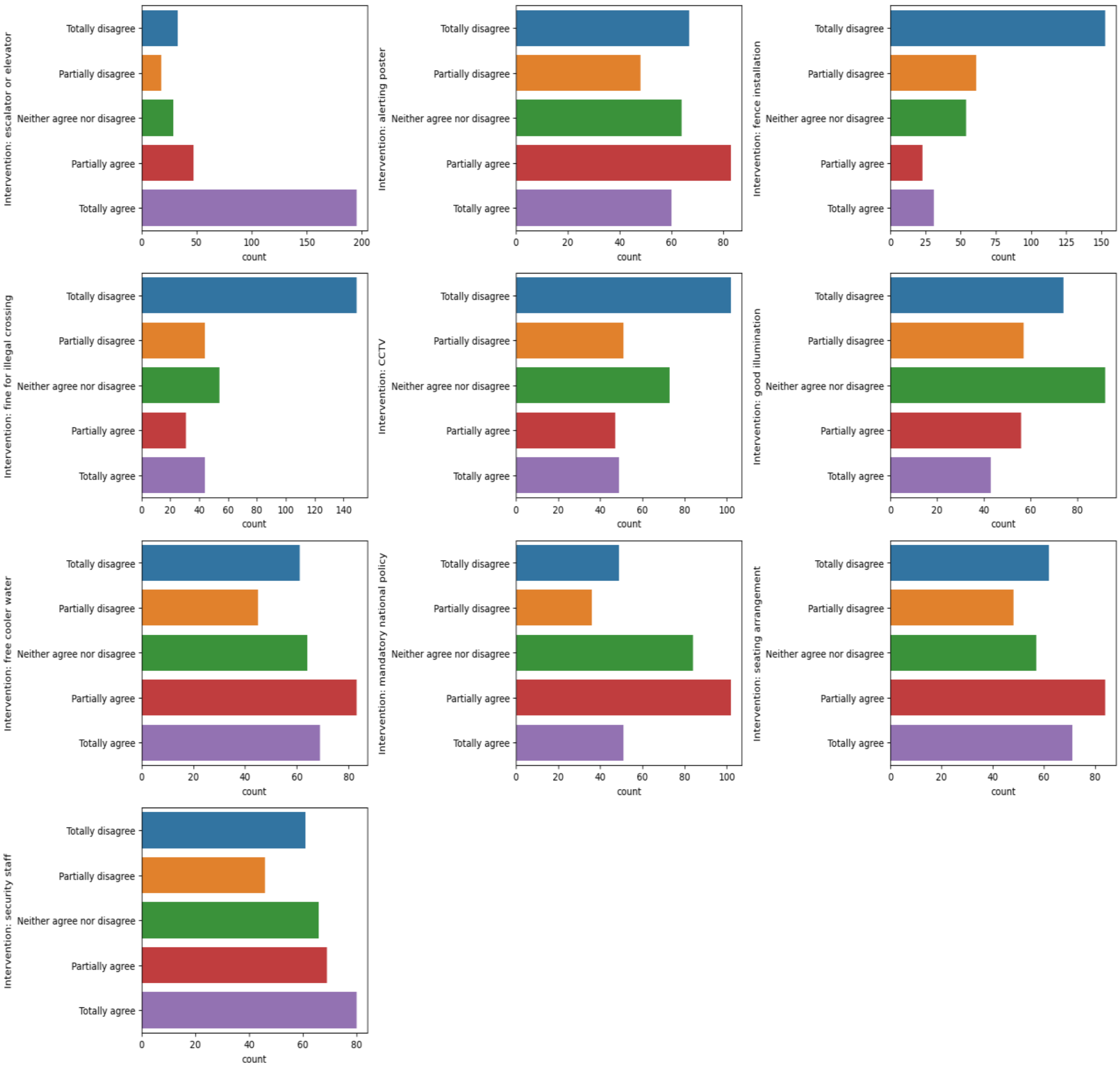

Figure 4.13 Count plot of Counter Measures

Following counter measures were suggested in order to analyze the change in usage of footbridge.

