Predicting Reliability of Response in Online Surveys

K. Nadarajah

2018

Predicting Reliability of Response in Online Surveys

A dissertation submitted for the Degree of Master of Computer Science

K. Nadarajah
University of Colombo School of Computing
2018

Declaration

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name: Kamalini Nadarajah

Registration Number: 2014/mcs/050

Index Number: 14440506

Signature: Date:

This is to certify that this thesis is based on the work of

Ms. Kamalini Nadarajah

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by:

Supervisor Name: Dr. Rasika Dayarathna

Signature: Date:

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SD	Standard Deviation Method
IQD	Interquartile Deviation Method
MAD	Median Absolute Deviation Method
MD	Mahalanobis Distance Method
SVM	Support Vector Machine
CFS	Correlation Feature Selection
CPM	Character Per Minute

Abstract

The use of online surveys is exponentially increasing day by day. From being less time consuming, to being easy to use, online surveys have become highly advantageous. However, the unreliability of participants' response has become a growing concern. A tool has been implemented, in the attempt of producing a detection mechanism to eliminate unreliable users' responses on merely the basis of their behaviour, while filling the online surveys such as time taken to answer the questions, clicks, excessive clicking, longer inactivity, changes on already given answers, time taken to answer open ended questions, changes on the screen and activation and changes of form elements like radio buttons, checkboxes and drop downs.

All the attributes considered are influential to an extent. Total answer clicks, number of responded questions, checkbox changes during the survey and the status of idle are the most influential attributes in identifying the reliable response of online surveys. The algorithms used, namely, Naïve Bayes, Logistic Regression, Decision Tree, Support Vector Machine and Random Forest are also quite reasonable considering that the accuracy for all of them were above 50%. The most influential algorithm was Logistic Regression.

Acknowledgements

Firstly, I would like to convey my sincere thanks to Dr. Rasika Dayarathna, for being an excellent guide who has been there to support and assist me throughout this phase. Thank you for the immense support in making this research a success.

I would also like to express my gratitude to the project coordinator Dr. Kasun Gunawardana, for his guidance and support over this time period. I am grateful to the undergraduate students of the University of Colombo School of Computing who participated in the survey and enabled me to thereby gather the data needed.

I would like to take this opportunity to extend my thanks to the Department of Statistics, especially Mr. Anajana Bandara and Ms. A.A.A.S.A. Keerthi, who helped with the analysis.

My sincere token of thanks to the feedback and comments from the examiners Dr. A R Weerasinghe and Dr. (Mrs.) M D J S Goonetillake who helped me to advance my research. All the senior lecturers, lecturers, instructors and staffs of University of Colombo School of Computing are appreciated for their valuable guidance.

A special thanks to my parents, brother, collegues and friends especially Luckshika, Nishanthy and Gomathy for their moral support throughout this journey.

1 Introduction

A survey is a method of collecting data from a targeted audience. Surveys are conducted in order to collect accurate data. Nowadays online surveys have become a viable tool for data collection resulting in most of the researches using it as their main data collection method. Online surveys are usually created as web forms where the users are given a set of questions to be answered online. Users choose their preferred devices including computers, tabs, iPads and mobile phones to participate in the survey. This enables them to have flexibility in accessing the survey. Online surveys are quicker, efficient, cost-minimizing methods of collecting data that are used by business models for research and development.

A few examples of well known online surveys software are LimeSurvey, SurveyMonkey, SurveyGizmo, LimeSurvey and PollDaddy. They are used to create online surveys which are distributed among the product/service users in order to conduct market research or to obtain their feedback and opinions. Surveys are also circulated among employees to get their feedback and comments on company's policies. Surveys are also frequently used to gather the opinion of concerned or parties of a generalized social concern. According to [13], the online survey software market is valued at US$4.065 billion in 2017 and is expected to grow by 11.25% to reach US$6.929 by the year 2022. This shows that the business models and other authorities have started using the data collected through online surveys for various purpose.

In online and offline surveys, reliability of user response is an important factor. This usually refers to the involvement or response of the users in answering the questions. Jiali Ye [14] stated that when researchers use an online survey, they can enjoy a number of benefits. However, they should also be prepared to face several challenges rising as a result of unsatisfactory response rates, sampling, and lack of control over the data collection environment. Possibilities of predicticing realiable user response in online survery is more efficient than offline surveys.

1.1 Statement of the Problem

The user response in the surveys can be categorized into reliable response and unreliable response based on user's behaviour. Reliable responses are the ones where user takes the survey seriously and gives the apt answers to the questions. The user puts their time and effort to provide the responses. Unreliable response is where the user does not pay enough attention to answer the questions. They don't take enough time to read or understand the questions. Instead, they merely submit the survey without realizing the impact of their participation in it. It is difficult to get the accurate survey results when a large number of participants do not provide genuine responses.

1.2 Motivation

In online and offline surveys, inattentive and careless responses are major concerns. Due to these unreliable responses by the users, the people who conduct the survey will not be able derive at the right conclusions.

Therefore, realiable responses should be identified and taken into consideration and unreliable responses should be identified and eliminated. In this way online survey data can be used in an effective manner.

1.3 Objectives

The main objective of this research is to predict the reliability of responses in online surveys.

The following are some secondary objectives which need to be fulfilled in order to achieve the primary objective of predicting the reliability of responses in online surveys.
a) Capture the behavioural data while filling online surveys.
b) Create a tool to capture behavioural data such as time taken to answer the questions, clicks, mouse movements, excessive clicking, longer inactivity, changes on already given answers, time taken to answer open ended questions, changes on the screen and activation and changes of form elements like radio buttons, checkboxes and drop downs.

1.4 Scope and Limitations

The users' behavioural data will be used to predict the reliability of user response in online surveys.

A browser based tool is used to capture and process behavioural data.

The limitations of the scope identified are as follows.

a) Due to a slow internet connection or system failures, users may take more time than required to finish the surveys. In such situations, accurate time cannot be calculated.

b) Some users give intentionally incorrect answers even though the question was read and understood. The detection of unreliable response due to intentional incorrect answers cannot be identified via behavioural data.

The thesis is structured as follows: Chapter 2, discusses the related work done to predict the reliability of responses by the users and the methods that are used to identify it. Chapter 3, describes how the data collection is done, the types of behavioural data collected, the participants, the finalized attributes for the dataset for this research and the data analysis methods to predict the reliability of responses while filling online surveys. Chapter 4, describes the results obtained using behavioural data that was taken from the analysis and evaluation methods for the proposed research. Chapter 5, describes the conclusion and future work for the proposed research.

2 Literature Review

2.1 Chapter Introduction

This chapter briefs the work done related to “Predicting the reliability of user response in filling online surveys”. In addition, it also gives an insight to the methods used in identifying the reliability of user responses.

In any survey, careless responses are a concern [9]. Hence, some researches have been done to identify the careless responses in online survey data.

Careless response, in general, is the cause of a lack of attention while answering to a question in a test, questionnaire or survey. This might not necessarily be intentional but has a definite negative impact on the test analyses that follows [16].

It is also known as inattentive, suboptimal, non-serious or insufficient effort. This attitude might result in low quality data. These low quality data can be excluded from further analyses to get better survey result by identifying the careless responses in survey data.

2.2 Factors Causing Unreliable Responses in Surveys

The fundamental requirement for accurate data in surveys is a motivated group of participants who are genuinely interested in attempting the survey. Psychologically, disinterested members are the most prominent contributors to careless responses. Distractions in the surrounding environment of a participant, irrespective of whether they were intended or accidental also play a role in the quality of the data entered by the user. A user who is multitasking or being distracted by external sources will not be able to pay the full attention expected while filling the survey. This results in a significant fall of the quality of data retrieved from the user.

Lengthy surveys or surveys with way too many details might also be another reason for careless responses since these traits might make the user quite impatient or bored after a certain point. The choice of participants for a survey in a particular domain is also very important. The participants should be able to understand the survey questions and respond in accordance. If the participants are new to the idea of surveys, there are very less chances that they could enter sensible answers.

Last but not least, the cause of poor quality data might unfortunately turn out to be the intentional behaviour of the user himself. This might be out of disinterest as well as unappreciable inner motives of the user.

There are several methods used to identify careless responses. While some of them are based on the content of the survey itself, others depend on the mere behaviour of the user which isn't content specific.

2.3 Content Dependent Methods to Identify Carelessness

2.3.1 Self-Reported Items

In this detection method, the user as a part of the survey is asked questions regarding the topic of concern which is mostly the objective of the survey itself. The conclusion on whether the survey has attained the purpose is purely derived from the user input.

This is a very economic measure to identify careless respondents where they directly asked the respondents to indicate the seriousness of their responses [10].

For example, "I carefully considered each item before responding". If they genuinely accept that their levels of involvement are insufficient, then their responses are not advised to be considered as part of the survey. The most fundamental way to validate if a user has put a decent effort on a survey is to simply ask him or her the same [5].

2.3.2 Instructed Items

As the name suggests, in this method the user who is attending the survey is given instructions to one or more questions within or before the survey. The fact that the user follows the given instructions reveals that the user is serious in filling the survey. The higher the number of instructed items, the higher the accuracy of the validation according to [5]. For example, it could be mentioned "Please leave this question blank". In this case, if the user , in contrary to the instruction, responds to the question it is likely that he hasn't seriously went through it.

2.3.3 Bogus Items

This is a method where a number of obvious truths and evidently fake statements are included as part of the survey. These statements should have a single correct answer which cannot be manipulated [9].

For example a statement like “I have only one nose” can never be disagreed upon. If a user has a varying opinion on such evident statements, it can be considered that the user might have participated in the survey in a careless manner. The advantage of such bogus items is that if the user chooses an incorrect response, there is little doubt that he is responding carelessly or dishonestly; thus, there is little chance of a false positive [9]. When user gives many incorrect responses for bogus items, obviously it can be stated that the user participated in the survey with no seriousness.

Because of the nature of self-reported, instructed and bogus items, respondents are likely to be aware of the purpose of these inserted items. This knowledge may motivate attentive respondents to avoid answering in an undesirable manner [5]. Although, there are chances to demotivate attentive respondents too. Mainly surveys are conducted to collect users’ opinions. Therefore responding to such items can make the user distracted from the survey and the user might also feel a sense of discomfort by these types of items that are added purposely to test them.

2.3.4 Semantic Synonyms and Semantic Antonyms

Semantic synonym method makes use of the fact that similar items are answered similarly. Any contrast in the way similar items are answered is considered inconsistent and thereby unreliable. An example of a semantically synonymous pair would be “I like reading” and “I read books”. The assumption here is that the user will not change his or her mind within the same survey a number of times [5].

Semantic antonym method, being similar to the semantic synonyms aims at finding users who provide like answers to unrelated items. This is certainly considered to be inconsistent and proves that the user did not invest the time and effort expected. “I am twenty years old” and “I have been working for twenty years” are two items that cannot be affirmative [5].

2.3.5 Psychometric Synonyms and Psychometric Antonyms

Psychometric synonyms are identified by analysing the inter item correlation matrix [5]. Item pairs with a correlation of above a certain threshold were defined as psychometric synonyms. Firstly, items that had a major correlation were categorized together. Next, each user's responses were taken into consideration. The way they have answered the psychometrically synonymous questions were evaluated and the responses of users with lower scores were ideally unreliable [9].

Psychometric antonyms which follow the same approach as the psychometric synonyms, differ by the fact that this method considers the items with the largest negative correlation amongst them as reliable responses.

For both psychometric synonyms and antonyms, the first step of this process is to identify the item pairs. This can be done both semantically and psychometrically. Next, the vectors of the responses of the first and second items of each set should be correlated. For semantic or psychometric synonyms, this computed screening index should be high. As for psychometric and semantic antonyms, this value should be a negative value.

2.3.6 Changes in the Input Fields

Though alteration of an option during a survey is quite normal, excessive changes with higher iterations are a definite concern. The alterations made after providing an initial answer to response-retrieval fields like the text fields, text areas, checkboxes, radio buttons and dropdown menus in a survey were taken into consideration throughout a session. Moreover, there was a difference noted between response alterations for factual questions and those for which users had to ponder over to derive at an opinion. Constant changes to factual questions resulted in a higher degree of negative influence on the quality of data, since the responses should ideally be a known factor and require minimum or null changes to it. Changes to opinion related questions were handled more leniently since the probability of change and need to rethink is fairly realistic in this case.

2.4 Content Independent Methods to Identify Carelessness

2.4.1 Response Pattern

This method checks responses of consecutive items. This technique is called as LongString. It considers the number of consecutive items with the same response option chosen by a participant.

The maximum LongString is the most number of times a user has consecutively chosen the same option. The average LongString is the average of total LongString values of an individual submission. A cutoff value is determined for the average and max long string by considering the submission of all users.

A cutoff value for the average and maximum LongString indices was formed based on clear break points in a frequency distribution. Based on the cutoff value, the response is flagged as reliable or unreliable response [9].

The LongString technique relies on the assumption that too many consecutive identical responses may indicate a lack of effort. Also, they recommend that LongString would be most useful when the items are randomly [5]. Like the response time index, the long string index does not have a well-defined cutoff [1].

2.4.2 Personal Reliability

Personal reliability relies on the respondent's consistency while responding to a particular survey, rather than the content. In this method, the survey is primarily split into two sections on the basis of odd and even positions of the questions. Then, each even response is paired with an odd entry sequentially, after which the correlation is calculated. Using the Spearman-Brown formula the results are later adjusted based on a scale. The lower the values, the greater the degree of carelessness [5].

The same method can be implemented taking into count two halves of the same survey instead of the odd and even segregation as well [1].

2.4.3 Non-response Items

This method is based on the reasoning that a reliable user will not skip responding to survey questions, at least not intentionally. Unreliable users are identified by evaluating the ratio between the total numbers of unanswered questions to the total number of survey questions that the user is supposed to attend to in the survey. If the user has answered all the questions or missed very less they are considered to have no negative influence on contributing to low quality data. Such user responses come under the "0" category that can otherwise be named as reliable user responses. On the contrary, if the user has not answered any questions at all or has answered an insignificant number of questions, it is considered to be an unreliable user response with a high influence on impacting the quality of data and fall under the category "2". Furthermore, in case the user has answered a decent number of questions considering the ratio but has missed on some too, then this response is considered as a low influence on the quality of data and is categorized as "1" [12].

Response pattern, personal reliability and no-response items do not require any special items. However, they need an exclusive analysis after the survey is complete [9].

2.4.4 Response Time

In order to use this technique, one must first come up with a sensible threshold for the time an average user should be spending on a question. Though there might be an argument that the speed might differ from person to person and the time taken depends on the type and standard of each question, it is obvious that a hasty careless response can be spotted from a regular average-paced one. It is "unlikely for participants to respond to survey items faster than the rate of 2s per item". This might result in hindrance of the overall accuracy of the data collected. This is factual since every question definitely needs a timeline to read, understand and then respond in accordance [12].

Computer-administered tests offer more precision since built-in timers can be used to measure the amount of time spent on a survey and work on a micro-analysis level based on the time for each question as well [5].

Response time can be calculated on an entire questionnaire or page-by-page, with the latter being more useful when attempting to identify sporadic or local random [1].

2.4.5 Paradata

Paradata, a vital part of modern survey research methodology, refers to data that is captured based on the behaviour established during the survey. The collection mostly happens via a browser-based facility. The browser based method can further be classified into: First, Collection by installing a particular program in the survey supporting computer (Webtracker) and second using the script languages (Web VIP). Using script languages is preferred since installing a program can have a negative impact both in terms of time consumption and a lack of technical awareness.

As per the research by Stieger and Reips [12], paradata has been retrieved and used to analyze the actions performed by the participants during the process of filling an online questionnaire.

a) Longer Inactivity

Longer inactivity while attempting a survey need not always be proportional to the contribution of low quality data. It might have even occurred due to a mere physical interruption or technical malfunction. However, there are chances that it was due to lack of interest too. Hence, this is considered only as a low influence on the quality of data and the threshold for inactivity will be predefined [12].

b) Excessive Clicking

According to [12], excessive clicking was defined as twice as many or more clicks than necessary used during the process. The users who have, clicked beyond twice the necessary total clicks and significant unnecessary scrolling were considered highly influential to the low quality data entry.

c) Excessive Mouse Movements

This method is based on the user behaviour when moving their mouse around the survey area. This is calculated on the basis of the overall length of the mouse track. By using standard deviation, the outliers of the result are considered to be high negative influences on the quality of data [12].

Based on the above mentioned efforts in this area, a further progress is planned by using the behavioral data collected at the time of the survey. An exclusive tool is to be used for the above mentioned enhancement.

Following are the areas of hypothesis analysed based on the outcome of this research:

H1. Deriving at the relevant attributes that have an effect on the online survey

H2. Most influential attributes in identifying the reliability of the responses in online surveys

H3. Moderately influential attributes in identifying the reliability of responses in online surveys

H4. Least influential attributes in identifying the unreliable responses in online surveys

H5. Classification amongst reliable and unreliable responses

H6. Most suitable algorithm to predict the reliability of responses in online surveys

H7. The attribute set that has the highest accuracy of impact amongst all the algorithms in identifying the reliability of responses in online surveys

2.5 Chapter Summary

The above chapter summarizes the detection methods used to identify the carelessness while filling online surveys and the basis on which they are built. This will function as a knowledge base on which, further development will be made and discussed in the following chapters.

3 Methodology

3.1 Chapter Introduction

This chapter explains the data collecting and data analyzing methods which are used to predict the reliability of user responses in online surveys. In this research, the behavioural data detects how users behave while filling online surveys.

3.2 Data Collection

In this research, the data collection plays an important role since behavioural data needs to be gathered in order to proceed. Basically, behavioural data includes significant characteristics of paradata. According to [12], paradata is auxiliary data which is collected during the process of data collection such as mouse clicks and response times.

A tool to capture behavioural data was not available online. The browser based tool created by Stefan Stieger & Ulf-Dietrich Reips was requested for since it was mentioned that it was available for research purposes. However, it was informed that it was outdated and was not compatible with any of the present browsers.

Hence, for this research, a tool was created and inserted into the survey in order to capture the users' behaviour. With the help of this above mentioned tool, the behavioural data is collected while users fill their surveys and it is used as the data for the proposed research analysis.

3.2.1 Online Survey Selection

There are some free online survey tools available. Some of them are LimeSurvey, SurveyMonkey, SurveyGizmo, Google Forms, etc. Although some are free online surveys, there are some restrictions such as number of respondents or the number of questions. By considering all the features, LimeSurvey was selected for the experiment since it allows the survey administrators to quickly create intuitive, powerful, online question and answer surveys that can work for tens to thousands of participants with less effort. To create an online survey for this research, LimeSurvey was installed on the server.

3.2.2 Survey Design

This survey has a collection of questions based on the topic, "Sri Lanka". The topic and the questions contained within are quite lenient, since the participants should be able to respond to all the questions with no major complications. There are 30 questions added to the survey. Four pages are in the survey including the welcome page. 30 questions are included within three pages. Equal number of questions are added in every page. Each page contains 10 questions. All types of questions are included namely, text area, text field, radio button, likert scale and checkbox type questions. The types of the questions are also of the same quantity in each page. Mostly, radio button type questions are added in order to make it convenient for the participants.

Survey Link: http://knreviews.com/limesurvey/index.php/212876?v=new

The questions screens are added in Appendix A.

3.2.3 Tool Creation and Capturing Features

The tool was created using JavaScript. JavaScript is a client-side powerful scripting language. The script of this tool is included into the survey page as an external JavaScript file. The created tool has the full compatibility to work with Lime survey. However, it can also work with other online surveys provided that minor code-specific modifications are done. This primarily includes changes to the script with regards to the relevant HTML elements. This tool works in all the standard browsers such as Microsoft Edge, Internet Explorer, Firefox, Chrome, Safari and Opera.

Following is an elaboration of how the tool works to capture the features.

Once the user clicks on a survey link, the first page of the survey loads and the tool starts running. The user who fills the survey will not be aware of or feel, that there is a tool running within the survey. The tool uses a Local Storage to store the captured features with the JSON structure. In this Local Storage, the data is stored across browser sessions. Local Storage is similar to Session Storage, except that while data stored in Local Storage has no expiration time. JSON is an open-standard file format that uses human-readable text to transmit data objects consisting of attribute–value pairs and array data types.

Once the first page of the survey is loaded, the tool checks whether there is a local storage with the defined name in the browser. If not, the tool adds a new local storage with the defined name in the browser and adds a main storage object to assign the data. If so, the tool will delete the local storage, add a new local storage and add a main storage object as mentioned.

Once local storage is created, an empty object and an empty array is added in the main storage object. The object, “userDetails” is used to add the user’s basic details such as screen size and browser details. The array called “pageDetails” is used to add user’s behavioural details and survey related details in each survey page such as page timings, clicking, scrolling etc.,. Once the first page is loaded, the user’s basic details are captured and added to “userDetails” and this will be added only once throughout the survey. Figure 1 shows the JSON structure for the user’s basic details and the page details:

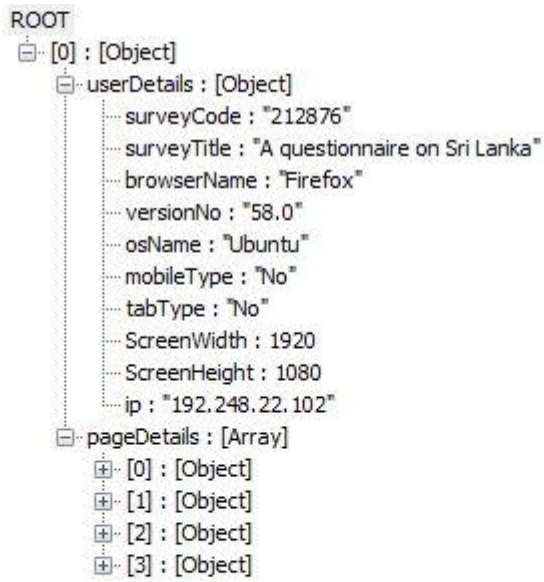


Figure 1: Capturing features with the help of the tool

Inside the “pageDetails” array, user’s behavioural details and survey related details are added page wise. When user visits a page, a separate object is added inside the “pageDetails” array. Note that this depends on the number of visits rather than the number of pages. Page details with all the actions performed by the user is added inside the object for the page. Under each object, arrays are added to capture the details such as scrolling, clicking, page idle details, page away details and answering selection. These actions can occur several times while the user stays in a particular page. Figure 2 depicts an expanded version of the “pageDetail” array:

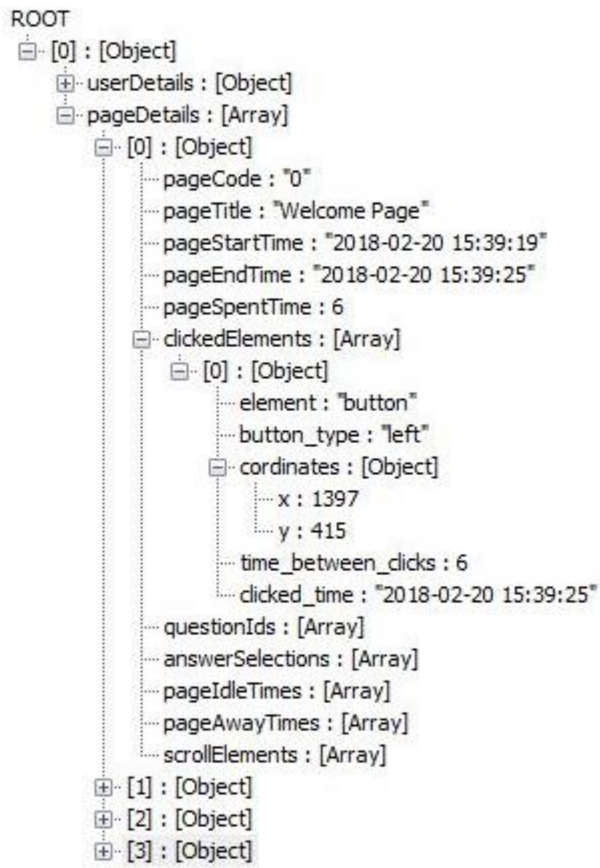


Figure 2: Page wise captured features

When user clicks on next button to move to the next page, the main storage object details are added to the existing Local storage. Likewise, when user moves to every page the local storage is updated in order to add the new details of the “pageDetails” array. Once user is done with the survey by clicking on the “submit” button, the Local storage data which is in the JSON format is passed through AJAX call to the server side. Then, the data is received by the server side and a random ID will be generated for each submission. Then the content which is received is written to the text file with the randomly generated ID in the server. The content of the text file will be in the JSON format. For each participant’s survey submission a separate text file gets created.

Capturing features by the tool, can be categorized into three, namely basic details, survey related details and behavioural details.

Table 1 reflects the basic details that can be captured by the tool.

Feature	Description
Browser Details	Name and version of the browser
Device Details	The type of the device (Mobile or desktop)
Operating System Details	Name and version of the Operating System
Screen Size	The width and height of the screen

Table 1: Basic details captured via the tool

Table 2 shows the survey related details that can be captured by the tool.

Feature	Description
Survey Details	The title and code of the survey.
Page Details	ID and title of the page
Question Details	Question IDs will be captured page wise.
Answer Details	When user types or chooses an answer the following details are captured. a) Question ID b) Question type such as text field, text area, radio, checkbox or dropdown c) Answer for the question d) Element ID for the answer e) Pasted Text – The texts which are pasted from some other area.

Table 2: Survey related details captured

The code for the tool is given in Appendix B.

Table 3, contains features which are the behavioural details which can be captured by the tool.

Feature	**Description**
Page Timings	The loading time and ending time of each page are captured. a) Page starting time – This time is calculated once the page is initially loaded b) Page ending time – The time, user moves to the next page by clicking “Next Button” or submits the survey by clicking the “submit button”
Clicked Elements	When user clicks on any area of the page, the following details are captured. a) X and Y coordinates of the page b) The element types such as button, div, label etc. c) The types of clicked button such as left, middle or right d) Clicked time e) The time difference between the current click and previous click
Idle Times	When user stays on the survey page without doing any activity for a long period of time the following detail is captured. To capture this, the minimum threshold is defined from which the idle count begins. a) The idle start time – The time the user reaches the defined idle time b) The idle end time – The time the user starts to do an activity on the survey page from the idle period
Away Time	When the user leaves the page when opening a new window or tab, the following details are captured. a) Away start time - The time the user is away from the survey page b) Away end time - The point of time the user again enters the survey page
Scrolling	When user scrolls and when it reaches the top or bottom, the following details are captured. a) Reached time - The time in which the scrollbar reaches the top or the bottom

	b) Reached type – This identifies whether the user has reached the extreme top or the extreme bottom
Answer selections	This feature is captured with the use of the input elements. When a page loads and user starts to select or type an answer for a question in a particular page that specific time will be taken as the answering start time for the particular question. When the user selects or starts to type another question or same question, that time is considered as the answering ending time for the previous question and the starting time for the current selection. Likewise all starting and ending times for every question are defined. For the text area type responses such as text area or text field, ending time is the time taken till the last character is entered. The following details are captured when the user answers the questions. a) Answered starting time b) Answered ending time c) X and Y coordinates of the page – the position the user keeps the cursor in to answer. e) Pasted count – The number of times the user pastes in the text area. f) Pasted character count – The number of characters pasted in the text area. g) Characters per minute – It is calculated according to the typing speed of the user.

Table 3: Behavioural data captured by the tool

3.2.4 Tool Validation

The features of the capturing tool are validated by installing a few open source software to validate the accuracy of the tool. There are several software used to capture the features. When a user starts to fill the online survey, the installed software should be run on the machine. At the same time screen video capturing is also taken. Once user submits the survey, the text file created for that attempt is taken and the details of that file is compared with the records that are taken by the open source software. Features recorded by the created tool are quite precise.

The validation details are included in the Appendix C.

3.2.5 Survey Participants

Undergraduate students from the University of Colombo were selected to participate in the survey. There were around 130 students participating in the survey. The students were given the access to the survey via a link and briefed upon the purpose of their participation. The students were instructed to fill the survey along with an instruction to specify an on-the-spot category that included filling the survey in a relaible manner or unreliable manner. Each participant’s submission details were saved as a separate text file in the server with the help of the tool created.

3.2.6 Survey Procedure

In this research, behavioural data is used for the analysis. Even though the answers for the survey are not necessary, the responses are collected as well for the sake of the fore coming evaluation processes. Students were devided in two groups. One group was asked to fill the survey with reliable response and other group was asked to fill the survey with unreliable response.

As part of the evaluation process, the outcomes from the analysis based on the behavioural features captured and the disclaimer of the users at the beginning of the surveys are compared. Moreover, the number of correct responses are also considered.

3.2.7 Dataset Preparation

Once the survey is complete; all the submissions are available in the form of a text file in the server. The data has to be converted to a common format for the analysis. Firstly, the data in each text file which is in the JSON format needs to be extracted. For example, each page's start time and end time is saved in the text file, but for the analysis the average time spent for a page is needed. Likewise, there are several calculations that need to be done to prepare the dataset for the analysis.

To make the calculations easier, the text file data is planned to be stored in the database tables. In order to achieve this, a small system was developed using PHP language and MySQL. The tables were selected to create, based on the captured features. The generated text files were uploaded to the system and each text file's data was read in order to add the details into the database. The data of each text file was retrieved and assigned to an array by using the JSON decode functions. The elements of the array (text file data) were added to the MySQL tables based on the features. According to the finalized attributes of the dataset, the calculations were done to the data with the help of MySQL Queries and PHP functions. The records with finalized attributes were listed in the html table format after which they were downloaded to the CSV format. Again, the records were pre-processed to make the final dataset. As a result of pre-processing which includes, eliminating incomplete records, the records further tuned to 120 out of 125 records. The derived final dataset contains 120 records with 31 attributes for the proposed research analysis.

Dataset preparing link: http://knreviews.com/project/submission-summary-list.php

System details are given in Appendix E and F.

Table 4 explains the selected attributes and descriptions for the final dataset. All the time related attributes are given in seconds.

ID	**Attribute Name**	**Attribute Description**
A1	Average time spent per question	The average time spent for each question by each user. (Spent time for all questions / No of questions)
A2	Total left button clicks	The total left button clicks of all the pages by each user

A3	Total right button clicks	The total left button clicks of all the pages by each user.
A4	Total middle button clicks	The total middle button clicks of all the pages by each user.
A5	Total answer clicks	The total clicks which are used to select or type the answers by each user.
A6	Total non-answer clicks	The total clicks which are not used to select or type the answers by each user.
A7	Total clicks	The total clicks of all the pages by each user.
A8	Average time spent between clicks	The average time spent between each click of all pages by each user.
A9	Number of time scrollbar reached top	The number of times the scroll bar reached the top of all the pages by each user.
A10	Number of time scrollbar reached bottom	The number of times the scroll bar reached the bottom of all the pages by each user.
A11	Page visited count	The number of times a page was visited by each user.
A12	Pasted text count	The number of times the text is pasted by each user.
A13	Pasted character count	The character count which are done through paste option by each user
A14	Average character per minute (cpm) for typing	The average time taken to type the characters on a text field or text area per attempt by each user. (Average cpm for typing = Total cpm for all typing attempts / no of typing attempts)
A15	The total time spent for out of page	The total time spent when user stays out of the survey page. This includes opening a new window or doing other activities.
A16	The total idle time	The total time participant stays in a page without doing any activity. If user inactivity lasts for more than 5 minutes it will be taken as idle time.

A17	Average time spent per page	The average time spent for each page by each user. (Average time spent per page=Total time spent for the survey / number of pages)
A18	Answer changes count	The number of times the answer is changed by each user.
A19	No of time changes on text areas	The numbers of times each user changes a text area answer.
A20	No of time changes on drop downs	The numbers of times each user changes a dropdown answer.
A21	No of time changes on text fields	The numbers of times each user changes a text field answer.
A22	No of time changes on radio buttons	The numbers of times each user changes a radio button answer.
A23	No of time changes on check boxes	The number of time user changes on check box answers by each user.
A24	Average time spent on text area questions	The average time spent on text area questions by each user. (Spent time for all text area questions / No of text area questions)
A25	Average time spent on drop down questions	The average time spent on drop down questions by each user. (Spent time for all drop down questions / No of drop down questions)
A26	Average time spent on radio button questions	The average time spent on radio button questions by each user. (Spent time for all radio button questions / No of radio button questions)
A27	Average time spent on text field questions	The average time spent on text field questions by each user. (Spent time for all text field questions / No of text field questions)

A28	Average time spent on check box questions	The average time spent on radio button questions by each user. (Spent time for all check box questions / No of check box questions)
A29	Average time taken between attempts on questions	The average time spent between attempts on questions by each user. (Spent time between attempts / No of attempts on questions)
A30	Number of response questions	The number of responsive questions in the survey.
A31	User type	Type of user (Instructed category to fill the survey: Reliable response:1, Unreliable response:0)

Table 4: Attributes gathered from the tool

3.3 Data Analysis

The behavioural data obtained is the key factor for this analysis. The data is analysed based on the selected 120 records with 31 attributes by using machine learning algorithms and certain statistical methods.

3.3.1 Tool Selection for the Analysis

For this research, R, Weka and SPSS were selected for the analysis. SPSS is used for the descriptive analysis. R and Weka are used for feature selection and user classification.

3.3.2 Descriptive Analysis

This analysis provides an insight on how each chosen attribute behaves. It was conducted by elaborating each attribute with the help of its mean, standard deviation, minimum value, maximum value and a graphical representation.

The graphs were drawn based on the category. (Reliable response and Unreliable response). In the below drawn graphs, A31 is divided into two, namely 1 and 0. (Reliable response – 1, Unreliable response 0).

A1) Average time spent per question

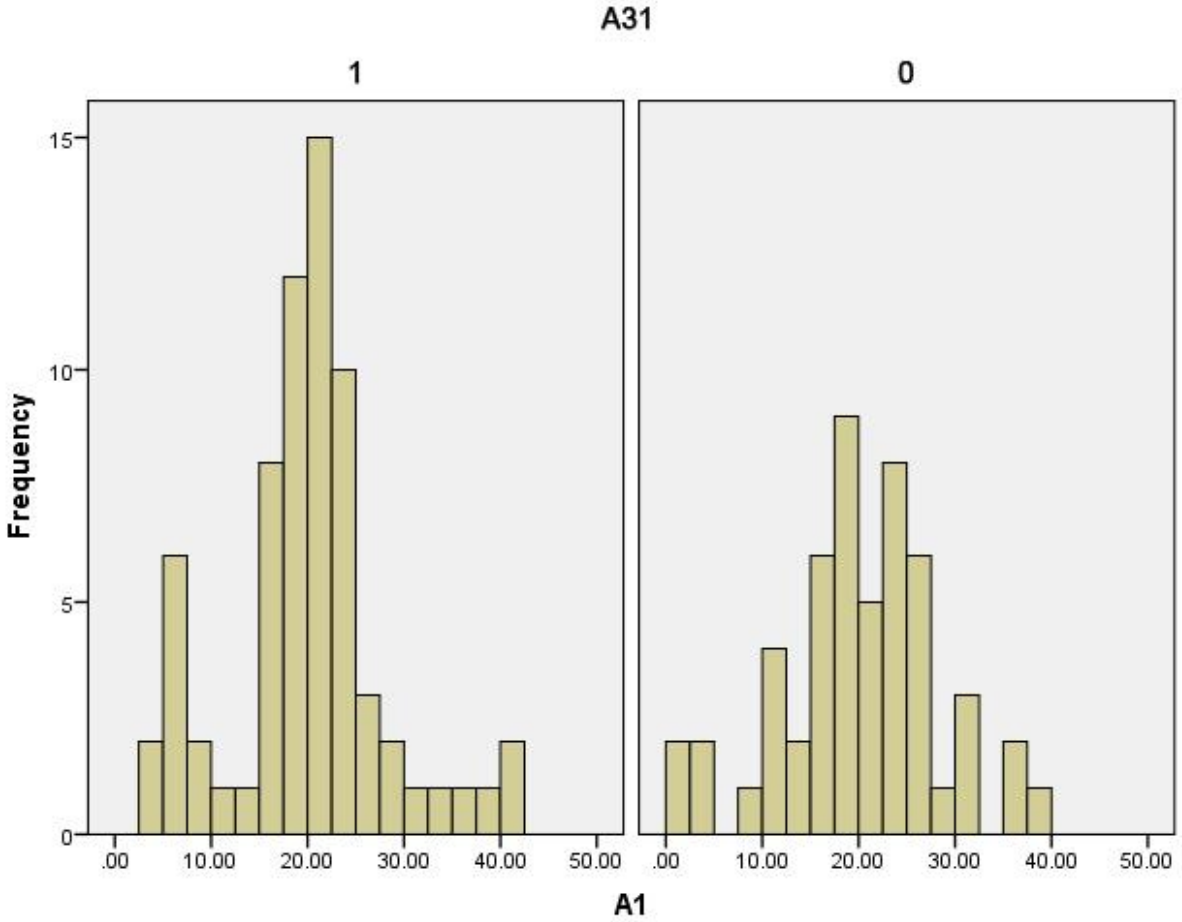


Figure 3: Graph for average time spent per question

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A1	68	3.93	40.73	19.9641	8.09313
	Valid N (listwise)	68				
0	A1	52	1.23	38.37	20.0675	8.40105
	Valid N (listwise)	52				

Table 5: Descriptive statistics for average time spent per question

Comparison:

The mean and standard deviation of the reliable and the unreliable response categories are close to each other for the attribute of average time spent per question. Although, mean and standard deviation of the unreliable response category are higher than reliable response category, it is in a small variation. The minimum average time of unreliable response category is lesser than the minimum average time of reliable response category. The maximum average time of reliable response category is lesser than the maximum average time of unreliable response category. Based on the graph, more values of the reliable response category is close to the mean.

A2) Total left button clicks

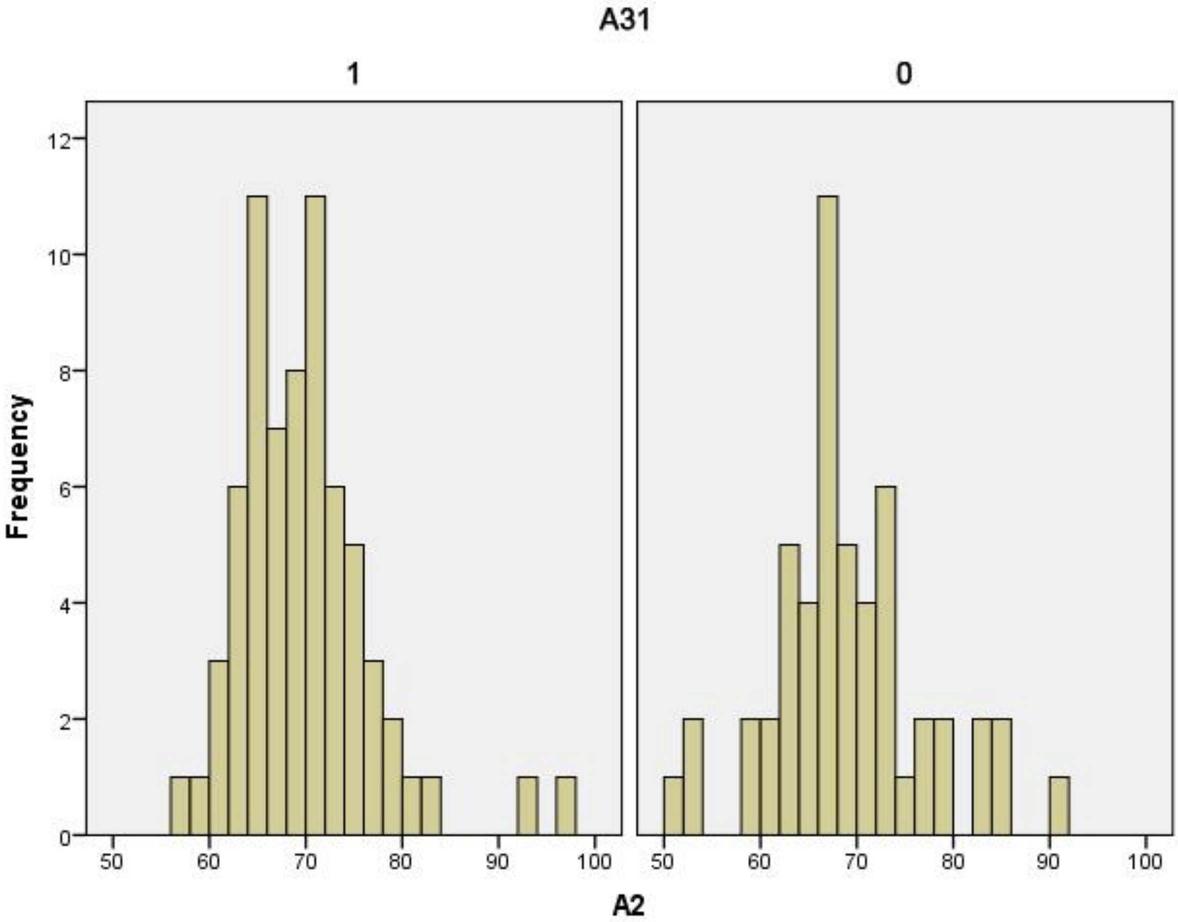


Figure 4: Graph for total left button clicks

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A2	68	56	97	69.31	6.875
	Valid N (listwise)	68				
0	A2	52	50	90	68.42	8.057
	Valid N (listwise)	52				

Table 6 : Descriptive statistics for total left button clicks

Comparison:

The mean of the reliable response and the response categories are close to each other for the attribute of total left button clicks. Both minimum and maximum left clicks of reliable response category are higher than unreliable response category. The standard deviation of the unreliable response category is higher than reliable response category. Based on the graph, much more values of reliable response category is close to the mean.

A3) Total right button clicks

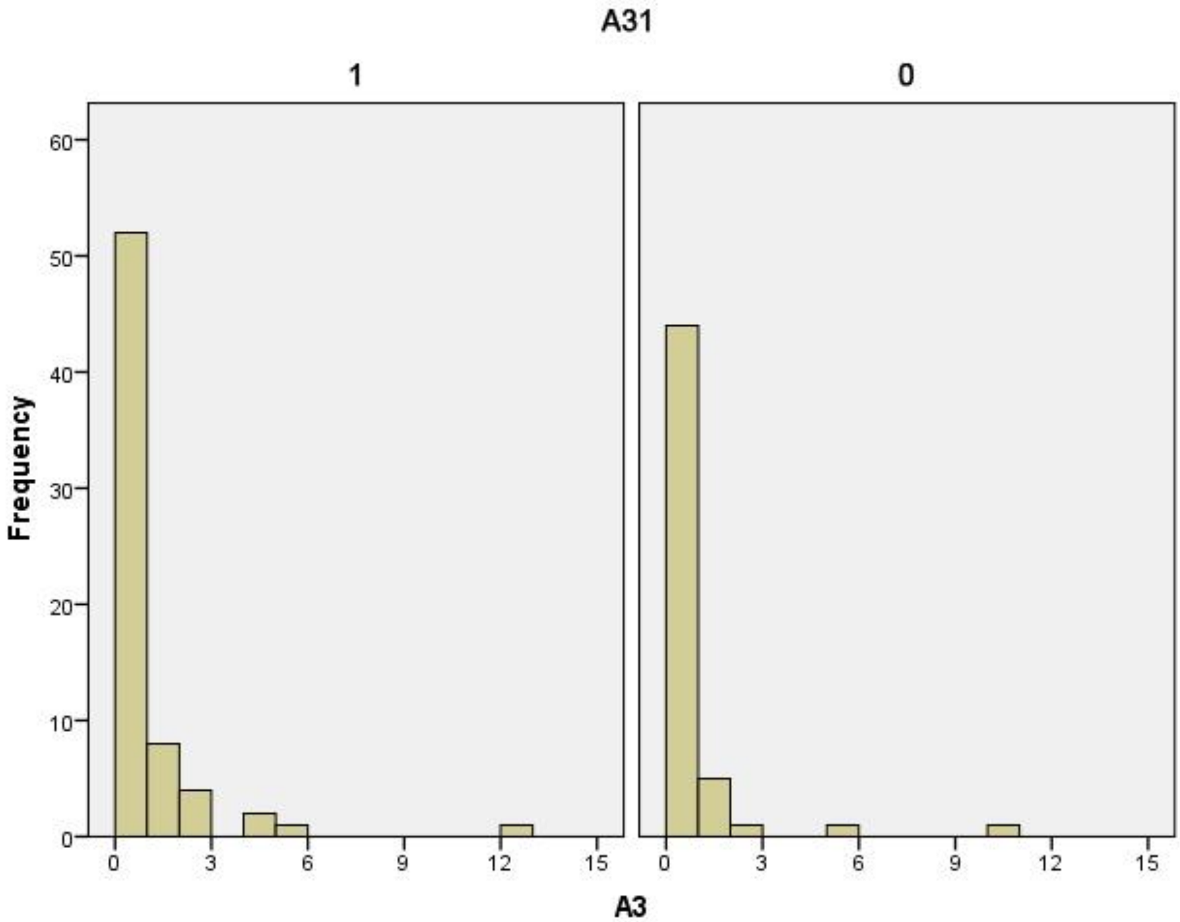


Figure 5: Graph for Total right button clicks

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A3	68	0	12	.60	1.729
	Valid N (listwise)	68				
0	A3	52	0	10	.42	1.564
	Valid N (listwise)	52				

Table 7: Descriptive statistics for total right button clicks

Comparison:

The minimum value for both the categories is zero. According to the graph the majority of both the category users didn't use the right click during the survey. The mean and standard deviation of the reliable response category are higher than the unreliable response category.

A4) The status of middle button click

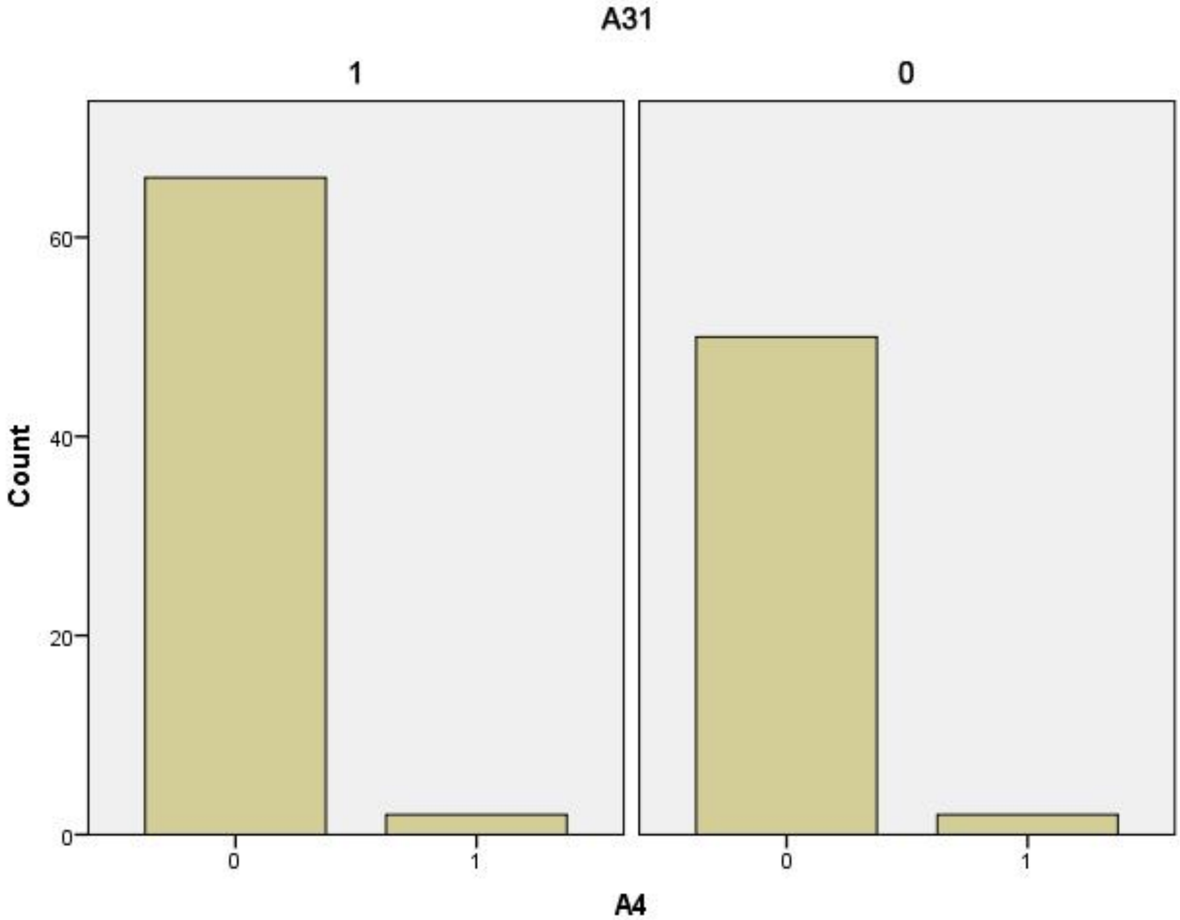


Figure 6: Graph for the status of middle button click

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A4	68	0	1	.03	.170
	Valid N (listwise)	68				
0	A4	52	0	1	.04	.194
	Valid N (listwise)	52				

Table 8: Descriptive statistics for the status of middle button click

Comparison:

According to the graph the majority of both the categories didn't use middle click during the survey. The mean and standard deviation of the unreliable response category are higher than the reliable response category in a small variation.