- Escalator, elevator
- Alerting poster
- Fence installation
- Fine for illegal crossing
- CCTV
- Good illumination
- Free cooler water
- Mandatory national policy

- Seating arrangement
- Security staff

From the count plot as shown in Figure 4.13, it is clear that pedestrians are willing to use footbridge if escalator or elevators are provided. In the modelling we already seen that pedestrians are not willing to use footbridge if it makes them feel tired. So, if we provide this facility, it actually makes them less effort and easy to use. Most of the pedestrians are agreeing to use footbridge if alerting poster are provided instead of advertising billboards. Fence installation and fine for illegal crossing does not make any sense. CCTV and good illumination does not seem to attract pedestrians, as we have already seen from the count plot of perception of pedestrians that they are not willing to use footbridge during night. Free cooler water, mandatory national policy, seating arrangement, security staff shows the willingness of pedestrians to use footbridge to an extent. Even though escalator or elevators is attracted to them mostly as compared to other facilities.

4.8 Summary

From the analysis it was found that random forest outperformed all other model with an accuracy of 83%. Improvements were suggested in order to know the willingness of pedestrian to use footbridge. It was found that most of the pedestrians are willing to use footbridge if escalator or elevator is provided. Escalator shows more positive response than all other counter measures.

CHAPTER 5

CONCLUSIONS

5.1 General

The safety of pedestrians must be a prime concern in all matters. For the safety of pedestrian's grade separated crossing facilities are provided but they are not widely accepted by pedestrians. The preferred crossing facility for most of the pedestrians is zebra crossing. Age plays a significant role in choosing available crossing facilities. It was found that as age increases the preference of footbridge decreases. Thus, schools can play a crucial part in making the students the frequent footbridge users. If using footbridge becomes their habitual action then they use it regularly by their own choice. The choice of making alternative solutions depends upon the situations. Most of the pedestrians are willing to use footbridge on extreme sunny or rainy day. They will not use footbridge if it makes them feel tired and stressful due to the physical effort needed in climbing up and down the stairs. Delay has an important role in choosing the alternative facilities. Pedestrians will not use the facility which consumes more time. Also, most of the pedestrians are willing to use footbridge if escalators or elevators are provided. This will reduce the physical effort and makes them feel easy and less time consuming. Based on the safety perception more pedestrians agreeing that appointing security staffs makes them feel safe while using the footbridge.

Machine learning techniques such as random forest, decision tree, support vector machine, k-nearest neighbor and logistic regression were compared to get the best model. The weightage of factors was computed using random forest since which is obtained as the best model. The results of these model can give information to the policy makers and planners regarding the factors depromoting the usage of footbridge. Make sure that the entry and exit of footbridge is easily accessible to the pedestrians, it should be constructed near bus stops for its efficient usage. By improving the existing facilities and by providing escalators or elevators we can attract more pedestrians to use the footbridge. This helps to reduce the pedestrian accidents while crossing through at-grade crossing facilities. For that the government should provide campaigns and awareness programs to the people regarding the importance of footbridge and the pedestrians crashes occurring while crossing through at grade crossing facilities. Thus, we need to make them frequent footbridge users to ensure the safety of pedestrians as well as free mobility of traffic for that the facilities should satisfy their needs.

5.2 Specific Conclusion

The conclusion made from the project are as follows:

- Random Forest outperformed other models with an accuracy of 83% in terms of classifying the test data into zebra crossing and footbridge users
- The weightage of factors was computed using Random Forest
- It was found that frequency of using footbridge has a high weightage among other factors
- It was followed by easiness, stressful, weather, tiresome, night, bad infrastructure, heavy traffic, height fear, unfamiliar location, accessibility, hurry, good illumination, education, age, pedestrian accident history and occupation
- As age increases the preference of footbridge decreases
- School students mostly use footbridge as compared to other age groups because schools give instructions to the students to use footbridge
- Pedestrians are willing to use footbridge if escalator or elevator is provided

5.3 Future Scope

- The current study focused on similar land use type. Further the study can be extended by adding more footbridge locations of different land use type
- Advanced machine learning techniques can be used for modelling

REFERENCES

1. Al Bargi, W. A., & Daniel, B. D. (2020). Modelling Pedestrians' utilization of crossing facilities along urban streets. *Case studies on transport policy*, *8*(2), 593-598.