A5) Total answer clicks

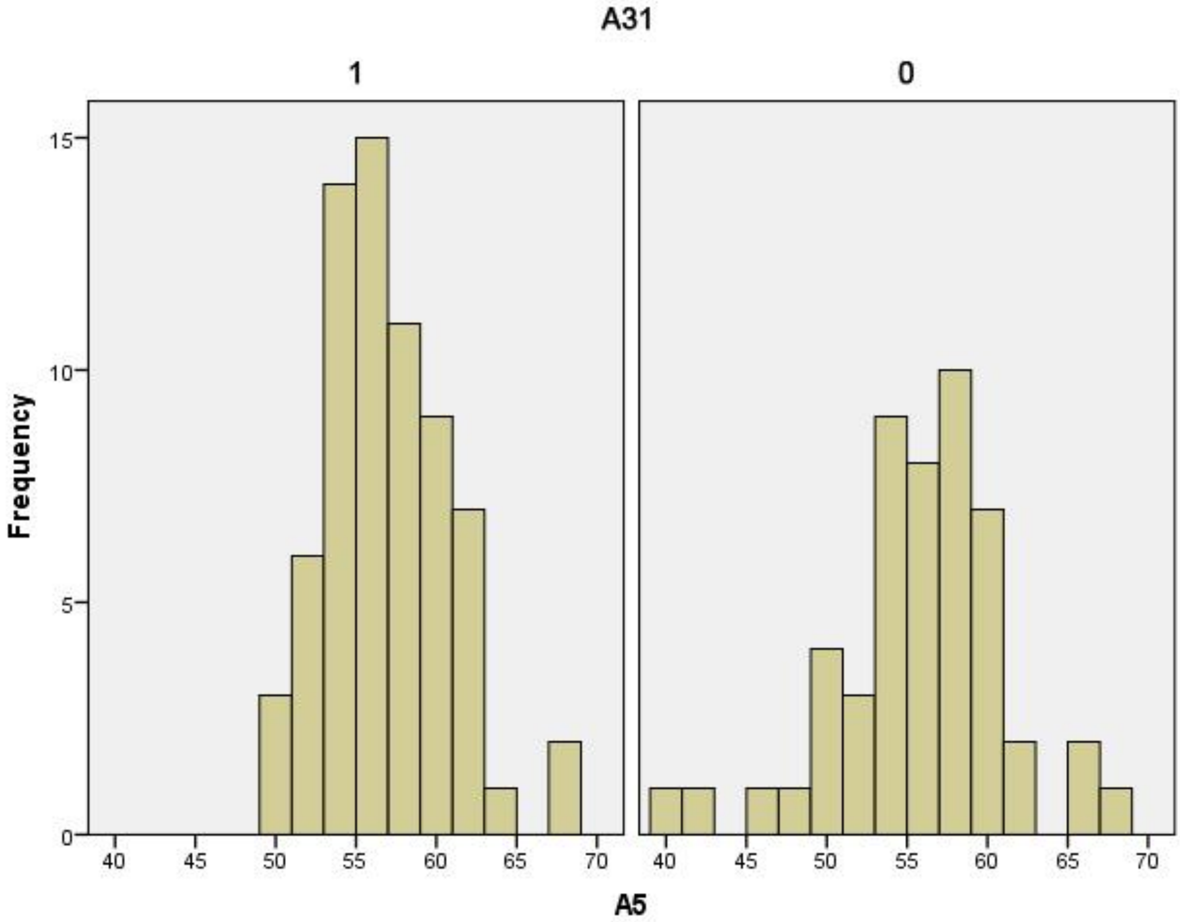


Figure 7: Graph for total answer clicks

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A5	68	50	68	56.49	3.858
	Valid N (listwise)	68				
0	A5	52	20	67	54.23	7.566
	Valid N (listwise)	52				

Table 9: Descriptive statistics for total answer clicks

Comparison:

The standard deviation of unreliable response category is higher than the reliable response category. There is a significant difference between the standard deviation of both categories. Based on the graph, more values of the reliable response category is close to the mean.

A6) Total non-answer clicks

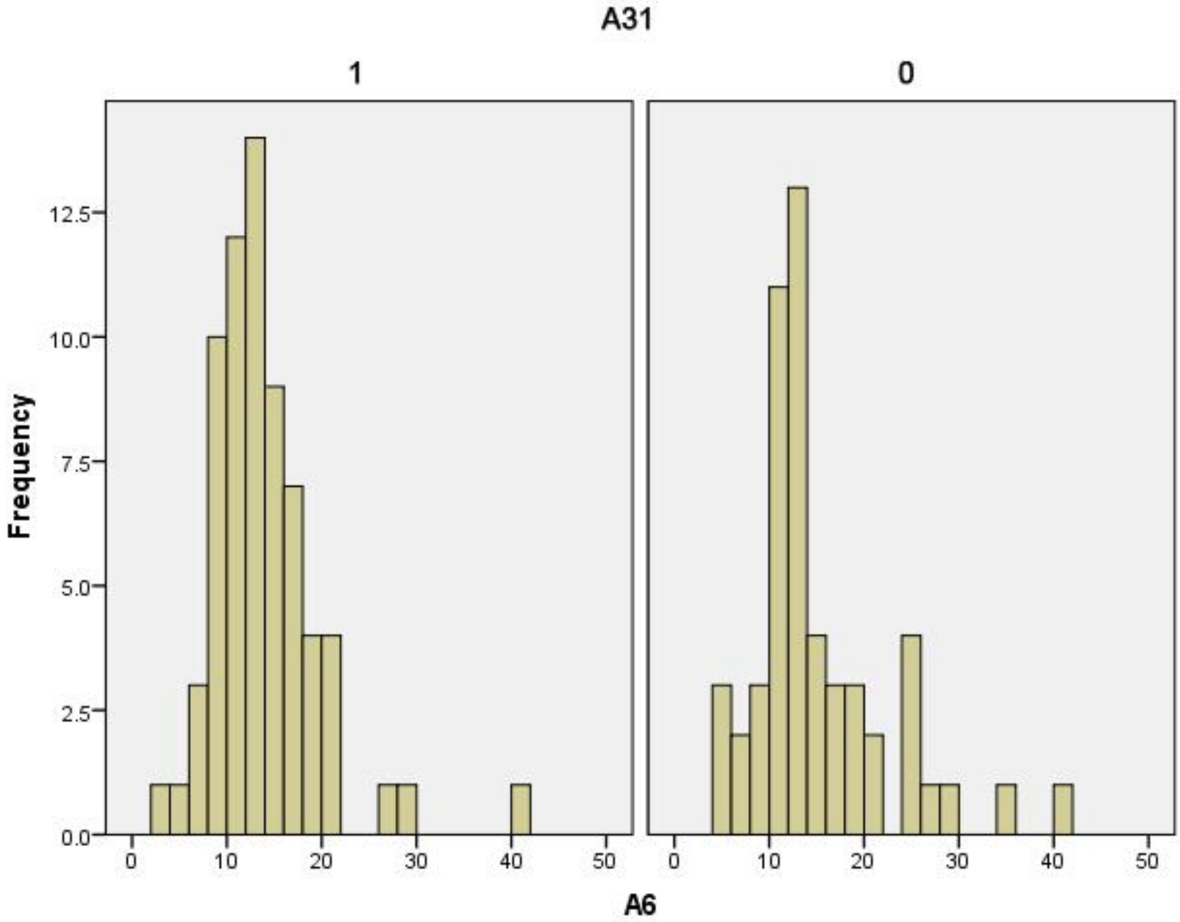


Figure 8: Graph for total non-answer clicks

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A6	68	3	41	13.47	5.679
	Valid N (listwise)	68				
0	A6	52	4	40	14.69	7.259
	Valid N (listwise)	52				

Table 10: Descriptive statistics for total non-answer clicks

Comparison:

The mean and standard deviation of the unreliable response category are higher than the reliable response category. The minimum and maximum non-answer click counts of both the categories are very close to each other.

A7) Total clicks

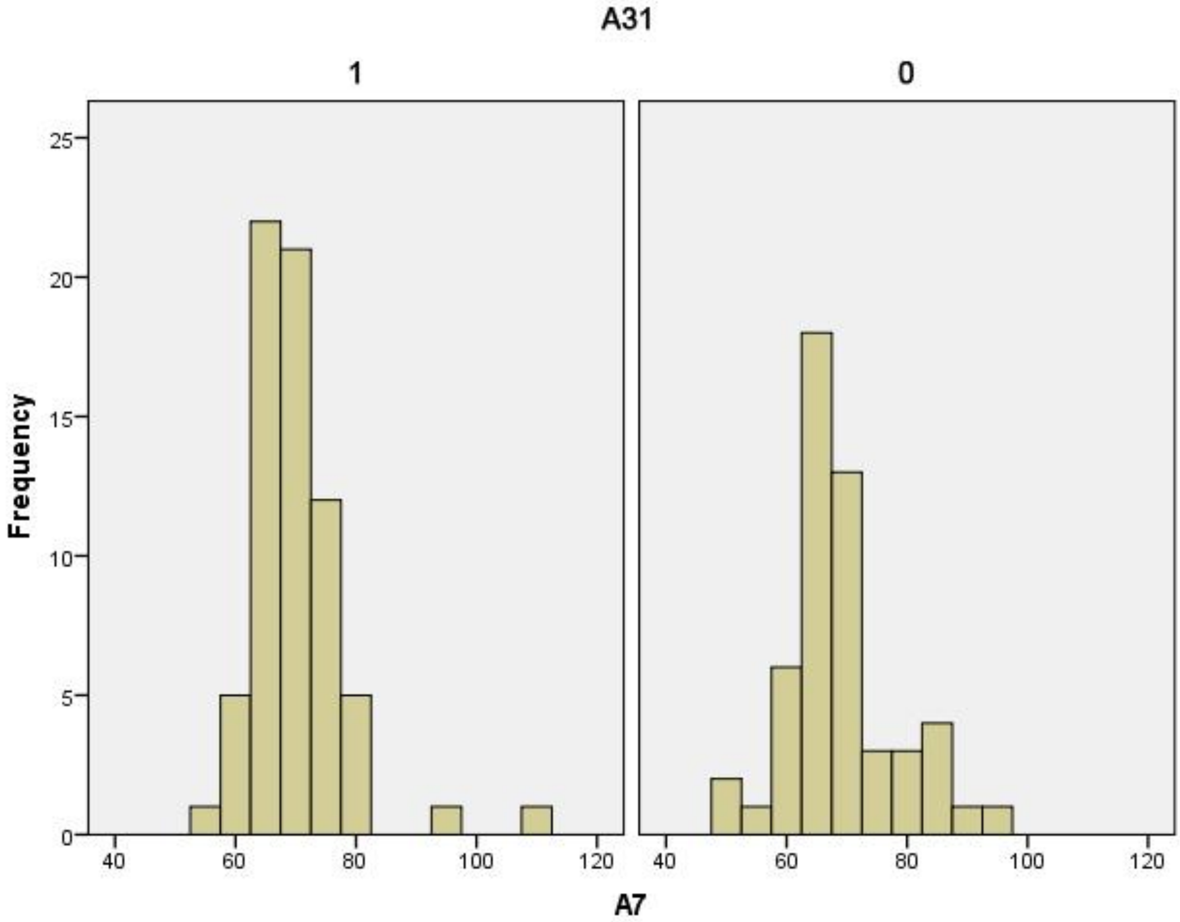


Figure 9: Graph for total clicks

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A7	68	56	109	69.96	7.948
	Valid N (listwise)	68				
0	A7	52	50	94	68.92	8.816
	Valid N (listwise)	52				

Table 11: Descriptive statistics for total clicks

Comparison:

The standard deviation of the unreliable response category is higher than the reliable response category. The mean of both the categories is very close to each other. Based on the graph, more values of the reliable response category is close to the mean.

A8) Average time spent between clicks

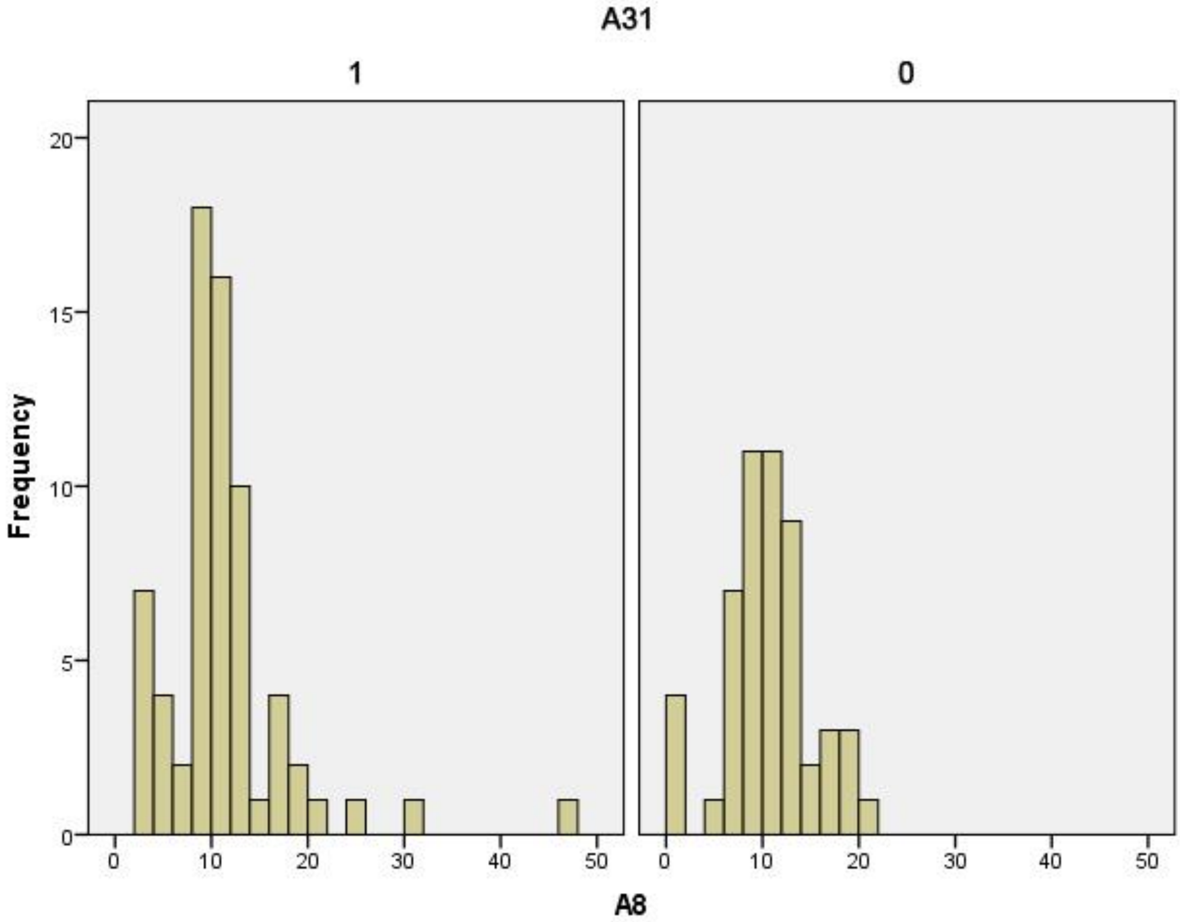


Figure 10: Graph for average time spent between clicks

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A8	68	2	47	10.85	6.674
	Valid N (listwise)	68				
0	A8	52	1	20	10.12	4.427
	Valid N (listwise)	52				

Table 12: Descriptive statistics for average time spent between clicks

Comparison:

The mean and standard deviation of the reliable response category are higher than the unreliable response category. The mean of both the categories are very close to each other. Based on the graph, more values of the reliable response category is close to the mean.

A9) Number of time scrollbar reached top

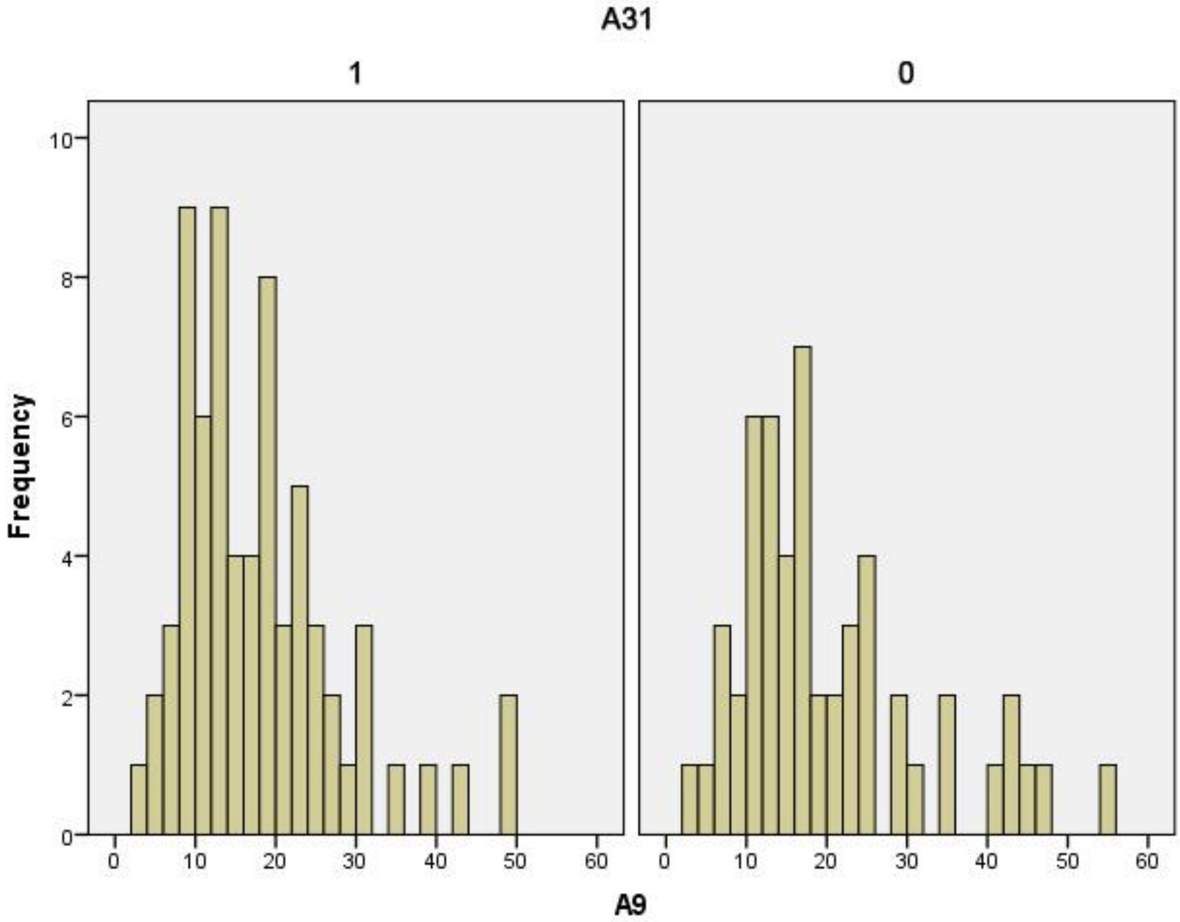


Figure 11: Graph for number of time scrollbar reached top

Descriptive Statistics

A33		N	Minimum	Maximum	Mean	Std. Deviation
1	A9	68	3	48	17.46	9.799
	Valid N (listwise)	68				
2	A9	52	3	54	19.73	11.867
	Valid N (listwise)	52				

Table 13: Descriptive statistics for number of time scrollbar reached top

Comparison:

The mean and standard deviation of the unreliable response category are higher than the reliable response category. The minimum count for both the categories are the same.

A10) Number of time scrollbar reached bottom

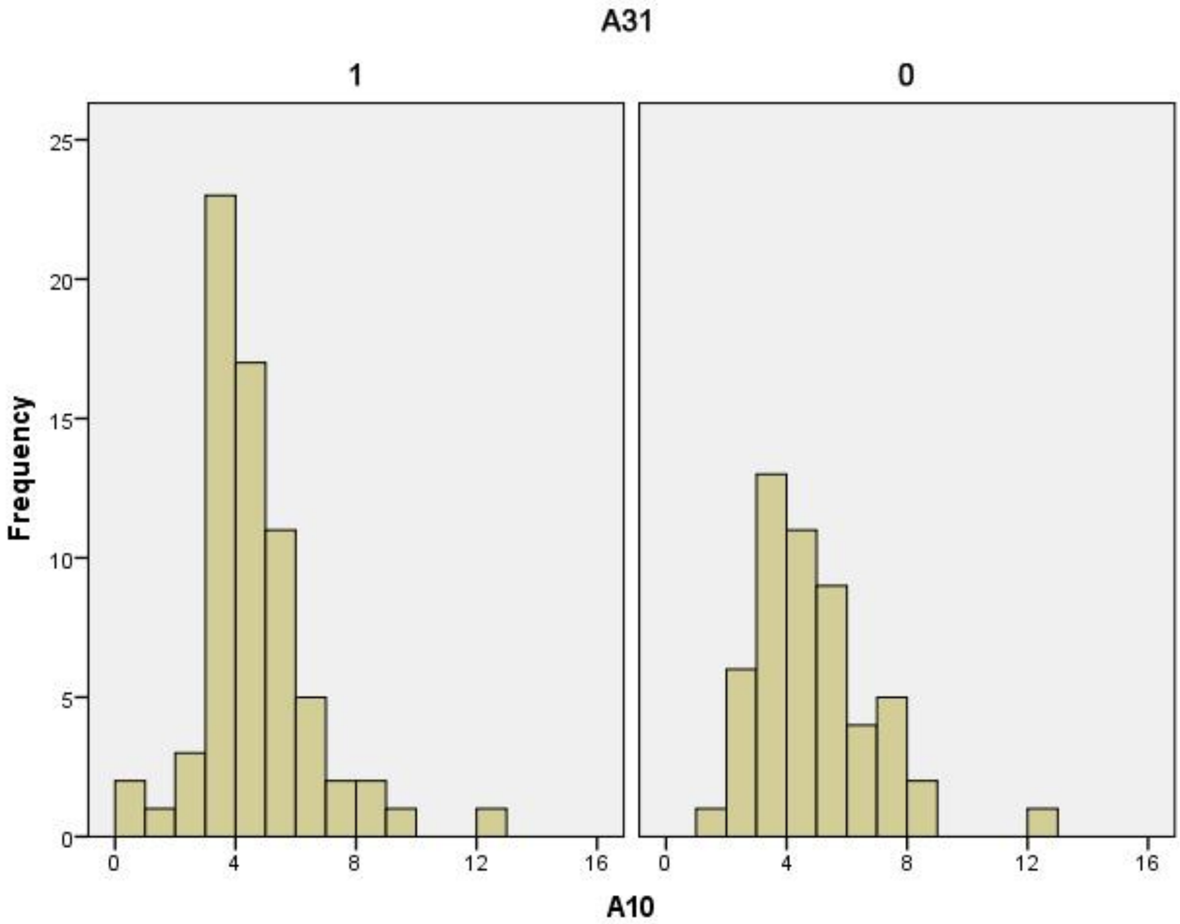


Figure 12: Graph for number of time scrollbar reached bottom

Descriptive Statistics

A33		N	Minimum	Maximum	Mean	Std. Deviation
1	A10	68	0	12	4.12	1.912
	Valid N (listwise)	68				
2	A10	52	1	12	4.38	2.002
	Valid N (listwise)	52				

Table 14: Descriptive statistics for number of time scrollbar reached bottom

Comparison:

The mean and standard deviation of the unreliable response category are higher than the reliable response category. Based on the graph, more values of reliable response category is close to the mean. The maximum count for both categories are the same.

A11) Page revisited status

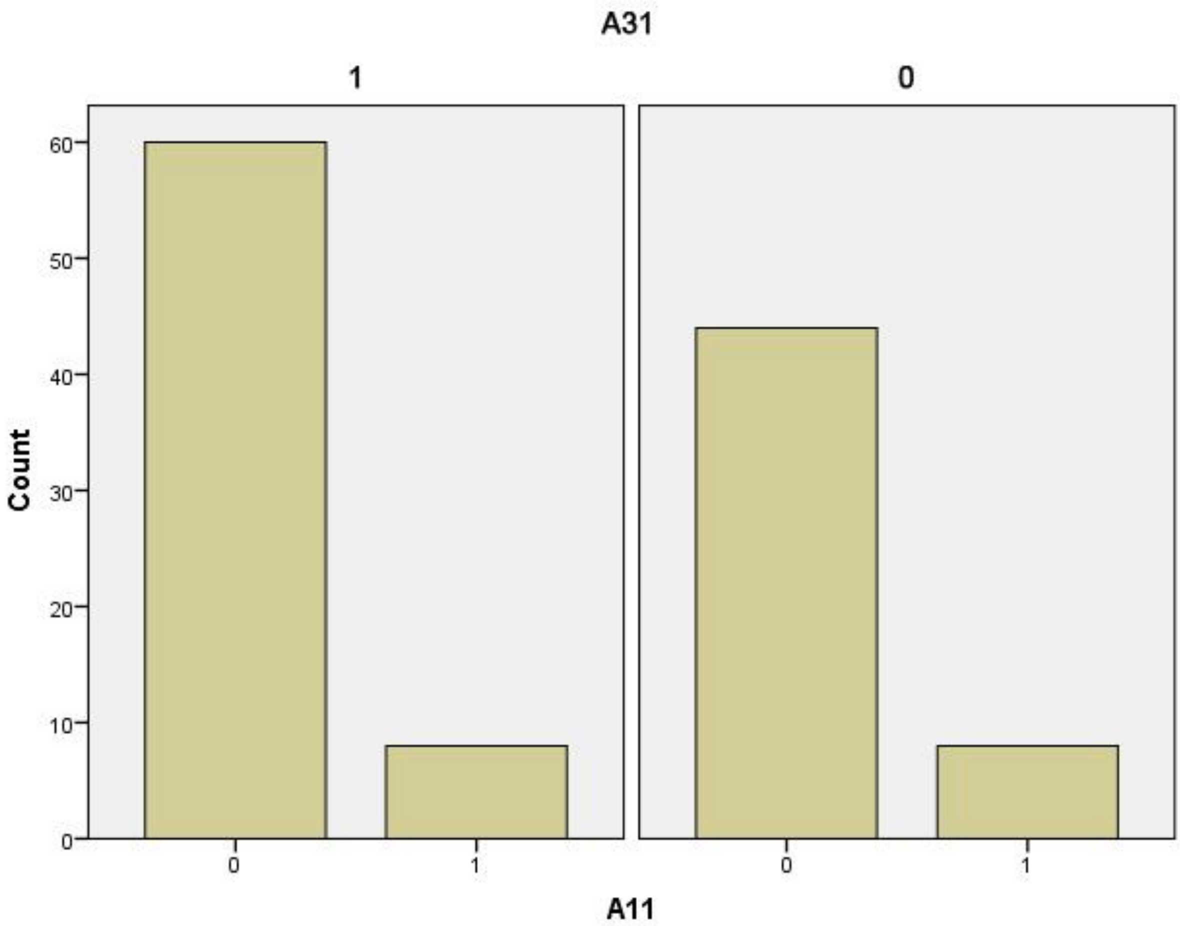


Figure 13: Graph for page revisited status

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A11	68	0	1	.12	.325
	Valid N (listwise)	68				
0	A11	52	0	1	.15	.364
	Valid N (listwise)	52				

Table 15: Descriptive statistics for page revisited status

Comparison:

According to the graph the majority of both the categories, users didn’t revisit the page during the survey. The mean and standard deviation of the unreliable response category are higher than the reliable response category in a small variation.

12) Text pasted status

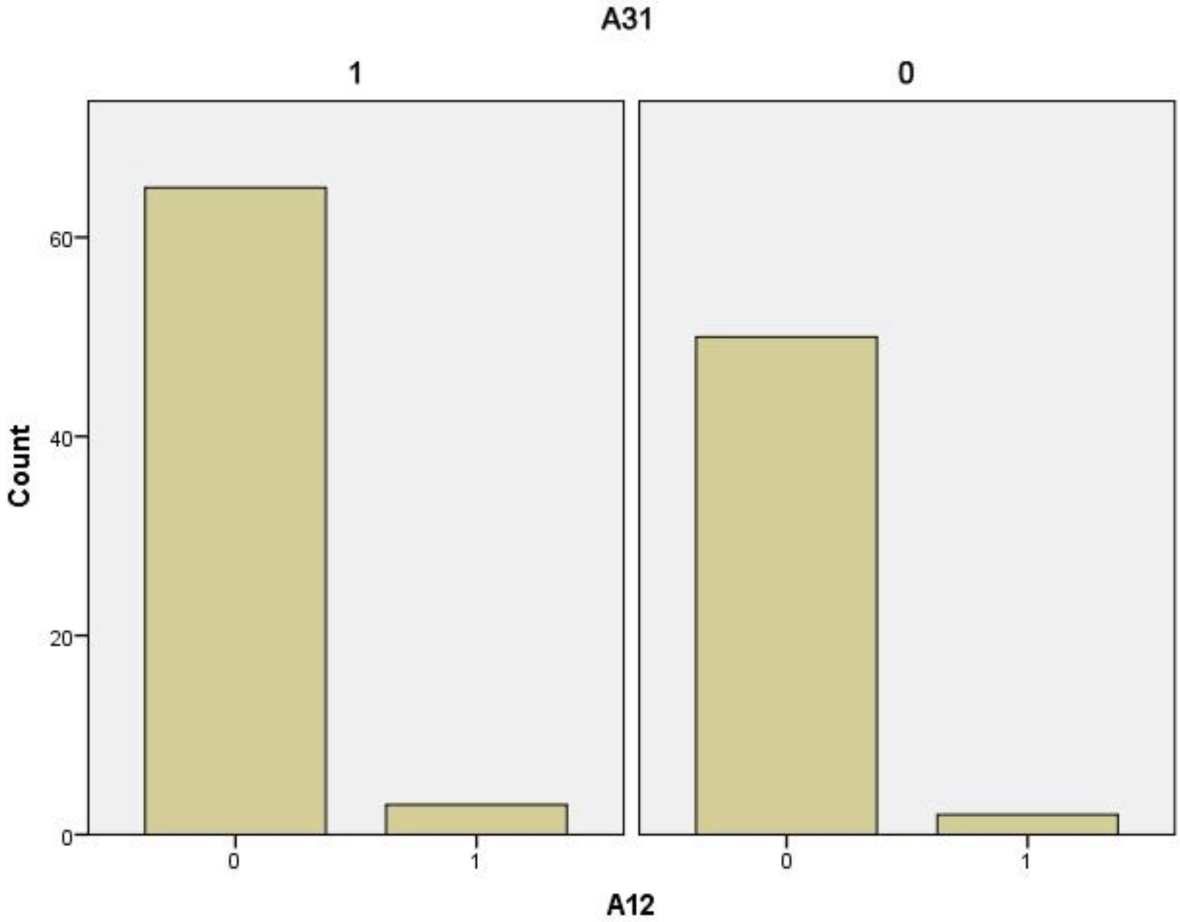


Figure 14: Graph for text pasted status

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A12	68	0	1	.04	.207
	Valid N (listwise)	68				
0	A12	52	0	1	.04	.194
	Valid N (listwise)	52				

Table 16: Descriptive statistics for text pasted status

Comparison:

According to the graph the majority of both the categories users didn't use the paste option during the survey. The standard deviation of the reliable response category is higher than the unreliable response category. The mean for both the categories is the same.

A13) Pasted character count

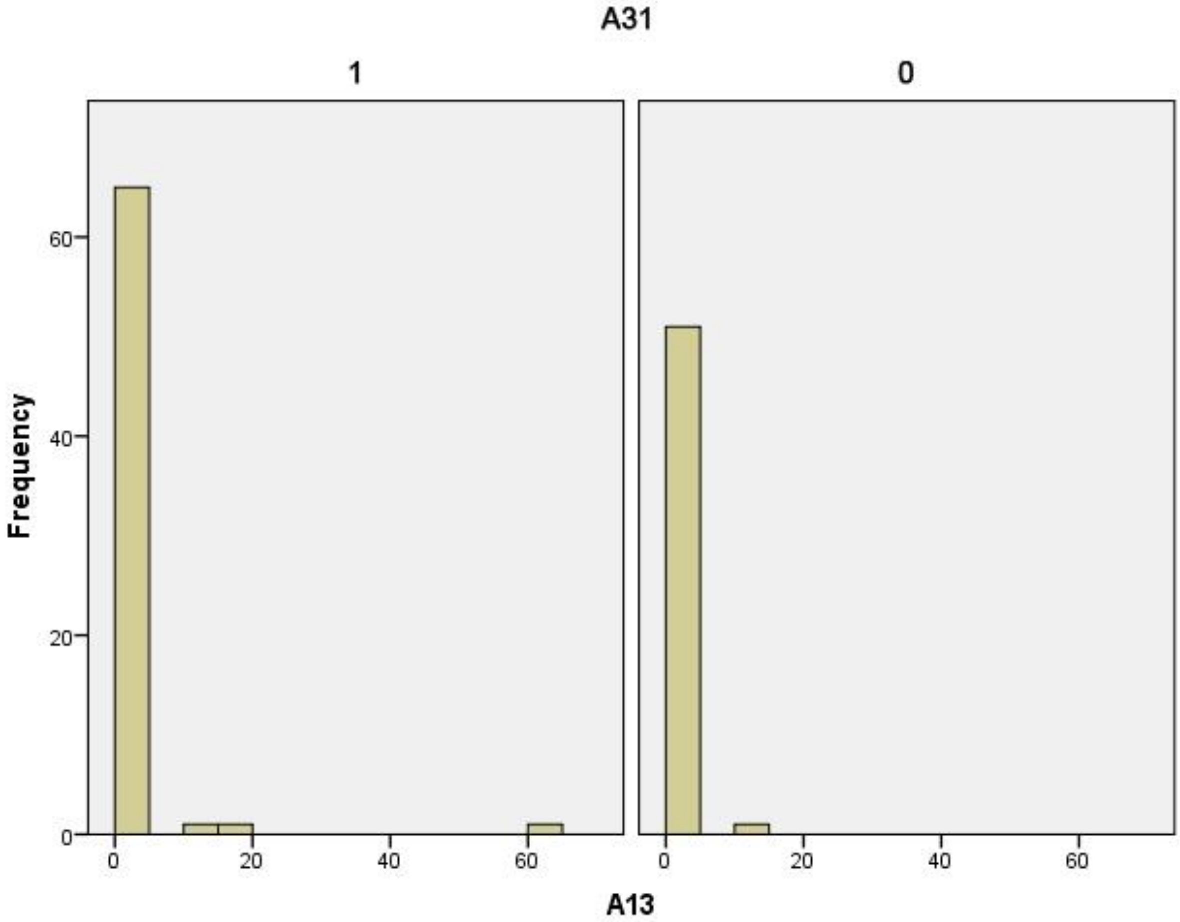


Figure 15: Graph for pasted character count

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A13	68	0	60	1.28	7.608
	Valid N (listwise)	68				
0	A13	52	0	12	.23	1.664
	Valid N (listwise)	52				

Table 17: Descriptive statistics for pasted character count

Comparison:

According to the graph the majority of both the categories, users didn't use paste option during the survey. Hence, pasted character count for both these categories are zero. The mean and standard deviation of the reliable response category are higher than the unreliable response category.

A14) **Average character per minute (cpm) for typing**

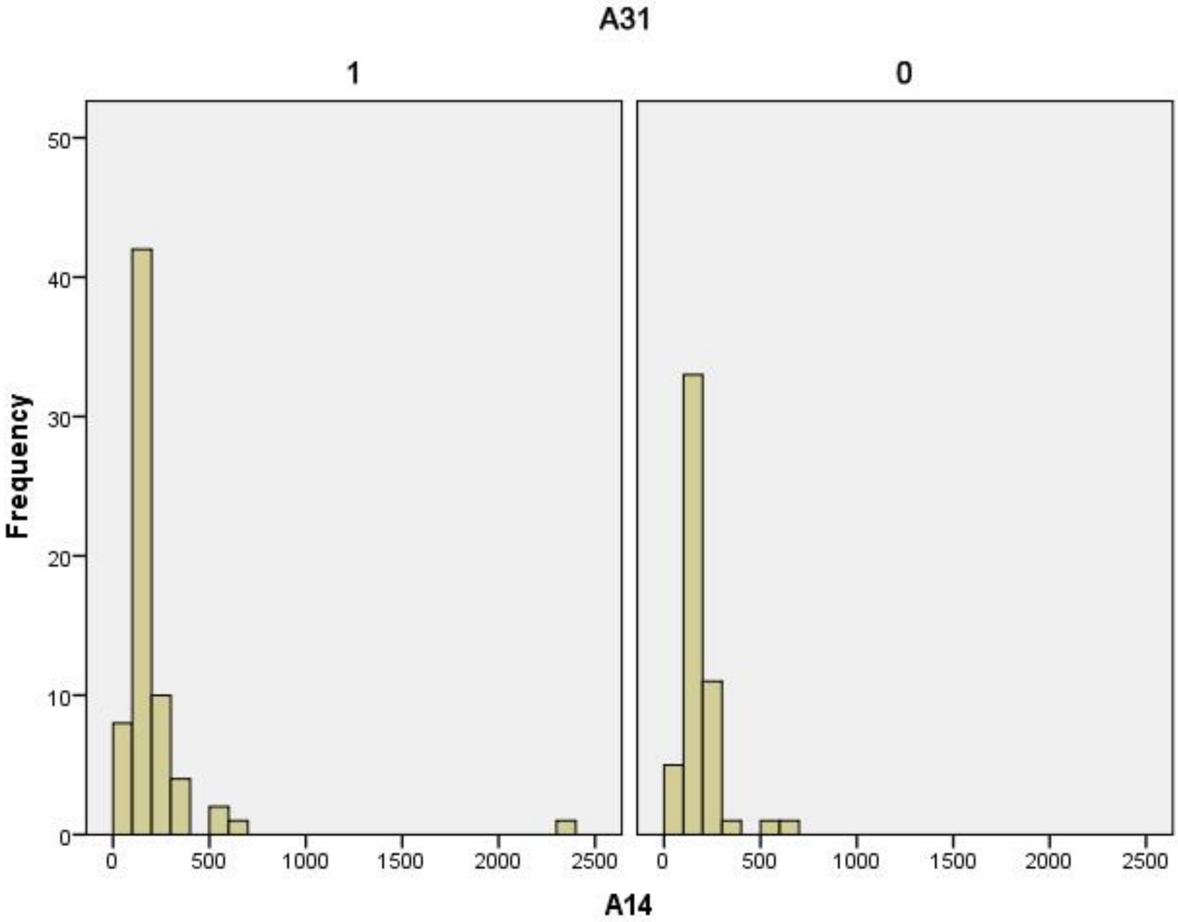


Figure 16: Graph for average character per minute (cpm) for typing

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A14	68	76	2356	218.82	284.929
	Valid N (listwise)	68				
0	A14	52	69	633	177.67	99.211
	Valid N (listwise)	52				

Table 18: Descriptive statistics for average character per minute (cpm) for typing

Comparison:

The mean and standard deviation of reliable response category are higher than the unreliable response category.

A15) The status of out of page during the survey

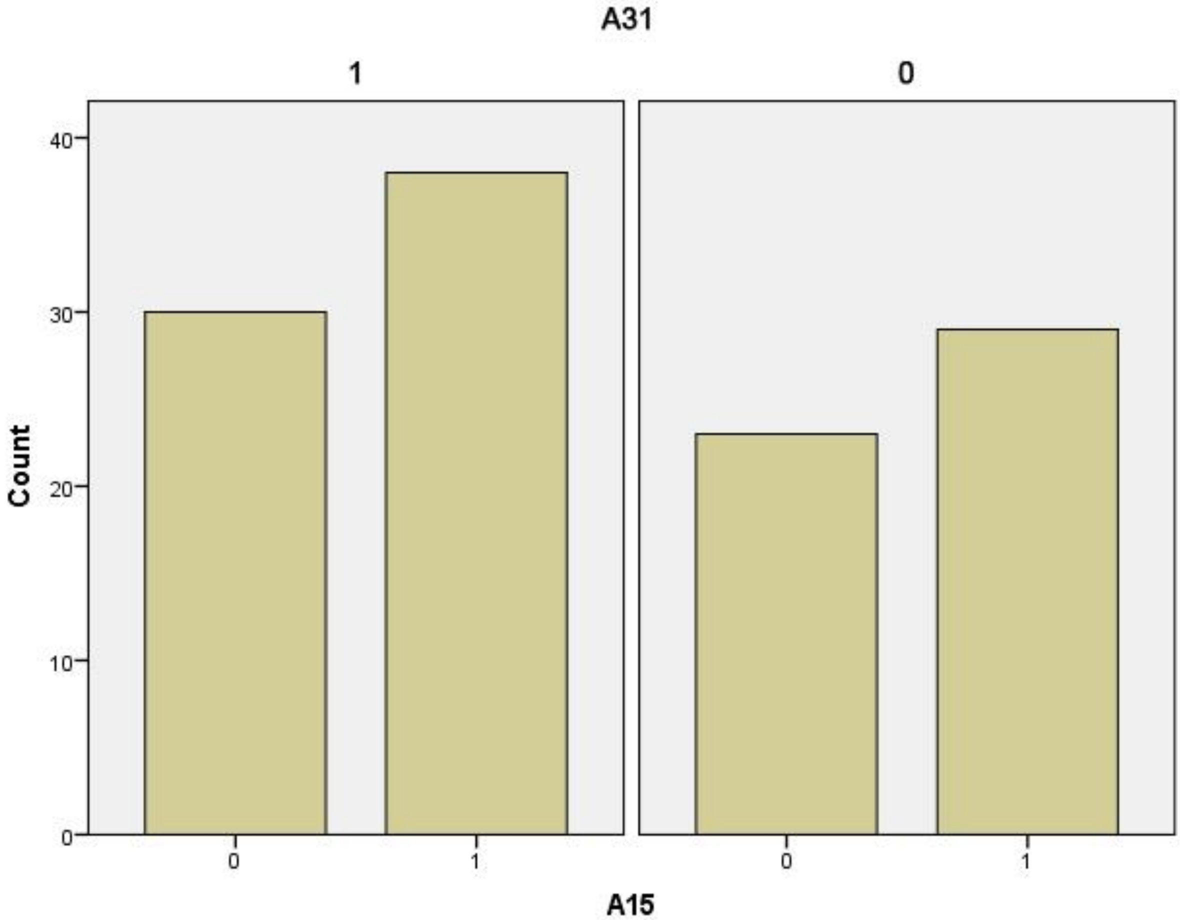


Figure 17: Graph for the status of out of page during the survey

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A15	68	0	1	.56	.500
	Valid N (listwise)	68				
0	A15	52	0	1	.56	.502
	Valid N (listwise)	52				

Table 19: Descriptive statistics for the status of out of page during the survey

Comparison:

According to the graph the majority of both the category responses stayed out of page at least once during the survey. The mean and standard deviation of both the categories are very close to each other.

A16) The status of idle during the survey

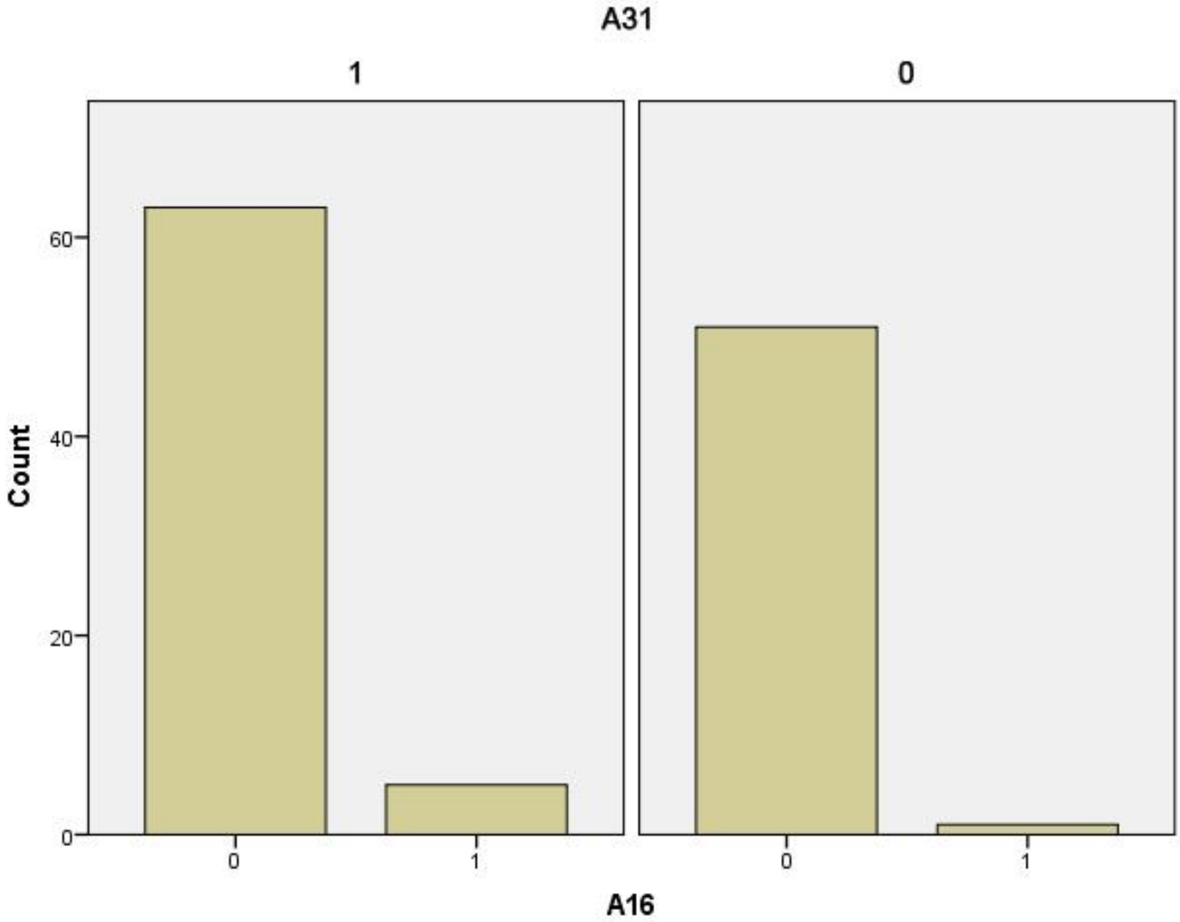


Figure 18: The status of idle during the survey

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A16	68	0	1	.07	.263
	Valid N (listwise)	68				
0	A16	52	0	1	.02	.139
	Valid N (listwise)	52				

Table 20: Descriptive statistics for the status of idle during the survey

Comparison:

According to the graph the majority of both the categories, users didn't stay idle at least once during the survey. The mean and standard deviation of the reliable response category is higher than the unreliable response category.

A17) Average time spent per page

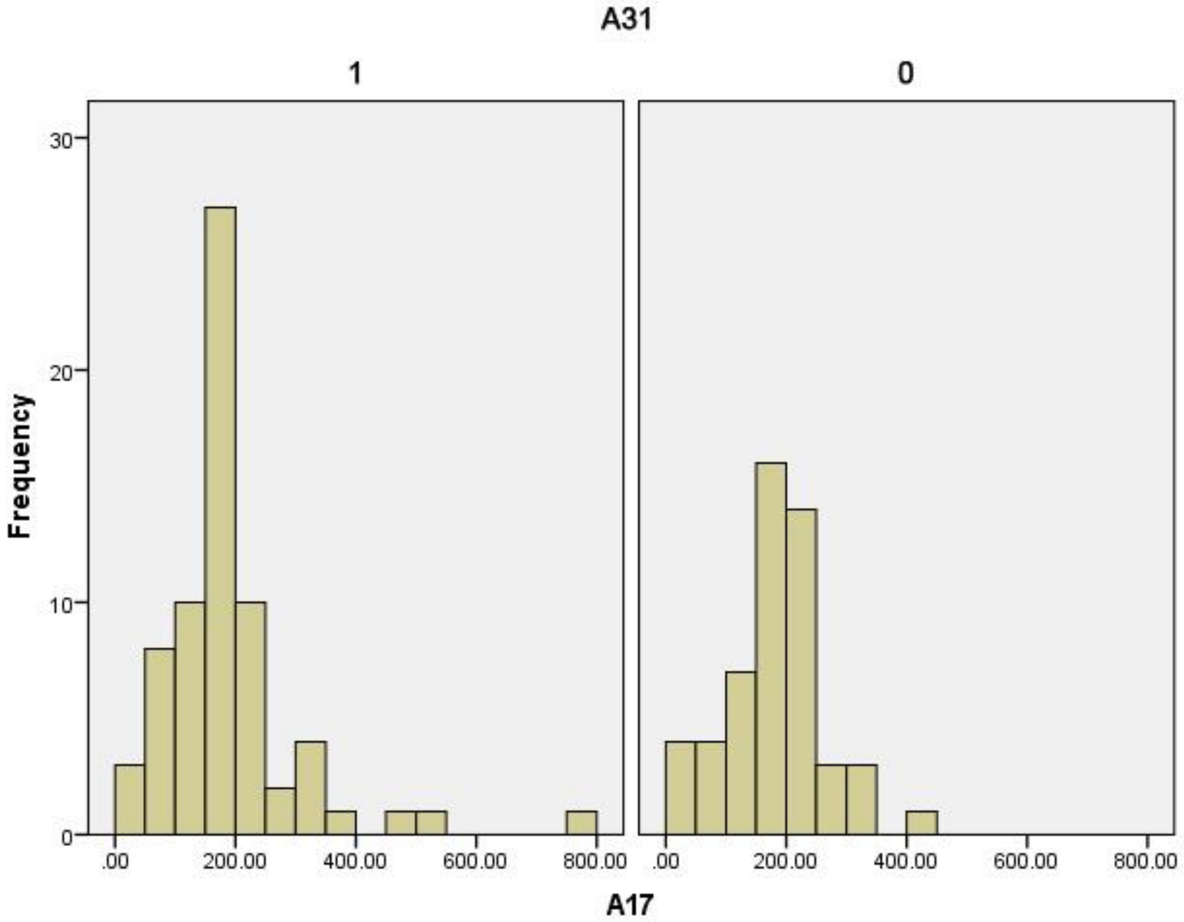


Figure 19: Average time spent per page

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A17	68	32.75	792.00	189.2537	118.57762
	Valid N (listwise)	68				
0	A17	52	16.25	406.50	175.4183	78.84428
	Valid N (listwise)	52				

Table 21: Descriptive statistics for average time spent per page

Comparison:

The mean and standard deviation of the reliable response category are higher than the unreliable response category. Based on the graph, more values of the reliable response category is close to the mean.