2. Anciaes, P. R., & Jones, P. (2018). Estimating preferences for different types of pedestrian crossing facilities. *Transportation research part F: traffic psychology and behaviour*, *52*, 222-237.

3. Asaithambi, G., Kuttan, M. O., & Chandra, S. (2016). Pedestrian road crossing behavior under mixed traffic conditions: a comparative study of an intersection before and after implementing control measures. *Transportation in developing economies*, *2*(2), 1-12.

4. Avinash, C., Gore, N., Shriniwas, A., Gaurang, J., & Manoranjan, P. (2020). Choice crossing behaviour model for Safety Margin of pedestrian at mid-blocks in India. *Transportation research procedia*, 48, 2329-2342.

5. Banerjee, A., & Maurya, A. K. (2020). A comparative study of pedestrian movement behavior over foot over bridges under similar land-use type. *Transportation research procedia,* 48, 3342-3354.

6. Banerjee, A., Raoniar, R., & Maurya, A. K. (2020). Pedestrian overpass utilization modeling based on mobility friction, safety and security, and connectivity using machine learning techniques. *Soft Computing*, *24*, 17467-17493.

7. Bhatia, S. Y., Patil, G. R., Chopadekar, A. D., & Sharma, S. (2022). Skywalk facility utilization and factors influencing its use: A case study of Mumbai Metropolitan Region, India. *Case Studies on Transport Policy, 10*(1), 396-405

8. Brosseau, M., Zangenehpour, S., Saunier, N., & Miranda-Moreno, L. (2013). The impact of waiting time and other factors on dangerous pedestrian crossings and violations at signalized intersections: A case study in Montreal. *Transportation researchpart F: traffic psychology and behaviour*, 21, 159-172.

9. Chandrappa, A. K., Bhattacharyya, K., & Maitra, B. (2021). Measures for improving pedestrian crossing facilities based on perceptions of urban commuters: An experience in Kolkata. *Case studies on transport policy*, *9*(2), 965-973.

10. Demiroz, Y. I., Onelcin, P., & Alver, Y. A. L. Ç. I. N. (2015). Illegal road crossing behavior of pedestrians at overpass locations: Factors affecting gap acceptance, crossing times and overpass use. *Accident Analysis & Prevention*, *80*, 220-228.

11. Ferenchak, N. N. (2016). Pedestrian age and gender in relation to crossing behavior at midblock crossings in India. *Journal of Traffic and Transportation Engineering (English Edition)*, *3*(4), 345-351.

12. Hasan, R., & Napiah, M. (2018). The perception of Malaysian pedestrians toward the use of footbridges. *Traffic injury prevention*, *19*(3), 292-297.

13. Hasan, R., Oviedo-Trespalacios, O., & Napiah, M. (2020). An intercept study of footbridge users and non-users in Malaysia. *Transportation research part F: traffic psychology and behaviour*, *73*, 66-79.

14. Jamal, A., Zahid, M., Tauhidur Rahman, M., Al-Ahmadi, H. M., Almoshaogeh, M., Farooq, D., & Ahmad, M. (2021). Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study. *International journal of injury control and safety promotion*, *28*(4), 408-427.

15. Landa-Blanco, M., & Ávila, J. (2020). Factors related to the use of pedestrian bridges in university students of Honduras. *Transportation research part F: traffic psychology and behaviour*, *71*, 220-228.

16. Ojo, T. K., Appiah, A. B., Obiri-Yeboah, A., Adebanji, A. O., Donkor, P., & Mock, C. (2022). An intercept survey of the use and non-use of footbridges in ghana. *Case Studies on Transport Policy*,10,1581-1590

17. Oviedo-Trespalacios, O., & Scott-Parker, B. (2017). Footbridge usage in high-traffic flow highways: The intersection of safety and security in pedestrian decision-making . *Transportation research part F: traffic psychology and behaviour*, *49*, 177-187.

18. Patra, M., Perumal, V., & Rao, K. K. (2020). Modelling the effects of risk factor and time savings on pedestrians' choice of crossing facilities at signalised intersections. *Case studies on transport policy*, 8(2), 460-470.