A18) Answer changes during the survey

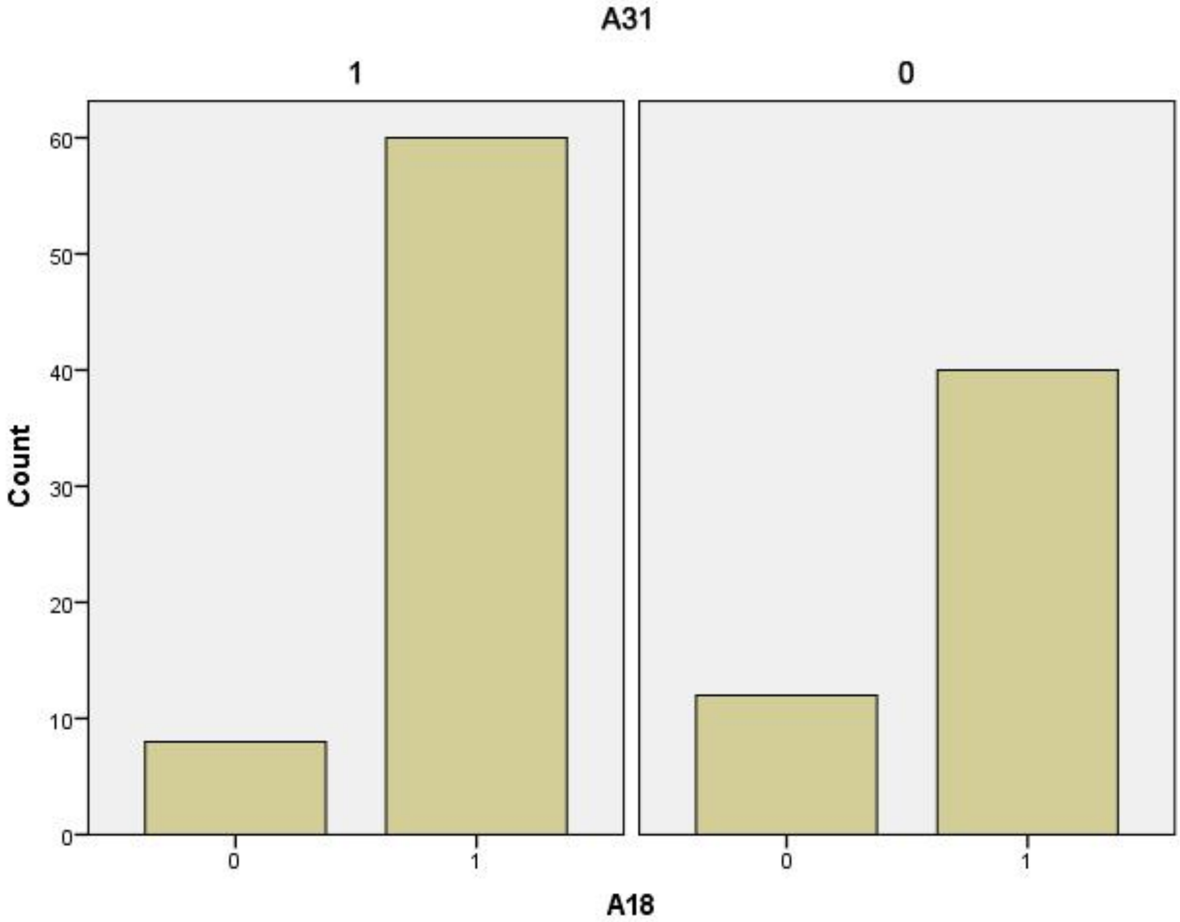


Figure 20: Answer changes during the survey

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A18	68	0	1	.88	.325
	Valid N (listwise)	68				
0	A18	52	0	1	.77	.425
	Valid N (listwise)	52				

Table 22: Descriptive statistics for answer changes during the survey

Comparison:

According to the graph majority of both the categories, users changed the answers at least once during the survey. The standard deviation of the unreliable response category is higher than the reliable response category. The mean of the reliable response category is higher than the unreliable response category.

A19) Text areas changes during the survey

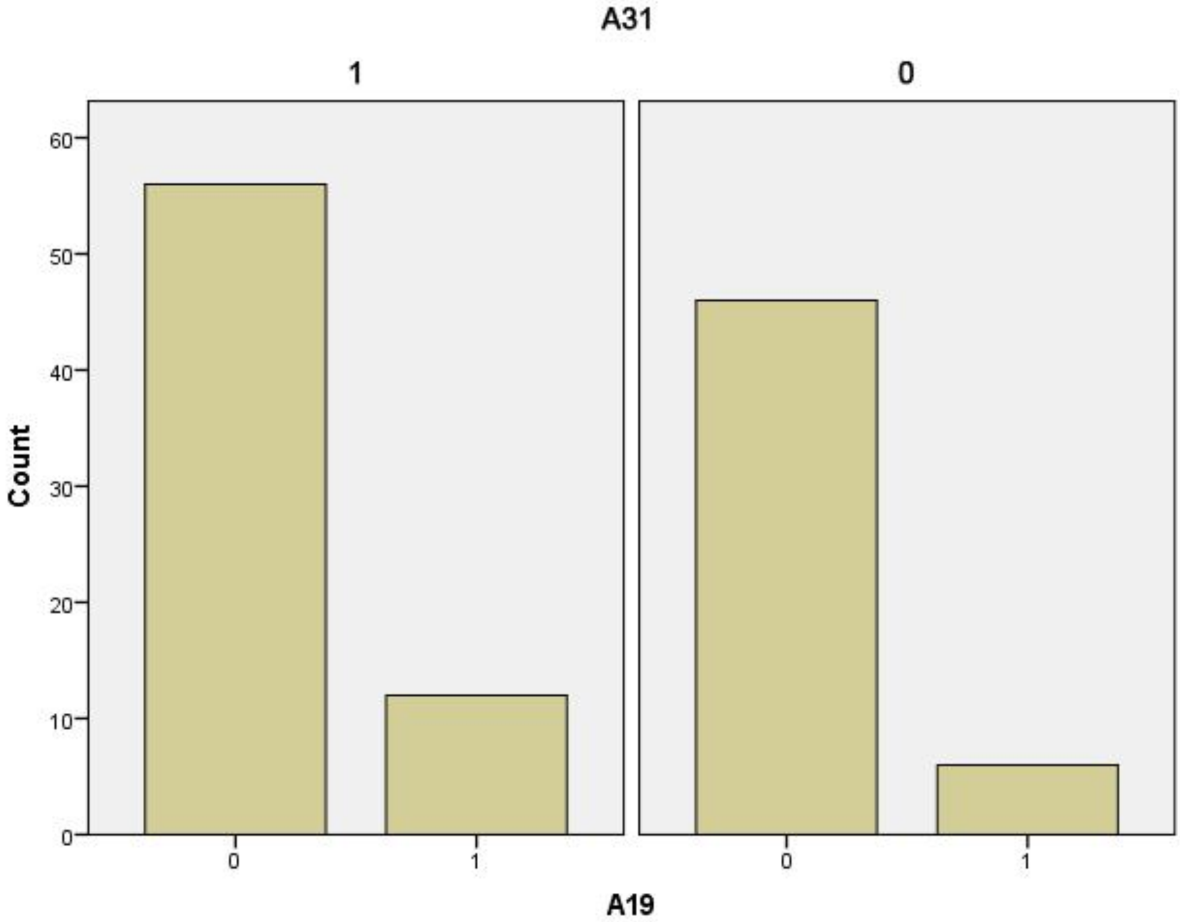


Figure 21: Graph for text areas changes during the survey

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A19	68	0	1	.18	.384
	Valid N (listwise)	68				
0	A19	52	0	1	.12	.323
	Valid N (listwise)	52				

Table 23: Descriptive statistics for text areas changes during the survey

Comparison:

According to the graph the majority of both the categories, users changed text area answers at least once during the survey. The mean and standard deviation of the reliable response category is higher than the unreliable response category.

A20) Drop down changes during the survey

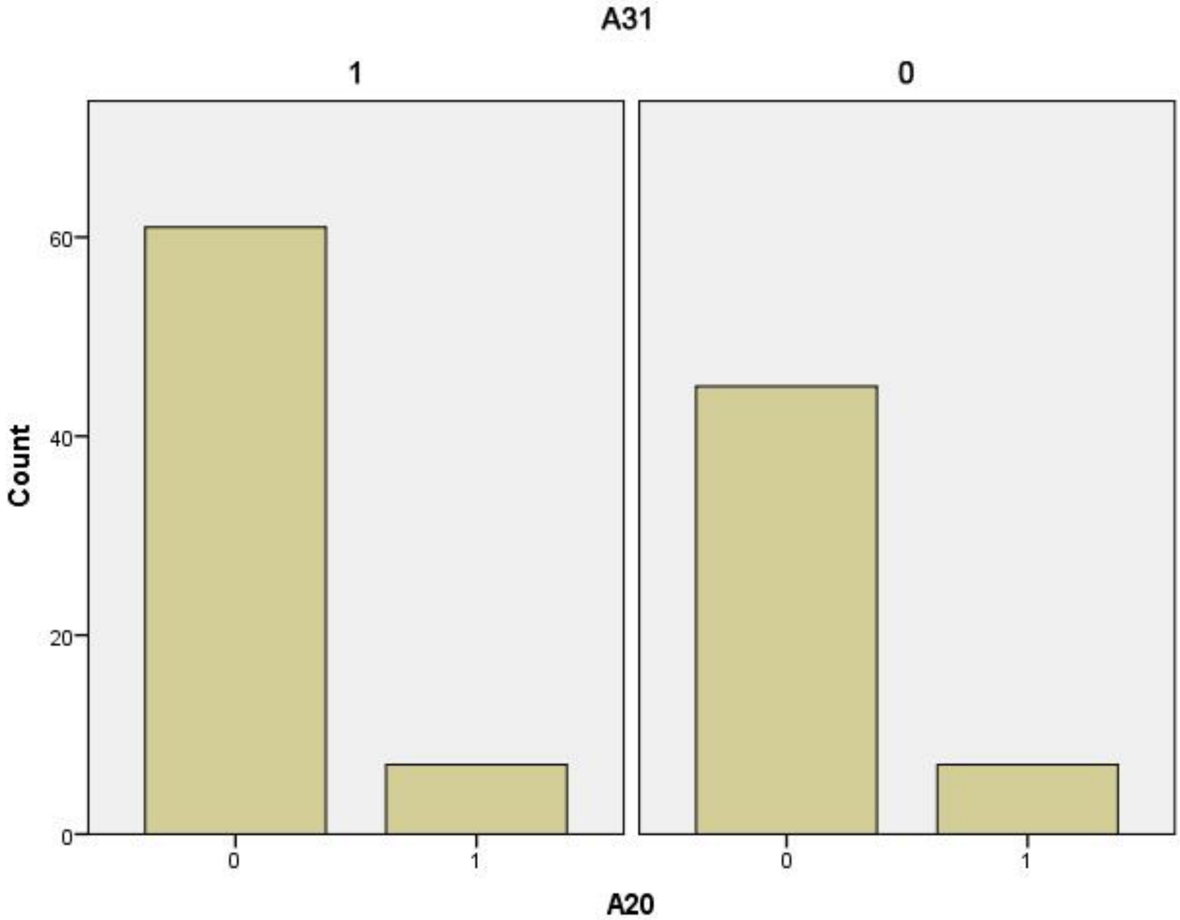


Figure 22: Graph for drop down changes during the survey

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A20	68	0	1	.10	.306
	Valid N (listwise)	68				
0	A20	52	0	1	.13	.345
	Valid N (listwise)	52				

Table 24: Descriptive statistics for drop down changes during the survey

Comparison:

According to the graph the majority of both the categories, users did not change dropdown answers at least once during the survey. The mean and standard deviation of unreliable response category is higher than the reliable response category.

A21) Text fields changes during the survey

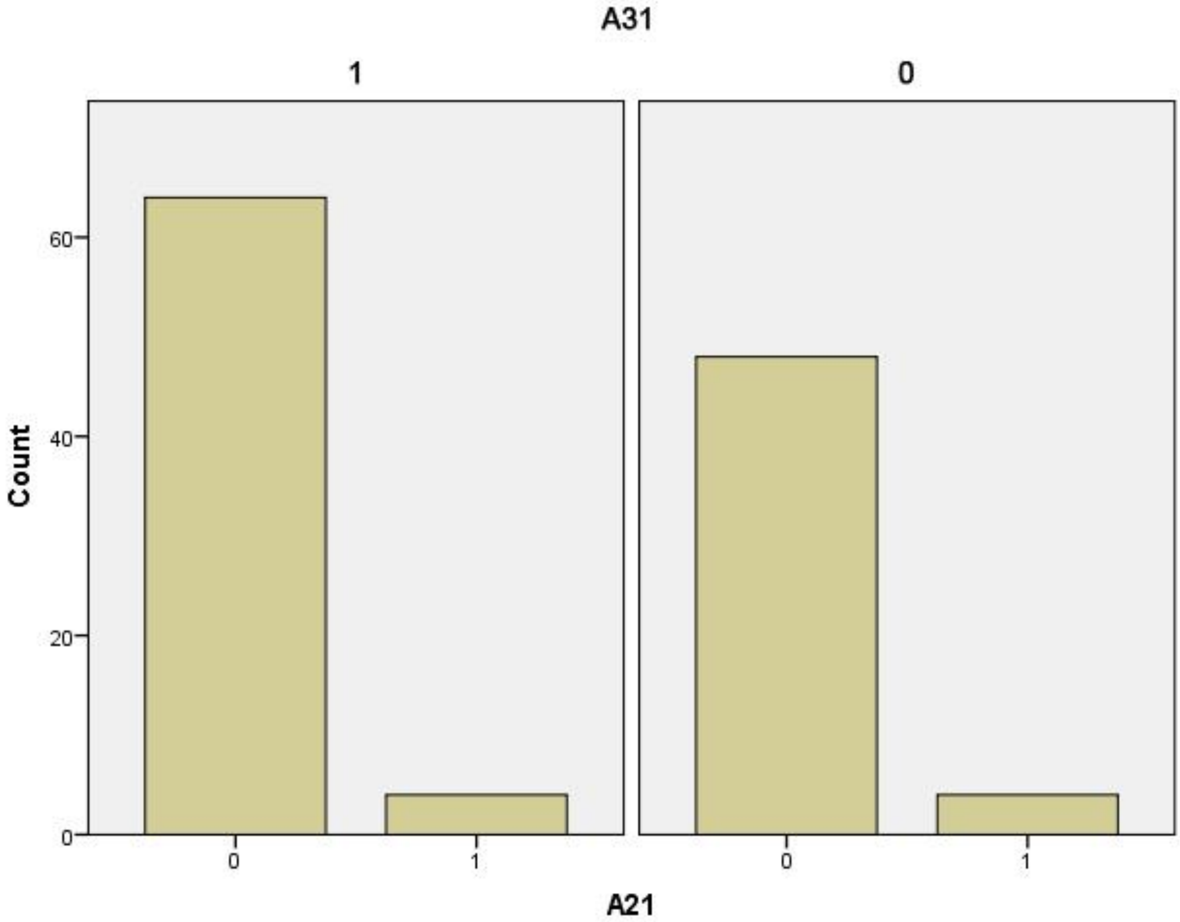


Figure 23: Graph for text fields changes during the survey

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A21	68	0	1	.06	.237
	Valid N (listwise)	68				
0	A21	52	0	1	.08	.269
	Valid N (listwise)	52				

Table 25: Descriptive statistics for text fields changes during the survey

Comparison:

According to the graph the majority of both the categories, users did not change the drop down answers at least once during the survey. The mean and standard deviation of the unreliable response category is higher than the reliable response category.

A22) Radio button changes during the survey

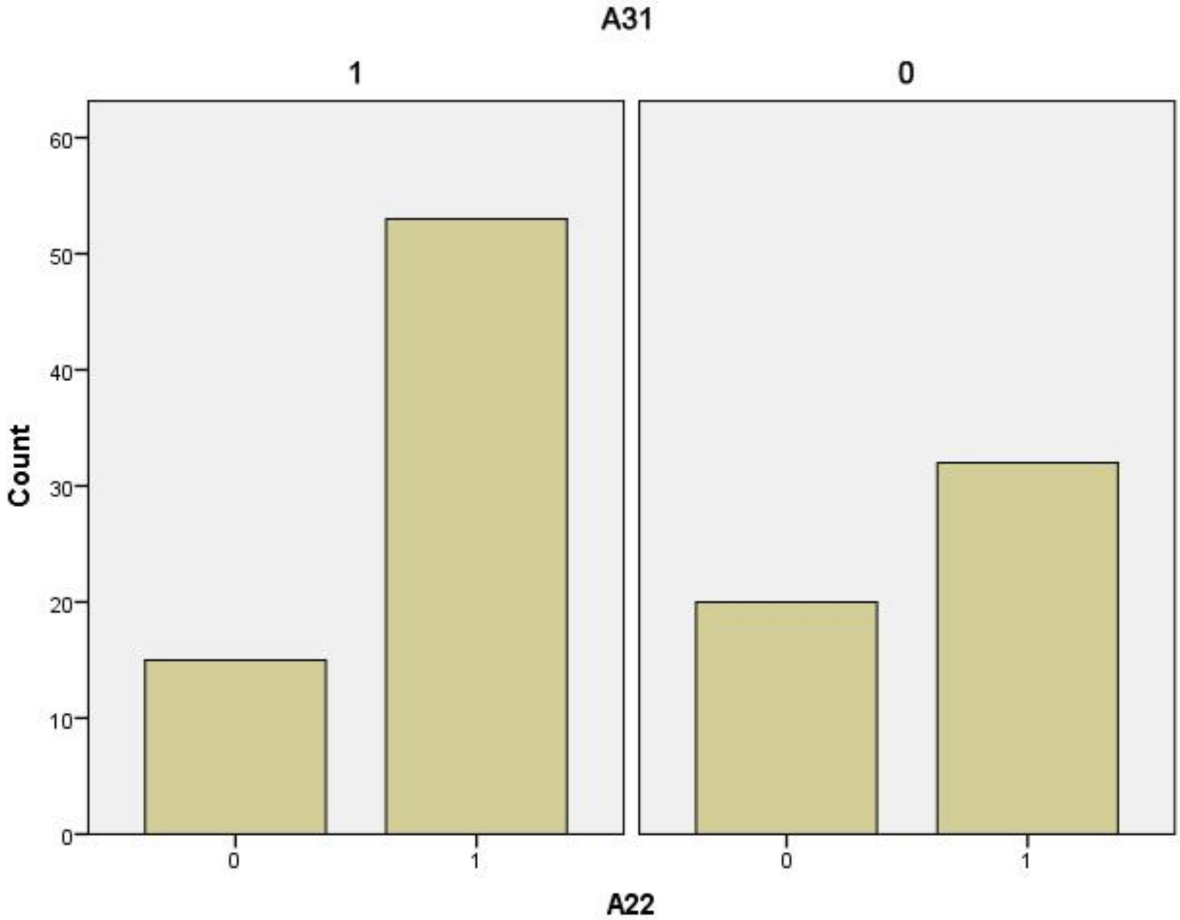


Figure 24: Graph for radio button changes during the survey

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A22	68	0	1	.78	.418
	Valid N (listwise)	68				
0	A22	52	0	1	.62	.491
	Valid N (listwise)	52				

Table 26: Descriptive statistics for radio button changes during the survey

Comparison:

According to the graph the majority of both the categories, users changed radio button answers at least once during the survey. The standard deviation of the unreliable response category is higher than the reliable response category. The mean of the reliable response category is higher than the unreliable response category.

A23) Check boxes changes during the survey

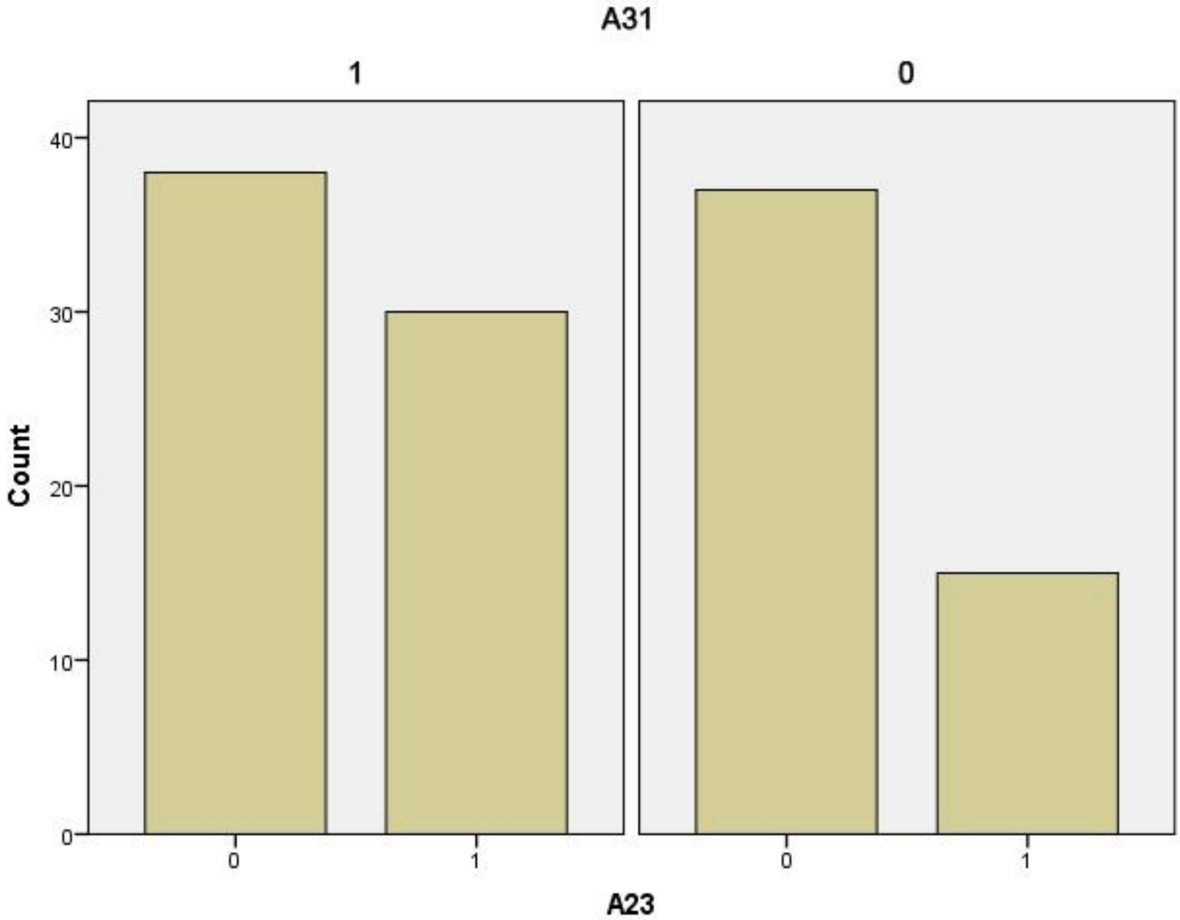


Figure 25: Graph for check boxes changes during the survey

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A23	68	0	1	.44	.500
	Valid N (listwise)	68				
0	A23	52	0	1	.29	.457
	Valid N (listwise)	52				

Table 27: Descriptive statistics for check boxes changes during the survey

Comparison:

According to the graph the majority of both the categories, users did not change check box answers at least once during the survey. The mean and standard deviation of the reliable response category is higher than the unreliable response category.

A24) Average time spent on text area questions

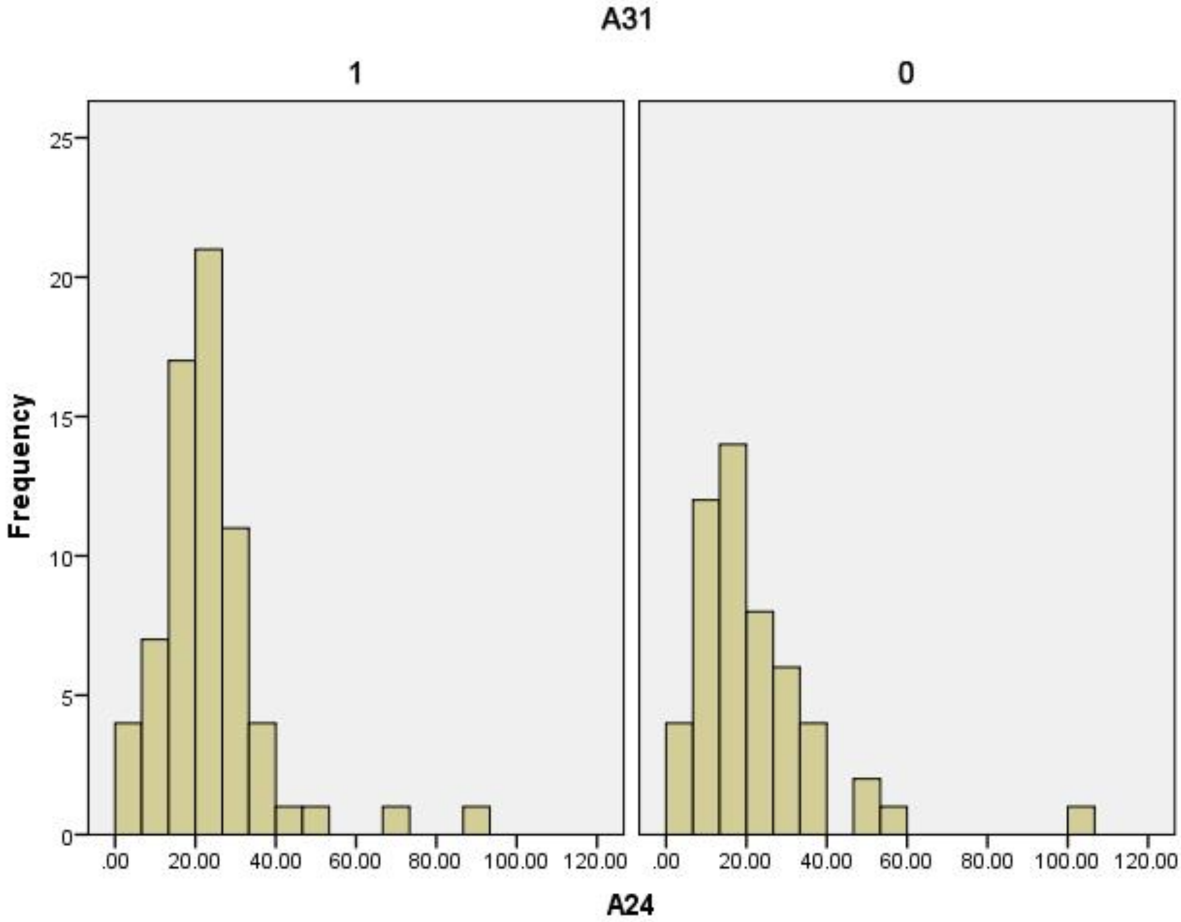


Figure 26: Graph for average time spent on text area questions

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A24	68	3.00	87.00	23.2696	13.54336
	Valid N (listwise)	68				
0	A24	52	.00	103.33	21.7625	16.27138
	Valid N (listwise)	52				

Table 28: Descriptive statistics for average time spent on text area questions

Comparison:

The mean of the reliable response category is higher than the unreliable response category. The standard deviation of the unreliable response category is higher than the reliable response category. Based on the graph, more values of the reliable response category is close to the mean.

A25) Average time spent on drop down questions

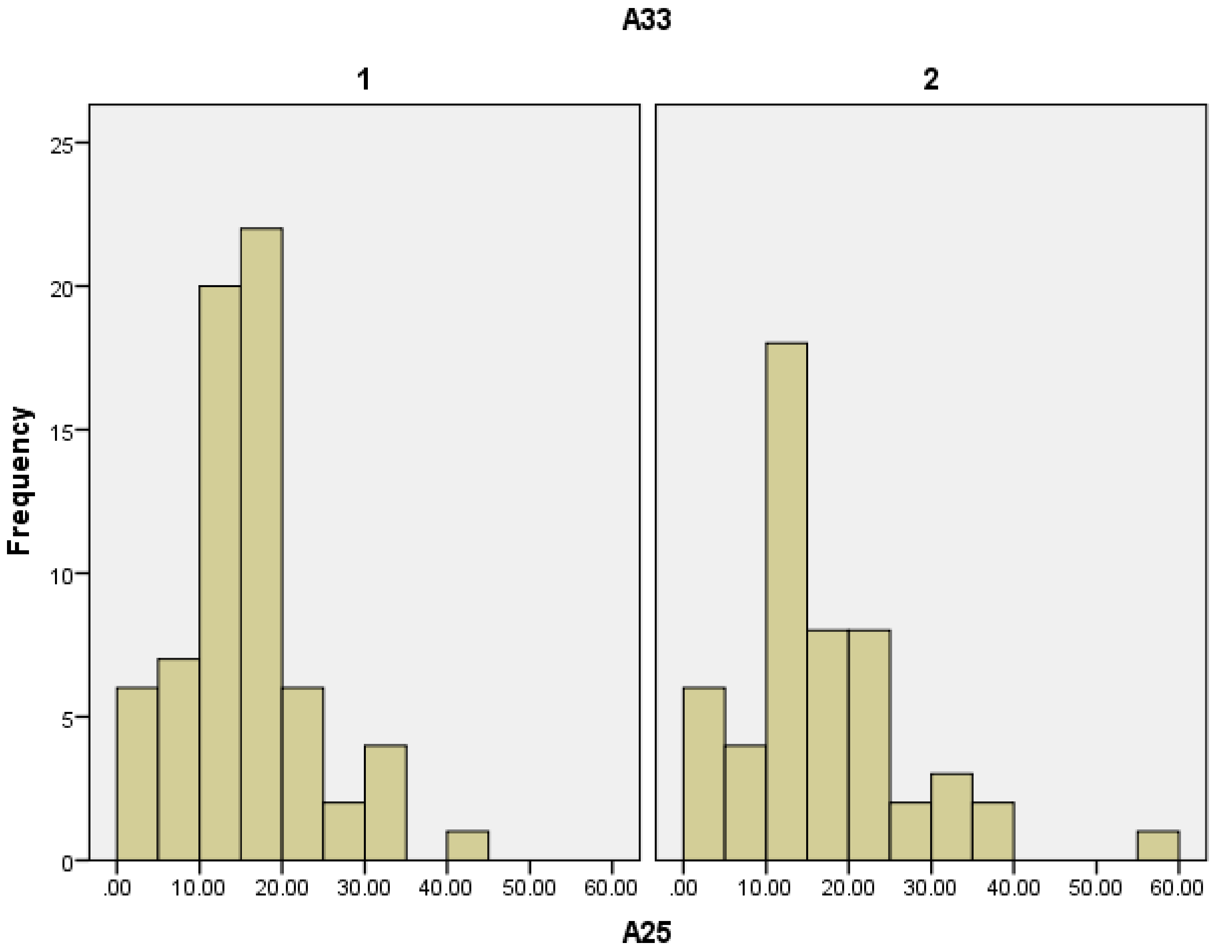


Figure 27: Graph for average time spent on drop down questions

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A25	68	.50	40.00	15.5074	7.75451
	Valid N (listwise)	68				
0	A25	52	.00	55.50	16.8317	10.50417
	Valid N (listwise)	52				

Table 29: Descriptive statistics for average time spent on drop down questions

Comparison:

The mean and standard deviation of the unreliable response category are higher than the reliable response category. Based on the graph, more values of reliable response category is close to the mean.

A26) Average time spent on radio button questions

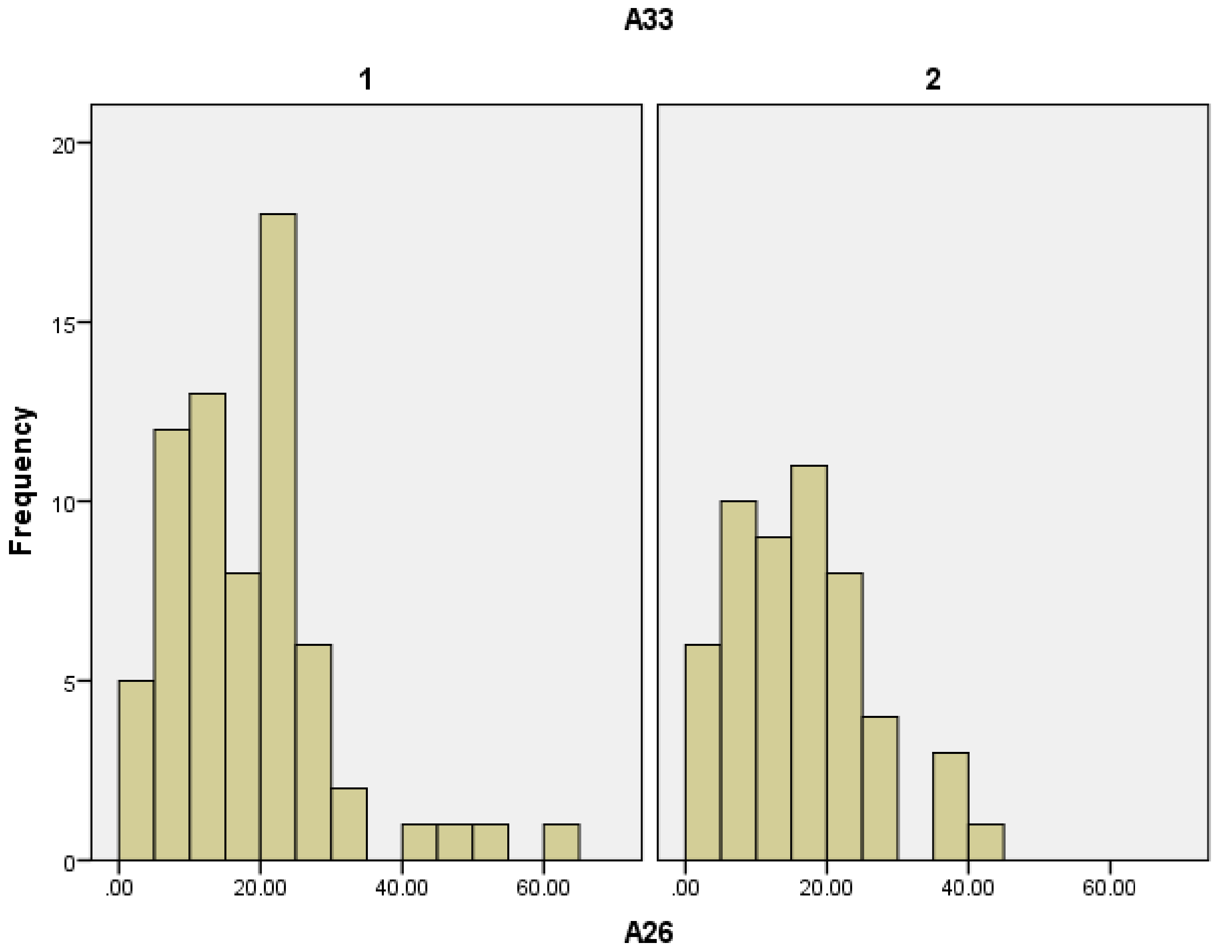


Figure 28: Graph for average time spent on radio button questions

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A26	68	2.67	63.00	18.2994	11.49405
	Valid N (listwise)	68				
0	A26	52	1.00	42.67	16.0448	9.45236
	Valid N (listwise)	52				

Table 30: Descriptive statistics for average time spent on radio button questions

Comparison:

The mean and standard deviation of the reliable response category are higher than the unreliable response category. Based on the graph, more values of the reliable response category is close to the mean.

A27) Average time spent on text field questions

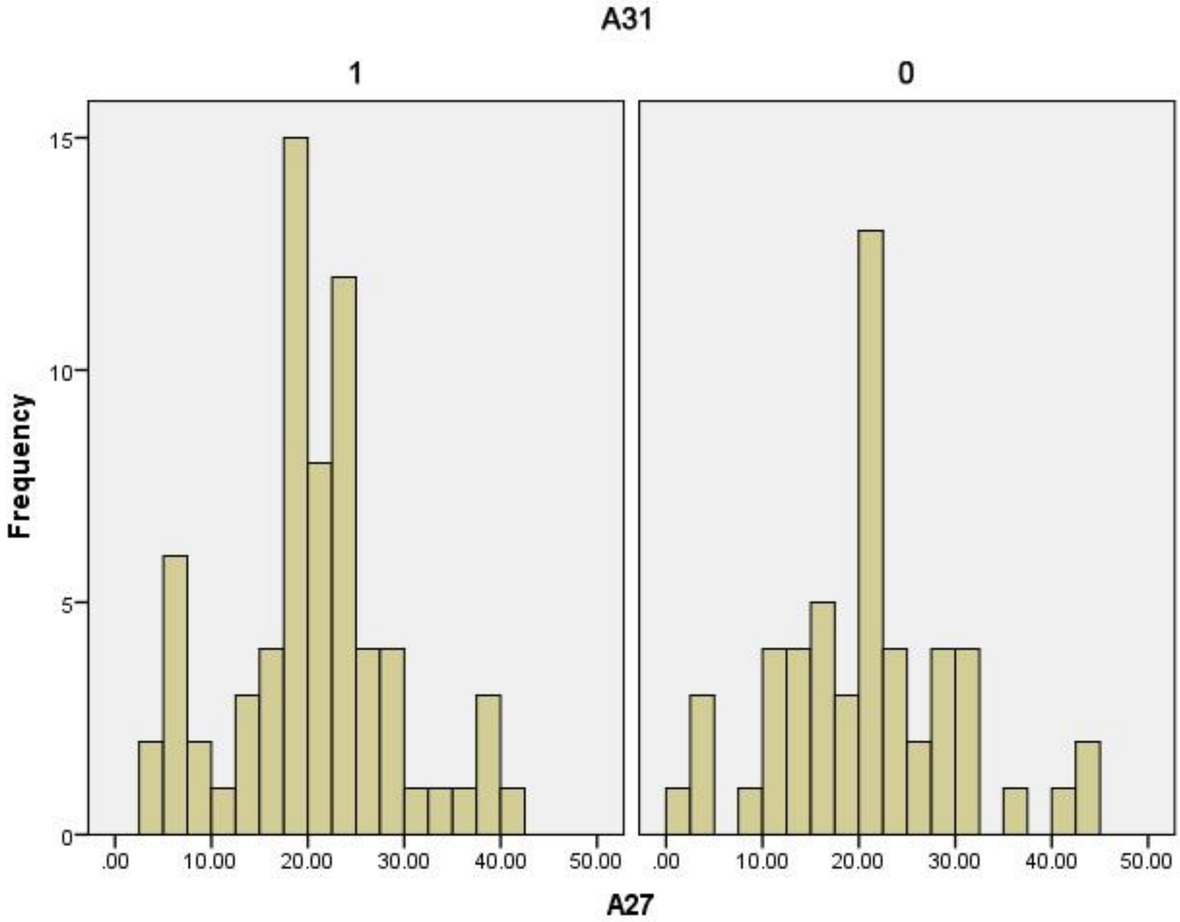


Figure 29: Graph for average time spent on text field questions

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A27	68	3.81	42.44	20.6746	8.70535
	Valid N (listwise)	68				
0	A27	52	1.56	44.69	21.0542	9.49355
	Valid N (listwise)	52				

Table 31: Descriptive statistics for average time spent on text field questions

Comparison:

The mean and standard deviation of the unreliable response category are higher than the reliable response category.

A28) Average time spent on check box questions

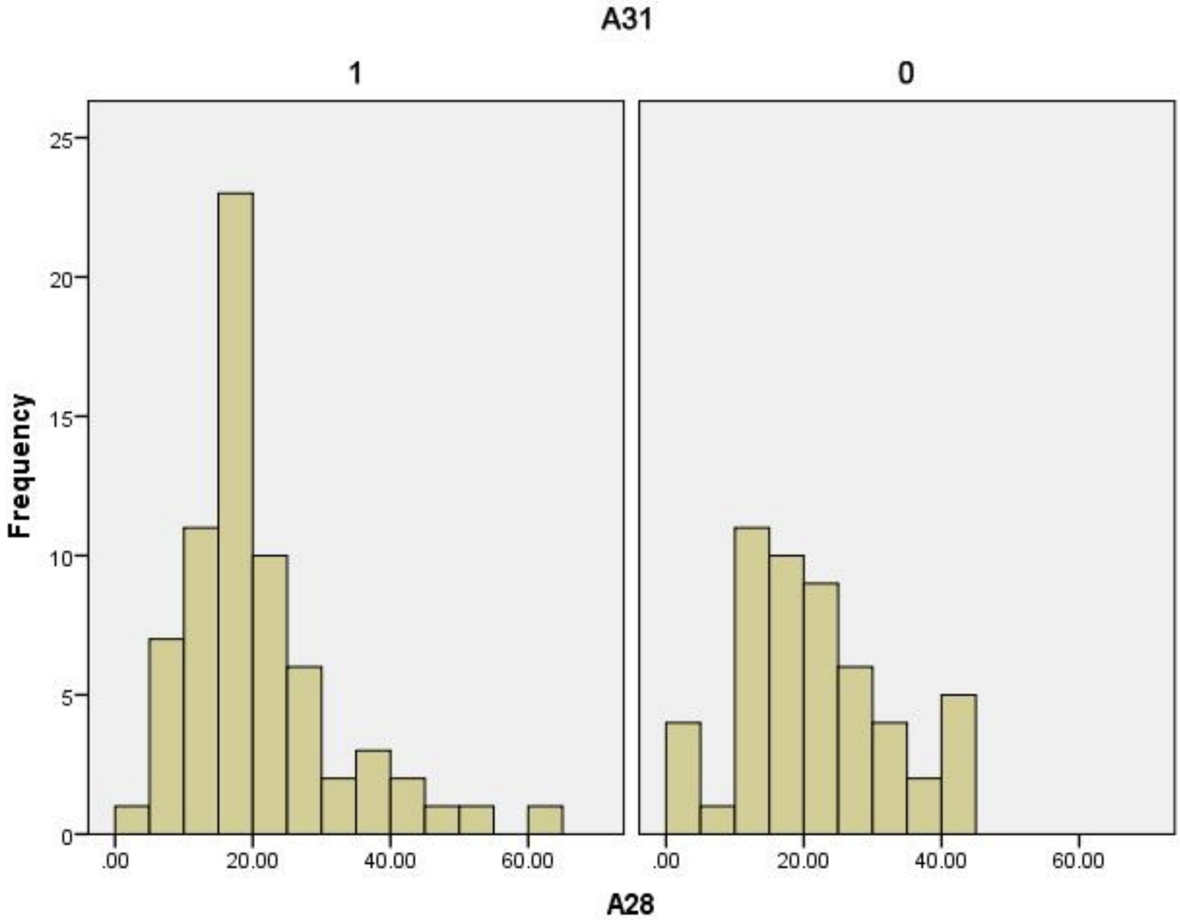


Figure 30: Graph for average time spent on check box questions

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A28	68	4.25	63.75	20.3566	11.26664
	Valid N (listwise)	68				
0	A28	52	.50	44.50	21.1058	10.95270
	Valid N (listwise)	52				

Table 32: Descriptive statistics for average time spent on check box questions

Comparison:

The mean of the unreliable response category is higher than the reliable response category. The standard deviation of the reliable response category is higher than the unreliable response category. Based on the graph, more values of the reliable response category is close to the mean.

A29) Average time taken between attempts on questions

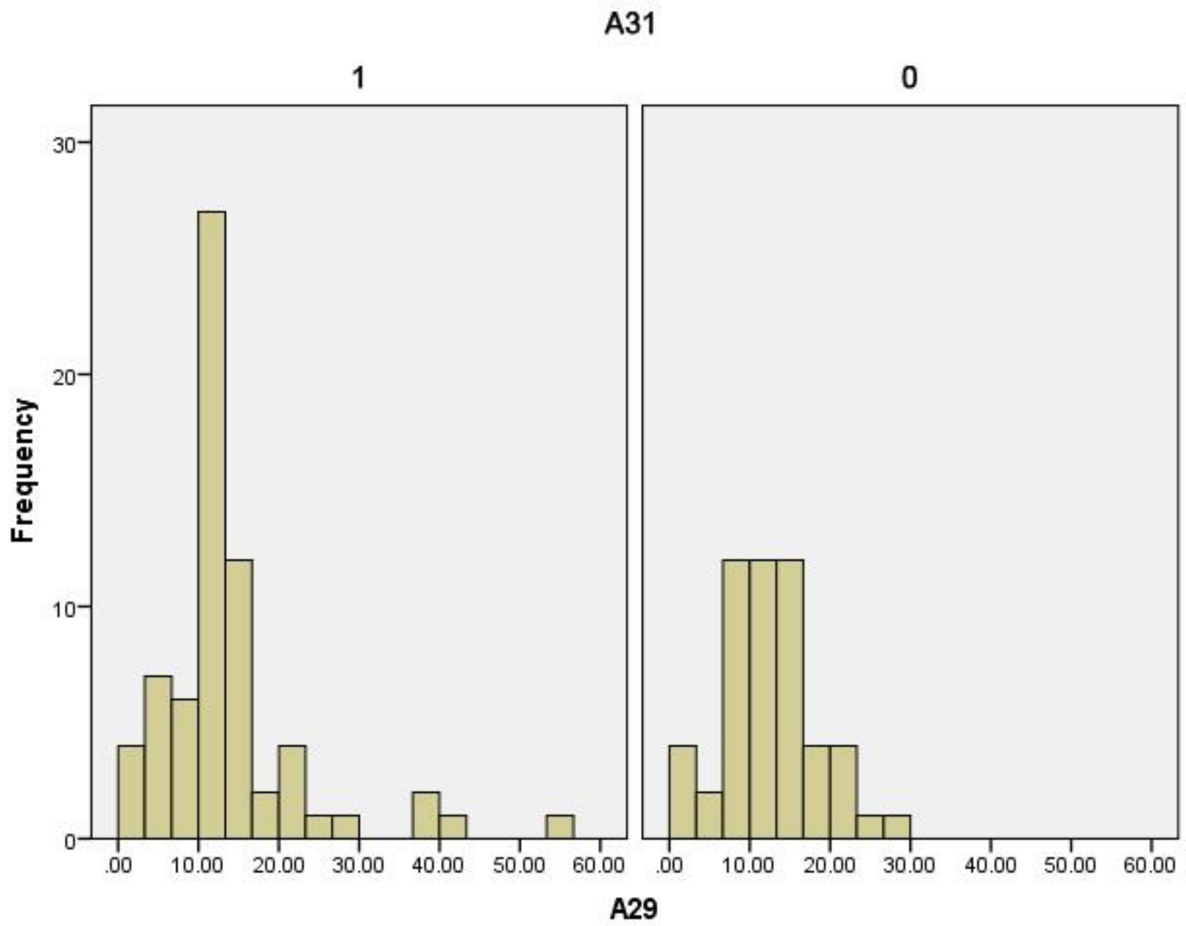


Figure 31: Graph for average time taken between attempts on questions

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A29	68	2.42	53.56	13.7437	9.08181
	Valid N (listwise)	68				
0	A29	52	1.80	28.53	12.6183	5.72319
	Valid N (listwise)	52				

Table 33: Descriptive statistics for average time taken between attempts on questions

Comparison:

The mean and standard deviation of the reliable response category are higher than the unreliable response category. The minimum and maximum value of the reliable response category is higher than the unreliable response category.

A30) Number of responded questions

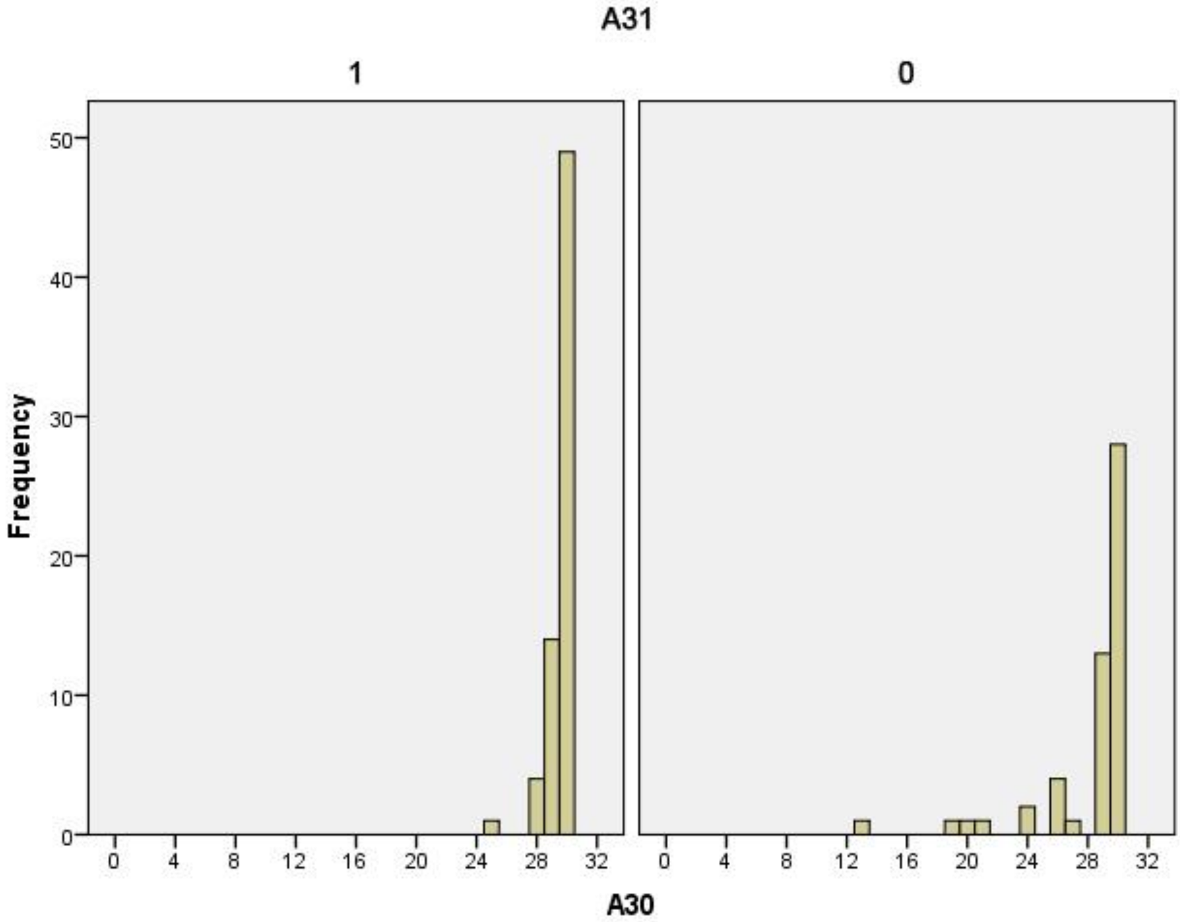


Figure 32: Graph for number of responded questions

Descriptive Statistics

A31		N	Minimum	Maximum	Mean	Std. Deviation
1	A30	68	25	30	29.60	.813
	Valid N (listwise)	68				
0	A30	52	13	30	28.25	3.401
	Valid N (listwise)	52				

Table 34: Descriptive statistics for number of responded questions

Comparison:

The mean of the reliable response category is higher than the unreliable response category. However the mean of both the categories are close to each other. The standard deviation of the unreliable response category is higher than the reliable response category with a significant difference.