19. Prajapati, M., Advani, M., & Parida, P. (2018). Identifying the most suitable road crossing infrastructure based on estimated delay to pedestrians and vehicle users. *Transportation letters*, 10(2), 113-127.

20. Rankavat, S., & Tiwari, G. (2016). Pedestrians perceptions for utilization of pedestrian facilities–Delhi, India. *Transportation research part F: traffic psychology and behaviour*, *42*, 495-499.

21. Rasanen, M., Lajunen, T., Alticafarbay, F., & Aydin, C. (2007). Pedestrian self-reports of factors influencing the use of pedestrian bridges. *Accident analysis & prevention*, *39*(5), 969-973.

22. Ravishankar, K. V. R., & Nair, P. M. (2018). Pedestrian risk analysis at uncontrolled midblock and unsignalised intersections. *Journal of traffic and transportation engineering (English edition)*, *5*(2), 137-147.

23. Sinclair, M., & Zuidgeest, M. (2016). Investigations into pedestrian crossing choices on Cape Town freeways. *Transportation research part F: traffic psychology and behaviour*, *42*, 479-494.

24. Truong, L. T., Nguyen, H. T., Nguyen, H. D., & Vu, H. V. (2019). Pedestrian overpass use and its relationships with digital and social distractions, and overpass characteristics. *Accident Analysis & Prevention*, *131*, 234-238.

25. Wu, Y., Lu, J., Chen, H., & Wu, L. (2014). Identification of contributing factors to pedestrian overpass selection. *Journal of traffic and transportation engineering (English edition)*, *1*(6), 415-423.

APPENDIX 1

Dear respondent,

I am a student of TKM college of Engineering Kollam, currently pursuing MTech Transportation. I am conducting a questionnaire survey for my project work on perception of pedestrians towards the use of footbridge. I will be thankful to you if you spend your precious time in filling this questionnaire. Your response will be kept confidential and only be used for academic purpose/Research works.

PART A

Socio-Economic Characteristics

1. Age

    a) <18 years
    b) 18- 25 years
    c) 26-50 years
    d) >50 years

2. Gender

    a) Male
    b) Female
    c) Others

3. Education

    a) Upto 10th standard
    b) Upto 12th standard
    c) Graduate
    d) Masters/above

Choice

1. Which one you prefer while crossing

    a) Footbridge
    b) Zebra Crossing

PART B

Perception regarding footbridge usage were asked among footbridge users and non-users

1. Frequency of using footbridge
    a) Never
    b) Sometimes
    c) Always

2. Past involvement in a traffic accident as a pedestrian
    a) Yes
    b) No

3. Do you prefer to use footbridge on following situations. Give your score out of 5 in which 1 indicate poor and 5 indicate good.
    a) Being in a hurry
    b) At night
    c) On extreme rainy or sunny day
    d) Unfamiliar location
    e) Heavy traffic

4. Perception of footbridge are rated based on the likert scale- Totally Disagree, Partially disagree, Neither agree nor Disagree, Partially agree, Totally Agree.
    a) Footbridges are easier to use than to cross the street
    b) Using footbridge is safe
    c) Using footbridge is tiresome
    d) Footbridge in my city have a bad infrastructure
    e) Footbridge have a good illumination
    f) Using footbridge is stressful
    g) Fear towards height while using footbridge
    h) Is footbridge comfort for physically disabled pedestrians
    i) Free from obstructions like seller
    j) Entry and exit accessible
    k) Free from stray dogs
    l) Crossing the street is stressful

m) Crossing the street is faster than using footbridge

5. Any suggestions

PART C

Counter Measures – only for footbridge non-users

1. Following are some of the interventions aimed at increasing footbridge usage. Will the proposed interventions encourage you to use the footbridge. These are rated based on the likert scale- Totally Disagree, Partially Disagree, Neither Agree nor Disagree, Partially Agree, Totally Agree.

a) Escalator or Elevator
b) Alerting poster
c) Fence installation
d) Fine for illegal crossing
e) CCTV
f) Good illumination
g) Free cooler water
h) Mandatory national policy
i) Seating arrangement
j) Security staff

2. Any Suggestions