Table 35 explains the attributes that have been converted to a binary format as a result of the descriptive analysis.

Attribute ID	**Previous**	**Present**
A4	Total middle button clicks	Middle click status enabled
A11	Page visited count	Page revisited status
A12	Pasted text count	Text pasted status
A15	The total time spent for out of page	The status of out of page
A16	The total idle time	The status of idle
A18	Answer changes count	Answer changes during the survey
A19	No of time changes on text areas	Text areas changes during the survey
A20	No of time changes on drop downs	Drop down during the survey
A21	No of time changes on text fields	Text field changes during the survey
A26	No of time changes on radio buttons	Radio button changes during the survey
A28	No of time changes on check boxes	Checkbox changes during the survey

Table 35: Converted Attributes

3.4 Feature Selection and Feature Ranking

Feature selection and ranking mechanism are used to identify the significant attributes among all the attributes (30) of the dataset. They aim to remove redundant and irrelevant features so that classification of new instances will be more accurate. Based on the selection done by users, it will be categorised as reliable or unreliable response. This will be taken as the categorical variable. The other 30 attributes are taken as the predictor variables. Weka and R programs are used to select and rank the attributes.

Feature Selection

Feature selection is used to select the most significant attributes among all the attributes.

Correlation Feature Selection

The CFS Subset Evaluator is used to select the attributes.

Attribute Evaluator	Attribute Name
CFS Subset Evaluator	A5,A30

Table 36: Selected attributes using CFS Subset Evaluator

Forward Feature Selection.

Forward feature selection is a mechanism in which features keeps appending from null unless and until the addition of a feature does not make a positive change in the model.

Attribute Evaluator	Attribute Name
Forward feature selection	A23,A9,A16,A30,A24

Table 37: Selected attributes using Forward Feature Selection

The relevant code is added to the appendix G.

Feature Ranking

Feature Ranking is used to rank the most significant attributes in order. Classifier Attribute Evaluator, Information Gain Attribute Evaluator, Correlation Attribute Evaluator, Relief Attribute Evaluator and Symmetrical Uncertainty Attribute Evaluator are used to rank the attributes.

The attributes that were ranked with the help of these evaluators are given in the table 38.

Attribute Evaluator	**Attribute Name**
Classifier Attribute Evaluator	A30,A11,A10,A12,A8,A13,A9,A7,A29,A3,A2,A4,A6,A5,A14,A15, A16,A26,A25,A27,A17,A28,A24,A23,A22,A21,A18,A19,A20,A1
Information Gain Attribute Evaluator	A30,A5,A8,A10,A11,A12,A13,A9,A7,A29,A6,A2,A3,A4,A14,A15, A16,A26,A25,A27,A17,A28,A24,A23,A22,AA21,A18,A19,A20,A1
Correlation Attribute Evaluator	A30,A5,A22,A23,A18,A16,A9,A26,A6,A14,A13,A19,A25,A29, A10,A17,A8,A7,A2,A3,A11,A24,A20,A21,A28,A4,A27,A12,A1,A15
Relief Attribute Evaluator	A23,A9,A16,A30,A24,A11,A2,A27,A5,A1,A7,A28,A6,A25,A12, A29,A17,A13,A22,A4,A8,A26,A3,A14,A10,A15,A21,A19,A18,A20
Symmetrical Uncertainty Attribute Evaluator	A30,A5,A8,A10,A11,A12,A13,A9,A7,A29,A6,A2,A3,A4,A14,A15, A16,A26,A25,A27,A17,A28,A24,A23,A22,A21,A18,A19,A20,A1

Table 38: Ranked attributes using attribute evaluators

With the help of feature selection and ranking, some attribute combinations are selected to predict the reliable and unreliable response using classifier algorithms. Table 39 gives the selected attributes.

Attribute Set Name	**Selected Attributes**
Attribute Set 1	All (A1-A30)
Attribute Set 2	A5,A30
Attribute Set 3	A30,A11,A10,A12,A8,A13,A9,A7,A29,A3
Attribute Set 4	A30,A11,A10,A12,A8
Attribute Set 5	A30,A5,A8,A10,A11,A12,A13,A9,A7,A29
Attribute Set 6	A30,A5,A8,A10,A11
Attribute Set 7	A30,A5,A22,A23,A18,A16,A9,A26,A6,A14
Attribute Set 8	A30,A5,A22,A23,A18
Attribute Set 9	A23,A9,A16,A30,A24,A11,A2,A27,A5,A1
Attribute Set 10	A23,A9,A16,A30,A24
Attribute Set 11	A5,A8,A9,A10,A11,A16,A22,A30
Attribute Set 12	A4,A16,A23,A25,A30

Table 39: Selected Attribute Sets

3.4.1 Identifying Reliable and Unreliable Response Using Classifier Algorithms

There are several algorithms to create a prediction model. This research uses five different algorithms: Logistic Regression, Decision Tree, Support Vector Machine, Random Forest and Naïve Bayes. All these algorithms are used to perform the same task, which is predicting a categorical variable based on predictor variables, using different mathematical methods.

Models are created to predict the categorical values. There are 30 predictor variables and a categorical variable. Categorial variable is taken through the answer given to the self reporting question "Are you instructed to fill the questionnaire in a serious manner?'. If the answer is yes then it is considered as a reliable response and if the answer is no it is considered as an unreliable response. Models are used to predict the type of response (reliable or unreliable response) for a given set of predictor variables.

The predictor variables are divided as per the results obtained from feature selection and ranked into twelve sets. Five algorithms chosen are applied to every set mentioned above. 70% of the instances are used for training while the remaining 30% is used for testing.

The classifier models come up with the results for the 30% data, based on the knowledge grasped with the training data. The number of correctly and incorrectly predicted instances are identified with the help of confusion matrix which is produced by each classifier model.

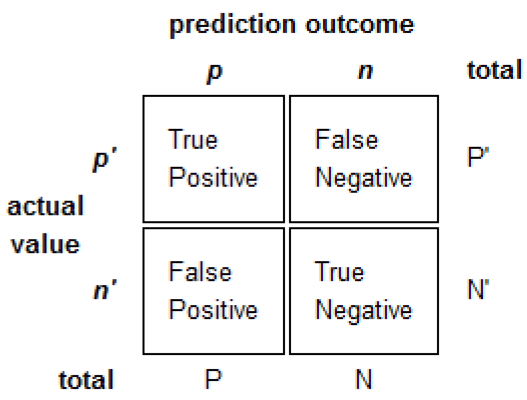


Figure 33: Confusion Matrix

Each classifier model produces the evaluation measures for each attribute set with the help of a confusion matrix. They are Accuracy, Precision, Recall and F-Measure.

TN (True Negative) : Number of correct predictions that an instance is irrelevant

FP (False Positive) : Number of incorrect predictions that an instance is relevant

FN (False Negative) : Number of incorrect predictions that an instance is irrelevant

TP (True Positive) : Number of correct predictions that an instance is relevant

Accuracy (ACC) – The proportion of the total number of predictions that were correct:
Accuracy (%) = (TN+TP)/(TN+FN+FP+TP)

Precision (PREC) – The proportion of the predicted relevant materials data sets that were correct:
Precision (%) = T/(FP+TP)

Recall (REC) – The proportion of the relevant materials data sets that were correctly identified:
Recall (%) = TP/(FN+TP)

F-Measure (FM) – Derives from precision and recall values:
F-Measure (%) = (2x RECxPREC)/(REC+PREC)

The highest accuracy of the model for the attribute set is identified and its attribute set is chosen as the best attribute set for the particular model. Likewise, the highest accuracy and the best attribute set are identified for each classifier model. Then, the highest accuracies and the best attribute sets are compared among all the five algorithms. Finally, the model which got the highest accuracy will be chosen as the best model to predict the reliable and unreliable response of the online survey data. And the attribute set which gets the highest accuracy will be taken as the best attribute set for the particular model. Also, the number of times the individual attribute is presented in entire selected attribute set for the classifier models is considered and that will be taken as the highly influencing attribute to predict the reliable and unreliable responses of the online surveys.

By considering the evaluation purpose, categorical variable is taken based on correct answers. If user has answered twenty-eight or more questions correctly out of thirty questions the user is considered as a reliable user and otherwise, an unreliable one. The classification process will be carried out for above mentioned categorical variable as well. Finally both the results, response based on self reporting and correct answers will be compared to predict the reliability of the response.

3.4.2 Identifying Reliable and Unreliable Response Using Outlier Values

The following methods are used to identify the reliable and unreliable response in online survey.

a) The identification of outlier values for the attributes

An outlier is an observation point that is distant from other observations in a data set. They are important to keep in mind when looking at pools of data because they can sometimes affect how the data is perceived on the whole.

In this research, outlier values of an attribute or mixed attributes were identified from the dataset. The following techniques were used to identify the outlier values from the attributes in the dataset.

1) Mean and Standard Deviation Method (SD)
2) Median and Interquartile Deviation Method (IQD)
3) Median Absolute Deviation Method (MAD)
4) Mahanobalis Distance Method (MD)

SD, IQD, and MAD Methods were used to identify the outlier values from individual attributes. Mahanobalis Distance Method was used to identify the outlier values from mixed attributes.

Once the data for an attribute or mixed attribute is checked for outlier detection, the obtaining level for that attribute or mixed attribute is assigned as;
a) Negative Influence on attribute or mixed attribute
b) No Negative Influence on attribute or mixed attribute

If the data for the attribute or mixed attribute identifies as an outlier, that data is considered as "negative influence" on that attribute or mixed attribute.
If the data for the attribute or mixed attribute does not identify as an outlier, that data is considered as "no negative influence".

With the help of the obtained level, the following value is assigned to the participant for that particular attribute.

a) 1- Negative Influence on attribute
b) 0 - No Negative Influence on attribute

1) Mean and Standard Deviation Method (SD)

For this outlier detection method, the mean and standard deviation of the residuals are calculated and compared. If a value is a certain number of standard deviations away from the mean, that data point is identified as an outlier.

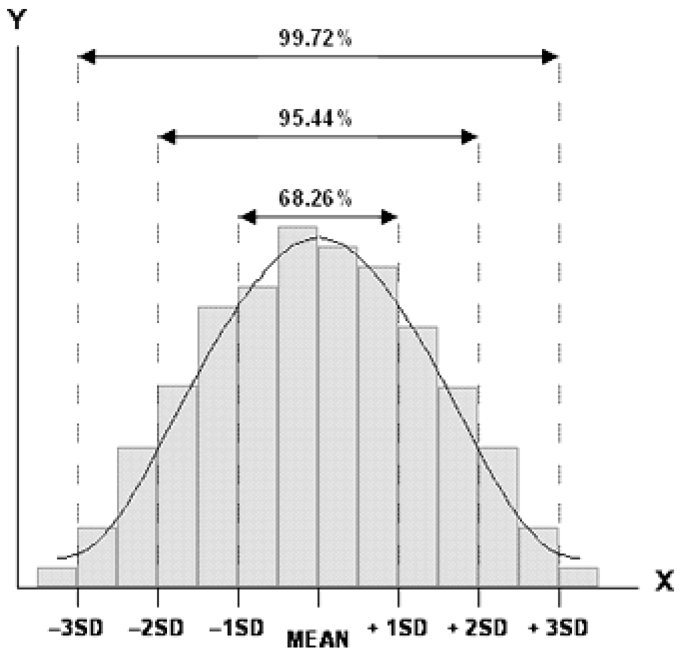


Figure 34: Mean and standard deviation method

Threshold values for the analysis are given below for SD method.

Upper Threshold: +2SD

Lower Threshold: -2SD

In this research, the attribute data which is away from a number of standard deviations from the mean or away from the Lower Threshold from the mean is taken as outlier value for that attribute. If the data falls above mentioned range, the particular user ID is coded as “1” for that attribute. If the data does not fall above mentioned range, the particular user ID is coded as “0” for that attribute.

2) Median and Interquartile Deviation Method (IQD)

For this outlier detection method, the median of the residuals is calculated, along with the 25th percentile (Q1) and the 75th percentile (Q3). The difference between the 25th and 75th percentile is the interquartile deviation (IQR). Then, the difference is calculated between each historical value and the residual median. If the historical value is a certain number of IQD away from the median of the residuals, that value is classified as an outlier. Box plots are based on this approach.

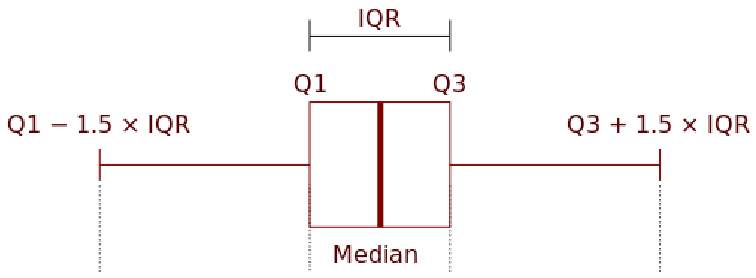


Figure 35: IQD method

IQR = Q3 − Q1

Threshold values for the analysis is given below for IQD method.

Upper Threshold: Q3+1.5 * IQR

Lower Threshold: Q1-1.5 * IQR

In this research, the attribute data which is away from Upper Threshold and beyond the Median or away from Lower Threshold and from the further side of the Median is taken as the outlier value for that attribute. If the data falls above mentioned range, the particular user ID is coded as “1” for that attribute. If the data does not fall above mentioned range, the particular user ID is coded as “0” for that attribute.

3) Median Absolute Deviation Method (MAD)

For this outlier detection method, the median of the residuals is calculated. Then, the difference is calculated between each historical value and this median. These differences are expressed as their absolute values, and a new median is calculated and multiplied by an empirically derived constant to yield the median absolute deviation (MAD). If a value is a certain number of MAD away from the median of the residuals, that value is classified as an outlier.

MAD = median(|Xi – median(Xi|)

Threshold values for the analysis is given below for MAD method.

Upper Threshold: Median + 2.5* MAD

Lower Threshold: Median + 2.5* MAD

In this research, the attribute data which is away from Upper Threshold from the Median or away from Lower Threshold from the Median is taken as outlier values for that attribute. If the data falls above mentioned range, the particular user ID is coded as "1" for that attribute. If the data does not fall above mentioned range, the particular user ID is coded as "0" for that attribute.

30 individual attributes from the dataset were selected and checked to find out the outlier values with the help of SD, IQD, and MAD Methods. The same attributes were checked for each method. (A1-A30 attributes were selected for the analysis – Attribute details are given in the attribute table in the data collection section).

4) Mahalanobis Distance

The Mahalanobis distance (MD) is the distance between two points in multivariate space. The Mahalanobis distance is a measure of the distance between a point P and a distribution D, introduced by P. C. Mahalanobis in 1936. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D. It is an appropriate method for use with survey data, a measure of the multivariate distance between an individual's response vector and the average response vector for all participants who took the questionnaire [11].

The Mahalanobis D technique is dependent on the characteristics of the sample, as it compares individual response vectors to sample mean response vectors. As a result, it is best suited for the identification of random and extreme response styles, as these tendencies are associated with increased response variation when compared with acquiescent or consistent responders. MD only uses independent variables in its calculations. By considering this, 8 mixed attribute sets were selected and checked to find out the outlier values.

Table 40 explains the selected mixed attributes for the outlier detection using MD method.

Mixed Attribute Set	**Mixed Attributes**
Mixed Attribute Set 1	A2,A3,A4
Mixed Attribute Set 2	A5,A6
Mixed Attribute Set 3	A19,A20,21,A22,A23
Mixed Attribute Set 4	A9,A10
Mixed Attribute Set 5	A24,A25,A26,A27,A28
Mixed Attribute Set 6	A15,A16,A17
Mixed Attribute Set 7	A12,A13,A14
Mixed Attribute Set 8	A1,A29
Mixed Attribute Set 9	A7,A8,A11,A18,A30

Table 40: Mixed attributes for the outlier detection using MD method

Initially, the MD values for mixed attribute sets are calculated separately. Then for each MD value the chi-square value P is calculated. If P value is lesser than 0.001, the attribute set is considered as an outlier.

Finally, based on the outliers, the following method is used to identify the reliability of response.

Method

Outliers are identified for the total number of negative influences of a response. If the total number of negative influences identifies as an outlier, that response is considered as an "unreliable" response.

In order to achieve this, the finalized dataset is run using R environment. SD, IQD, MAD and MD methods are used in R to get the outliers for the attributes and attribute sets. Finally, a CSV is produced with outlier details and types of response. (reliable or unreliable) according to the outlier detection methods and the defined threshold values for each method.

3.5 Chapter Summary

This chapter summarizes the data collection and data analysis methods for the proposed research. Survey selection, tool creation with capturing features, finalizing the dataset attributes and data analysis methods are explained in detail in the above mentioned chapters. The following chapter will reflect the results of the analysis performed.

4 Results and Evaluations

4.1 Chapter Introduction

This chapter elaborates on the results of the data analysis obtained from the classifier models chosen. The above mentioned results are evaluated based on the standard metrics of accuracy, precision, recall and F-measure. The most relevant attribute set and classifier model in predicting the reliability of an online survey are also identified. The results from outlier values are considered as well.

4.2 Results of the Classifier Models

Reliable and unreliable responses are classified using the classifier models, Naïve Bayes, Random Forest, Decision Tree, Support Vector Machine and Logistic Regression.

The results of the feature selection, that is, the twelve attribute sets obtained are used as the predictor variables and passed on to each of these above mentioned algorithms.

4.2.1 Results of the Classifier Models Based on Self-Reporting

Categorical variable taken from the response, based on self reporting (reliable or unreliable) are taken into consideration to predict the results.

Table 41 contains details of the training and testing data of the classifier models.

Classifier Models	Naïve Bayes Support Vector Machine Logistic Regression Decision Tree Random Forest
Total Instances	120
Training (70%)	84
Testing (30%)	36
Number of Unreliable Response in the Testing Data	15
Number of Reliable Response in the Testing Data	21

Table 41: Training and testing data details - based on self-reporting

The results based on each algorithm for the chosen attribute sets are given in the following tables (42-46). The attribute set with the highest accuracy as given by the algorithm is considered as the best attribute set.

Naïve Bayes

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	Unreliable Response	Reliable Response	Unreliable Response	
All (A1-A30)	14	8	7	7	61.1111 %
A5,A30	21	3	12	0	66.6667 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	15	5	10	6	55.5556 %
A30,A11,A10,A12,A8	17	3	12	4	55.5556 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	16	5	10	5	58.3333 %
A30,A5,A8,A10,A11	18	3	12	3	58.3333 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	19	5	10	2	66.6667 %
A30,A5,A22,A23,A18	21	5	10	0	72.2222 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	19	5	10	2	66.6667 %
A23,A9,A16,A30,A24	18	5	10	3	63.8889 %
A5,A8,A9,A10,A11, A16,A22,A30	16	4	11	5	55.5556 %
A4,A16,A23,A25,A30	20	5	10	1	69.4444 %

Table 42: Naïve Bayes - Results based on self-reporting

The best attribute set based on Naive Bayes algorithm is A30,A5,A22,A23,A18

Support Vector Machine

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	Unreliable Response	Reliable Response	Unreliable Response	
All (A1-A30)	14	7	8	7	58.3333 %
A5,A30	21	0	15	0	58.3333 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	19	1	14	2	55.5556 %
A30,A11,A10,A12,A8	21	0	15	0	58.3333 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	17	1	14	4	50 %
A30,A5,A8,A10,A11	21	1	14	0	61.1111 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	17	5	10	4	61.1111 %
A30,A5,A22,A23,A18	19	3	12	2	61.1111 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	19	1	14	2	55.5556 %
A23,A9,A16,A30,A24	19	2	13	2	58.3333 %
A5,A8,A9,A10,A11, A16,A22,A30	19	3	12	2	61.1111 %
A4,A16,A23,A25,A30	21	0	15	0	58.3333 %

Table 43: Support Vector Machine - Results based on self-reporting

The best attribute sets using svm are A30,A5,A8,A10,A11 and A30,A5,A22,A23,A18.

There are four attribute sets, which produced the same accuracy. However, two of these have been eliminated considering the larger number of attributes that had to be used in order to arrive at this accuracy.

Logistic Regression

Selected Attributes	**Correctly Classified as**		**Incorrectly Classified as**		**Accuracy**
	Reliable Response	**Unreliable Response**	**Reliable Response**	**Unreliable Response**	
All (A1-A30)	15	9	6	6	66.6667 %
A5,A30	19	3	12	2	61.1111 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	14	7	8	7	58.3333 %
A30,A11,A10,A12,A8	11	7	8	10	50 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	12	7	8	9	52.7778 %
A30,A5,A8,A10,A11	11	7	8	10	50 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	15	6	9	6	58.3333 %
A30,A5,A22,A23,A18	17	7	8	4	66.6667 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	13	9	6	8	61.1111 %
A23,A9,A16,A30,A24	15	6	9	6	58.3333 %
A5,A8,A9,A10,A11, A16,A22,A30	14	9	6	7	63.8889 %
A4,A16,A23,A25,A30	15	11	4	6	72.2222 %

Table 44: Logistic Regression - Results based on self-reporting

The best attribute set using logistic regression is A4,A16,A23,A25,A30.

Decision Tree

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	Unreliable Response	Reliable Response	Unreliable Response	
All (A1-A30)	7	6	9	14	36.1111 %
A5,A30	21	0	15	0	58.3333 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	19	3	12	2	61.1111 %
A30,A11,A10,A12,A8	19	3	12	2	61.1111 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	15	7	8	6	61.1111 %
A30,A5,A8,A10,A11	15	7	8	6	61.1111 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	14	3	12	7	47.2222 %
A30,A5,A22,A23,A18	14	3	12	7	47.2222 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	13	6	9	8	52.7778 %
A23,A9,A16,A30,A24	21	3	12	0	66.6667 %
A5,A8,A9,A10,A11, A16,A22,A30	11	4	11	10	41.6667 %
A4,A16,A23,A25,A30	20	3	12	1	63.8889 %

Table 45: Decision Tree - Results based on self-selection

The best attribute set using decision tree is A23,A9,A16,A30,A24.

Random Forest

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	**Unreliable Response**	**Reliable Response**	**Unreliable Response**	
All (A1-A30)	10	9	6	11	52.7778 %
A5,A30	10	6	9	11	44.4444 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	14	9	6	7	63.8889 %
A30,A11,A10,A12,A8	12	8	7	9	55.5556 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	12	8	7	9	55.5556 %
A30,A5,A8,A10,A11	7	9	6	14	44.4444 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	7	9	6	14	44.4444 %
A30,A5,A22,A23,A18	8	8	7	13	44.4444 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	12	8	7	9	55.5556 %
A23,A9,A16,A30,A24	11	9	6	10	55.5556 %
A5,A8,A9,A10,A11, A16,A22,A30	5	7	8	16	33.3333 %
A4,A16,A23,A25,A30	12	8	7	9	55.5556 %

Table 46: Random Forest - Results based on self-selection

The best attribute set based on Random Forest is A30,A11,A10,A12,A8,A13,A9,A7,A29,A3.

Table 47 compares the highest accuracies of each algorithm chosen.

Selected Algorithm	**Attributes**	**Accuracy**
Naïve Bayes	A30, A5, A22, A23, A18	72.2222%
SVM	A30,A5,A8,A10,A11	61.1111%
SVM	A30,A5,A22,A23,A18	61.1111%
Logistic Regression	A4,A16,A23,A25,A30	72.2222%
Decision Tree	A23,A9,A16,A30,A24	66.6667%
Random Forest	A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	63.8889%

Table 47: Highest accuracies of each algorithm - based on self-reporting

Naïve Bayes and Logistic regression give the highest accuracies in deriving at attributes that are the most influential in finding out the reliability of responses in online surveys.

Hence, as per the analysis mentioned above the attributes are categorised as high influential, moderately influential and low influential in identifying the reliability of responses in online surveys. If an individual attribute that is presented the most frequent number of times in above chosen attribute sets as shown in table 47, that individual attribute is considered as more significant ones compared to the rest of the attributes. For instance, Attribute A30 is presented in every attribute set for the algorithms.

Table 48 gives the attributes and their levels of influence.

High influential in identifying the reliability of response	Total answer clicks, Checkbox changes during the survey and Number of responded questions.
Moderate influential in identifying reliability of response	Number of time scrollbar reached bottom, Answer changes during the survey, Radio button changes during the survey, Page revisited status, Average time spent between clicks, The status of idle and Number of times scrollbar reached top

Low influential in identifying the reliability of response	Total middle button clicks, Average time spent on drop down questions, Average time spent on text area questions, Text pasted status, Pasted character count, Total clicks, Average time takes between attempts on questions and Total right button clicks

Table 48: Attributes and their levels of influence - based on self-reporting

4.2.2 Results of the Classifier Models Based on the Correct Answers

Categorical variable taken from the responses, based on correct answers (reliable or unreliable) are taken into consideration to predict the results.

If the user's correct number of answers is above or equal to 28 that user's response will be considered as a reliable response, else it will be considered as an unreliable response.

Table 49 contains details of the training and testing data of the classifier models.

Classifier Models	Naïve Bayes Support Vector Machine Logistic Regression Decision Tree Random Forest
Total Instances	120
Training (70%)	84
Testing (30%)	36
Number of Unreliable Users in the Testing Data	26
Number of Reliable Users in the Testing Data	10

Table 49: Training and testing data details - based on correct answers

The results based on each algorithm for the chosen attribute sets are given in the following tables (50-54).

Naïve Bayes

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	Unreliable Response	Reliable Response	Unreliable Response	
All (A1-A30)	9	12	14	1	58.3333 %
A5,A30	10	14	12	0	66.6667 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	8	16	10	2	66.6667 %
A30,A11,A10,A12,A8	9	12	1	14	58.3333 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	8	14	12	2	61.1111 %
A30,A5,A8,A10,A11	9	13	13	1	61.1111 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	10	10	16	0	55.5556 %
A30,A5,A22,A23,A18	10	14	12	0	66.6667 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	8	14	12	2	61.1111 %
A23,A9,A16,A30,A24	8	17	9	2	69.4444 %
A5,A8,A9,A10,A11, A16,A22,A30	8	16	10	2	66.6667 %
A4,A16,A23,A25,A30	9	15	11	1	66.6667 %

Table 50: Naïve Bayes - Results based on correct answers

The best attribute set based on Naïve Bayes algorithm is A23,A9,A16,A30,A24.

Support Vector Machine

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	Unreliable Response	Reliable Response	Unreliable Response	
All (A1-A30)	4	22	4	6	72.2222 %
A5,A30	0	26	0	10	72.2222 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	0	26	0	10	72.2222 %
A30,A11,A10,A12,A8	0	26	0	10	72.2222 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	0	26	0	10	72.2222 %
A30,A5,A8,A10,A11	0	26	0	10	72.2222 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	0	26	0	10	72.2222 %
A30,A5,A22,A23,A18	0	26	0	10	72.2222 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	0	26	0	10	72.2222 %
A23,A9,A16,A30,A24	0	26	0	10	72.2222 %
A5,A8,A9,A10,A11, A16,A22,A30	0	26	0	10	72.2222 %
A4,A16,A23,A25,A30	0	26	0	10	72.2222 %

Table 51: Support Vector Machine - Results based on correct answers

The accuracies found out using Support Vector Machine are the same for all the attribute sets. Moreover, it did not identify any of the unreliable responses. Hence, this algorithm is not considered for further analysis.

Logistic Regression

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	Unreliable Response	Reliable Response	Unreliable Response	
All (A1-A30)	5	14	12	5	52.7778 %
A5,A30	0	26	0	10	72.2222 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	2	18	8	8	55.5556 %
A30,A11,A10,A12,A8	2	22	4	8	66.6667 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	3	17	9	7	55.5556 %
A30,A5,A8,A10,A11	1	26	0	9	75 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	6	22	4	4	77.7778 %
A30,A5,A22,A23,A18	4	26	0	6	83.3333 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	1	17	9	9	50 %
A23,A9,A16,A30,A24	2	22	4	8	66.6667 %
A5,A8,A9,A10,A11, A16,A22,A30	6	22	4	4	77.7778 %
A4,A16,A23,A25,A30	1	23	3	9	66.6667 %

Table 52: Logistic Regression - Results based on correct answers

The best attribute set using Logistic Regression **is** A30,A5,A22,A23,A18.

Decision Tree

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	Unreliable Response	Reliable Response	Unreliable Response	
All (A1-A30)	9	8	18	1	47.2222 %
A5,A30	0	26	0	10	72.2222 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	0	26	0	10	72.2222 %
A30,A11,A10,A12,A8	0	26	0	10	72.2222 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	0	26	0	10	72.2222 %
A30,A5,A8,A10,A11	0	26	0	10	72.2222 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	8	21	5	2	80.5556 %
A30,A5,A22,A23,A18	0	26	0	10	72.2222 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	2	23	3	8	69.4444 %
A23,A9,A16,A30,A24	2	26	0	8	77.7778 %
A5,A8,A9,A10,A11, A16,A22,A30	3	22	4	7	69.4444 %
A4,A16,A23,A25,A30	6	24	2	4	83.3333 %

Table 53: Decision Tree - Results based on correct answers

The best attribute set using Decision Tree algorithm is A4,A16,A23,A25,A30

Random Forest

Selected Attributes	Correctly Classified as		Incorrectly Classified as		Accuracy
	Reliable Response	Unreliable Response	Reliable Response	Unreliable Response	
All (A1-A30)	6	24	2	4	83.3333 %
A5,A30	4	16	10	6	55.5556 %
A30,A11,A10,A12,A8, A13,A9,A7,A29,A3	4	22	4	6	72.2222 %
A30,A11,A10,A12,A8	3	6	20	7	63.8889 %
A30,A5,A8,A10,A11, A12,A13,A9,A7,A29	5	19	7	5	66.6667 %
A30,A5,A8,A10,A11	4	18	8	6	61.1111 %
A30,A5,A22,A23,A18, A16,A9,A26,A6,A14	4	22	4	6	72.2222 %
A30,A5,A22,A23,A18	4	20	6	6	66.6667 %
A23,A9,A16,A30,A24, A11,A2,A27,A5,A1	3	20	6	7	63.8889 %
A23,A9,A16,A30,A24	4	21	5	6	69.4444 %
A5,A8,A9,A10,A11, A16,A22,A30	3	19	7	7	61.1111 %
A4,A16,A23,A25,A30	8	19	7	2	75 %

Table 54: Random Forest - Results based on correct answers

The best attribute set using Decision Tree algorithm is All (A1-A30).

Table 55 compares the highest accuracies of each algorithm chosen.

Selected Algorithm	**Attributes**	**Accuracy**
Naïve Bayes	A23,A9,A16,A30,A24.	69.4444 %
Logistic Regression	A30,A5,A22,A23,A18	83.3333 %
Decision Tree	A4,A16,A23,A25,A30	83.3333 %
Random Forest	All (A1-A30).	83.3333 %

Table 55: Highest accuracies of each algorithm - based on correct answers

Decision Tree, Random Forest and Logistic Regression give the highest accuracies in deriving at attributes that are the most influential in finding out the reliability of response in online survey.

Hence, as per the analysis mentioned above the attributes are categorised as high influential, moderate influential and low influential in identifying the reliability of response in online surveys.

The table 56 gives the attributes and their levels of influence.

High influential in identifying the reliability of response	Checkbox changes during the survey, Number of responded questions and The status of idle.
Moderate influential in identifying the reliability of response	Number of time scrollbar reached bottom, Average time spent on text area questions, Total clicks, Radio button changes during the survey, Answer changes during the survey, Total middle button clicks and Average time spent on drop down questions

Less influential in identifying the reliability of response	Average time spent per question, Total left button clicks,Total right button clicks, Total non-answer clicks, Total clicks, Average time spent between clicks, Number of time scrollbar reached bottom, Page revisited status, Text pasted status, Pasted character count, Average character per minute (cpm) for typing, The status of out of page, Average time spent per page, Text areas changes during the survey, Drop down during the survey, Text field changes during the survey, Average time spent on radio button, Average time spent on text field questions, Average time spent on check box questions and Average time taken between attempts on questions

Table 56: Attributes and their levels of influence - based on correct answers

4.3 Evaluation of the Results from Classifier Models

Table 57 compares the results obtained from self-reporting and correct answers to identify attributes' levels of significance.

Attribute Name	Response Category Based on Self-Reporting	Response Category Based on Correct Answers
Average time spent per question	Insignificant	Low
Total left button clicks	Insignificant	Low
Total right button clicks	Low	Low
Total middle button clicks	Low	Moderate
Total answer clicks	High	Moderate
Total non-answer clicks	Insignificant	Low
Total clicks	Low	Moderate
Average time spent between clicks	Moderate	Low
Number of time scrollbar reached top	Moderate	Low
Number of time scrollbar reached bottom	Moderate	Moderate
Page revisited status	Moderate	Low
Text pasted status	Low	Low
Pasted character count	Low	Low
Average character per minute (cpm) for typing	Insignificant	Low
The status of out of page	Insignificant	Low
The status of idle	Moderate	High
Average time spent per page	Insignificant	Low
Answer changes during the survey	Moderate	Moderate
Text areas changes during the survey	Insignificant	Low
Drop down during the survey	Insignificant	Low
Text field changes during the survey	Insignificant	Low
Radio button changes during the survey	Moderate	Moderate
Checkbox changes during the survey	High	High

Average time spent on text area questions	Low	Moderate
Average time spent on drop down questions	Low	Moderate
Average time spent on radio button questions	Insignificant	Low
Average time spent on text field questions	Insignificant	Low
Average time spent on check box questions	Insignificant	Low
Average time taken between attempts on questions	Low	Low
Number of responded questions	High	High

Table 57: Attributes - Levels of significance

The attributes that denote the checkbox changes during the survey and the number of responded questions prove to be the most influential attributes in both the methods.

The algorithms are evaluated using its accuracy, precision, recall and F-Measure with the help of confusion matrix.

Table 58 shows the accuracy, precision, recall and F-measure of attribute sets using user category self-reporting method.

Selected Algorithms	**Attributes**	**Accuracy**	**Precision**	**Recall**	**F-Measure**
Naïve Bayes	A30,A5,A22, A23, A18	72.2222%	0.812	0.722	0.679
SVM	A30,A5,A8, A10,A11	61.1111%	0.767	0.611	0.490
SVM	A30,A5,A22, A23,A18	61.1111%	0.608	0.611	0.551
Logistic Regression	A4,A16,A23, A25,A30	72.2222%	0.730	0.722	0.724
Decision Tree	A23,A9,A16, A30,A24	66.6667%	0.788	0.667	0.593
Random Forest	A30,A11,A10, A12,A8,A13,A9, A7,A29,A3	63.8889%	0.643	0.639	0.640

Table 58: Evaluation meaures - based on self-reporting

Table 59 shows the accuracy, precision, recall and F-measure of attribute sets using response category correct answer method.

Selected Algorithms	**Attributes**	**Accuracy**	**Precision**	**Recall**	**F-Measure**
Naïve Bayes	A23,A9,A16, A30,A24	69.44 %	77.70 %	69.40%	71.00%
Logistic Regression	A30,A5,A22, A23,A18	83.33 %	86.50 %	83.30%	80.60%
Decision Tree	A4,A16,A23, A25,A30	83.33 %	82.70%	83.30%	82.70%
Random Forest	All (A1-A30)	83.33 %	82.70%	83.30%	82.70%

Table 59: Evaluation meaures - based on self-reporting

All the algorithms provide an accuracy of greater than 50%.

According to the results obtained in self-reporting method, Logistic Regression and Naïve Bayes are the most significant algorithms giving the same accuracy.

According to the results obtained in correct answer method, Logistic Regression, Decision Tree and Random Forest are equally significant giving the same accuracy.

Hence, based on the overall results, Logistic Regression model is the most suitable to predict the reliability of responses in online surveys.

The attributes in the logistic alogorithm’s attribute set based on the correct answers, that showed the highest accuracy, includes number of responded questions, total answer clicks, radio button changes during the survey, checkbox changes during the survey and answer changes during the survey.

The attributes in the logistic alogorithm’s attribute set based on self-reporting, that showed the highest accuracy, includes total middle button clicks, the status of idle, checkbox changes during the survey, average time spent on drop down questions, number of responded questions and average time spent on text area questions.

The result screens are added to the appendix H.

4.4 Results from Outlier Detection Method

Based on the total outliers which were detected by SD, IQD, MAD and MD methods, the type of responses are identified. In order to achieve this, Standard Deviation method was used for total outliers. If the total number of outliers or the total number of negative influences is identified as outlier, the particular response will be identified as an unreliable response. If not, the particular response will be identified as a reliable response.

The field names of the result table are given in table 60.

ID	**Description**
C1	User ID
C2	The total number of outliers of a participant identified by Standard Deviation Method (SD) for 30 attributes.
C3	The total number of outliers of a participant identified by Interquartile Deviation Method (IQD) for 30 attributes.
C4	The total number of outliers of a participant identified by Median Absolute Deviation Method (MAD) for 30 attributes.
C5	The total number of outliers of a participant identified by Mahalanobis Distance Method for 8 attribute sets.
C6	The total number of outliers or the total number of negative influences.
C7	The type of response (reliable response-1, unreliable response-0) which are identified by outlier detection method.
C8	The user category marked by participant (reliable response -1, unreliable response -0) when filling the questionnaire.
C9	The response category based on correct answers (reliable response -1, unreliable response -0)

Table 60: Fields in the results table

The result which were arrived from the statistical data analysis is given in table 61. (descriptions of the title are given in table 60)

C1	C2	C3	C4	C5	C6	C7	C8	C9
1	3	4	6	1	14	1	1	0
2	1	1	2	0	4	1	1	1
3	1	1	1	0	3	1	1	0
4	0	0	2	0	2	1	0	0
5	2	2	3	1	8	1	1	1
6	2	2	8	1	13	1	1	0
7	2	2	6	0	10	1	1	0
8	0	0	1	0	1	1	1	0
9	0	0	0	0	0	1	1	0
10	5	5	6	3	19	1	1	0
11	1	2	6	0	9	1	0	0
12	0	0	2	0	2	1	0	0
13	0	0	1	0	1	1	1	0
14	1	1	3	0	5	1	0	0
15	1	1	2	0	4	1	0	1
16	5	5	8	1	19	1	0	1
17	0	0	1	0	1	1	0	1
18	1	2	6	0	9	1	1	0
19	2	3	6	1	12	1	1	0
20	1	1	2	0	4	1	0	1
21	2	2	4	0	8	1	1	0
22	0	0	0	0	0	1	1	0
23	0	0	1	0	1	1	0	0
24	12	15	17	2	46	0	1	0
25	3	5	6	1	15	1	1	0
26	0	0	2	0	2	1	0	1
27	2	2	4	0	8	1	0	0
28	1	1	1	0	3	1	1	0
29	0	0	1	0	1	1	1	1
30	1	1	1	1	4	1	0	0
31	0	2	4	0	6	1	0	0
32	0	0	2	0	2	1	1	0
33	4	3	5	1	13	1	0	1
34	1	1	4	0	6	1	0	1
35	2	2	3	0	7	1	1	0
36	1	2	3	0	6	1	1	1
37	6	8	9	1	24	0	1	0
38	1	1	3	0	5	1	0	0

C1	C2	C3	C4	C5	C6	C7	C8	C9
39	1	1	3	0	5	1	0	1
40	7	7	11	1	26	0	1	0
41	1	1	3	0	5	1	1	1
42	3	2	6	1	12	1	0	0
43	1	3	5	0	9	1	1	0
44	1	1	4	0	6	1	1	1
45	1	1	3	0	5	1	1	1
46	0	0	1	0	1	1	1	0
47	3	3	4	0	10	1	0	0
48	2	3	4	0	9	1	0	1
49	0	0	1	0	1	1	1	1
50	0	0	0	0	0	1	0	1
51	1	2	4	0	7	1	1	1
52	2	2	6	0	10	1	0	0
53	0	0	1	0	1	1	1	1
54	4	3	5	0	12	1	0	0
55	0	0	2	0	2	1	0	1
56	0	1	3	0	4	1	1	0
57	3	3	3	0	9	1	1	0
58	0	0	0	0	0	1	1	0
59	0	0	2	0	2	1	1	0
60	1	1	3	0	5	1	0	1
61	5	5	7	1	18	1	0	0
62	1	1	3	0	5	1	0	0
63	1	1	3	0	5	1	0	0
64	5	9	11	1	26	0	0	0
65	1	1	2	0	4	1	0	0
66	0	0	2	0	2	1	1	0
67	1	1	3	0	5	1	0	1
68	1	1	3	0	5	1	0	1
69	2	3	5	0	10	1	1	0
70	1	1	2	0	4	1	1	1
71	1	1	3	0	5	1	0	0
72	2	2	4	0	8	1	1	0
73	6	7	8	2	23	0	1	0
74	2	2	4	0	8	1	1	0
75	0	0	1	0	1	1	1	1
76	2	2	6	0	10	1	1	0
77	1	1	1	0	3	1	0	1
78	1	1	4	0	6	1	0	1
79	0	0	1	0	1	1	1	0

C1	C2	C3	C4	C5	C6	C7	C8	C9
80	0	1	2	0	3	1	1	0
81	3	4	5	1	13	1	0	1
82	0	0	2	0	2	1	1	0
83	3	3	6	1	13	1	1	0
84	1	1	2	1	5	1	1	1
85	0	0	1	0	1	1	1	0
86	1	1	5	0	7	1	1	0
87	0	0	2	0	2	1	1	0
88	2	2	5	0	9	1	1	0
89	4	5	8	0	17	1	0	0
90	0	0	2	0	2	1	0	0
91	0	0	1	0	1	1	1	0
92	1	1	2	0	4	1	1	1
93	0	1	2	0	3	1	1	0
94	1	1	2	0	4	1	0	0
95	8	7	9	4	28	0	1	1
96	2	3	4	0	9	1	0	0
97	0	0	1	0	1	1	1	0
98	1	1	3	0	5	1	1	0
99	2	4	5	1	12	1	1	0
100	0	0	1	0	1	1	1	0
101	5	9	9	2	25	0	0	0
102	2	3	4	0	9	1	0	0
103	0	1	2	0	3	1	0	0
104	2	2	4	1	9	1	0	0
105	3	2	6	1	12	1	0	0
106	2	2	5	1	10	1	1	0
107	1	1	1	0	3	1	1	0
108	0	1	3	0	4	1	1	1
109	1	1	3	0	5	1	0	0
110	3	3	4	1	11	1	0	0
111	7	8	11	2	28	0	0	0
112	0	2	4	0	6	1	1	0
113	1	1	3	0	5	1	0	0
114	1	2	3	0	6	1	1	1
115	0	1	2	0	3	1	1	1
116	0	0	1	0	1	1	1	1
117	0	1	3	0	4	1	1	1
118	3	7	10	0	20	1	0	0
119	1	5	6	0	12	1	0	0
120	7	9	10	0	26	0	0	0

Table 61: Results - Statistical data analysis

The responses that are identified as unreliable responses from the total outliers, detected by SD, IQD, MAD and MD methods, when finding the outlier for attribute and mixed attribute are given in table 62.

C1	C2	C3	C4	C5	C6	C7	C8	C9
24	12	15	17	2	46	0	1	0
37	6	8	9	1	24	0	1	0
40	7	7	11	1	26	0	1	0
64	5	9	11	1	26	0	0	0
73	6	7	8	2	23	0	1	0
95	8	7	9	4	28	0	1	1
101	5	9	9	2	25	0	0	0
111	7	8	11	2	28	0	0	0
120	7	9	10	0	26	0	0	0

Table 62: Results - Unreliable users from the total outliers

The total outliers for each attribute by all participants, which are identified by SD, IQD and MAD are given in table 63.

Attribute	SD	IQR	MAD	Total	Attribute	SD	IQR	MAD	Total
A1	9	10	12	31	A16	6	6	6	18
A2	8	4	10	22	A17	4	10	11	25
A3	6	24	24	54	A18	20	20	20	60
A4	4	4	4	12	A19	18	18	18	54
A5	6	7	6	19	A20	14	14	14	42
A6	6	7	10	23	A21	8	8	8	24
A7	6	6	11	23	A22	0	0	35	35
A8	3	11	15	29	A23	0	0	45	45
A9	9	8	8	25	A24	5	6	7	18
A10	5	3	9	17	A25	6	8	11	25
A11	16	16	16	48	A26	5	3	5	13
A12	5	5	5	15	A27	8	11	14	33
A13	2	4	4	10	A28	6	5	10	21
A14	1	8	10	19	A29	4	8	9	21
A15	0	0	53	53	A30	6	12	43	61

Table 63: The total outliers for each attribute by SD, IQD and MAD Methods

Mostly, outliers are detected in the following attributes.

A11) Page revisited status

A15) The status of out of page

A22) Radio button changes during the survey

A23) Checkbox changes during the survey

A30) Number of responded questions

The total outliers for each attribute set by all participants, which are identified by MD are given in table 64.

Attribute Set	No of Outliers
A2,A3,A4	6
A5,A6	2
A19,A20,21,A22,A23	2
A9,A10	2
A24,A25,A26,A27,A28	5
A15,A16,A17	6
A12,A13,A14	5
A1,A29	3
A7,A8,A11,A18,A30	5

Table 64: The total outliers by MD method

Mostly, outliers are detected for the following attributes.

A2) Total left button clicks

A3) Total right button clicks

A4) Total middle button clicks

A15) The status of out of page

A16) The status of idle

A17) Average time spent per page

Figure 36 depicts identified outliers from the total number of outliers. Identified outliers are shown by filled circle.

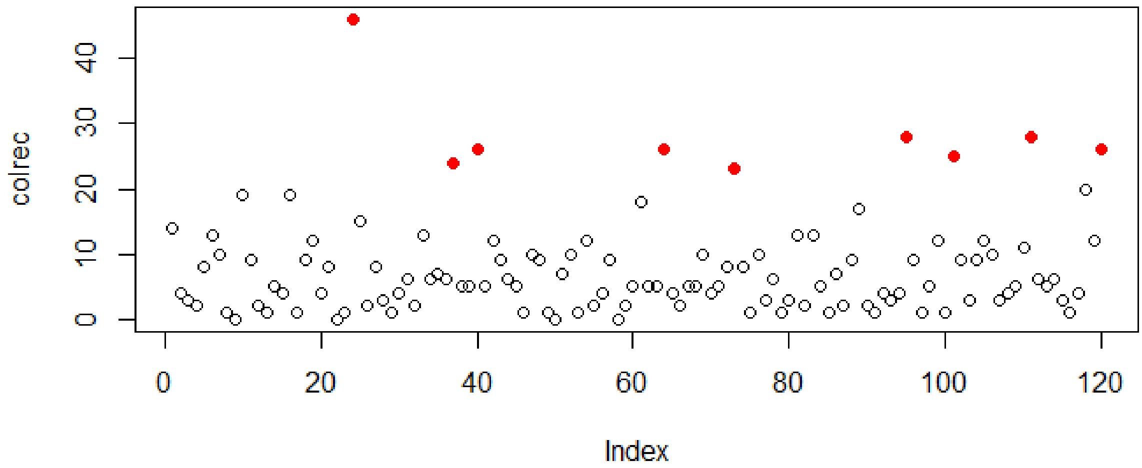


Figure 36 : Identified outliers from the total outliers

Figure 37 also depicts identified outliers from the total number of outliers using boxplot.

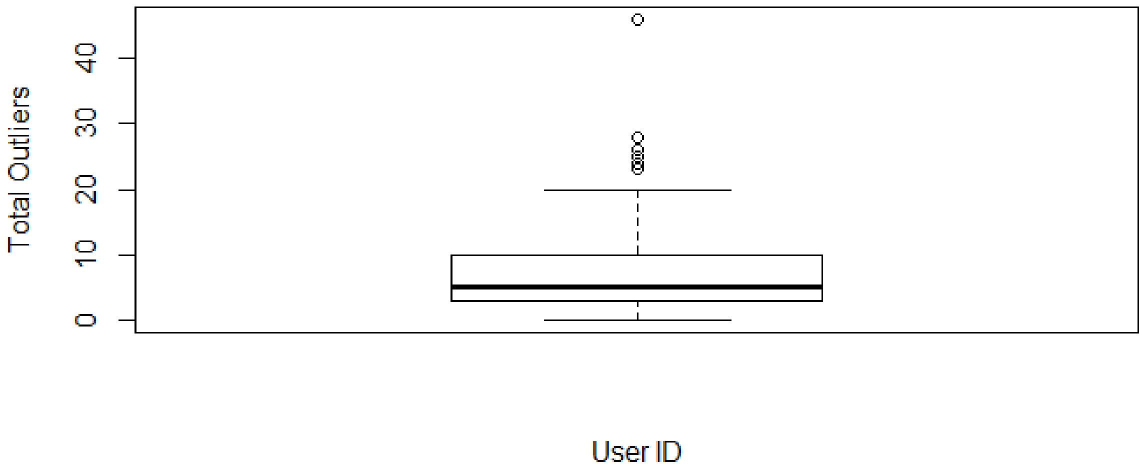


Figure 37 : With outliers

Figure 38 depicts without outliers (after removing the outliers from dataset) using boxplot.

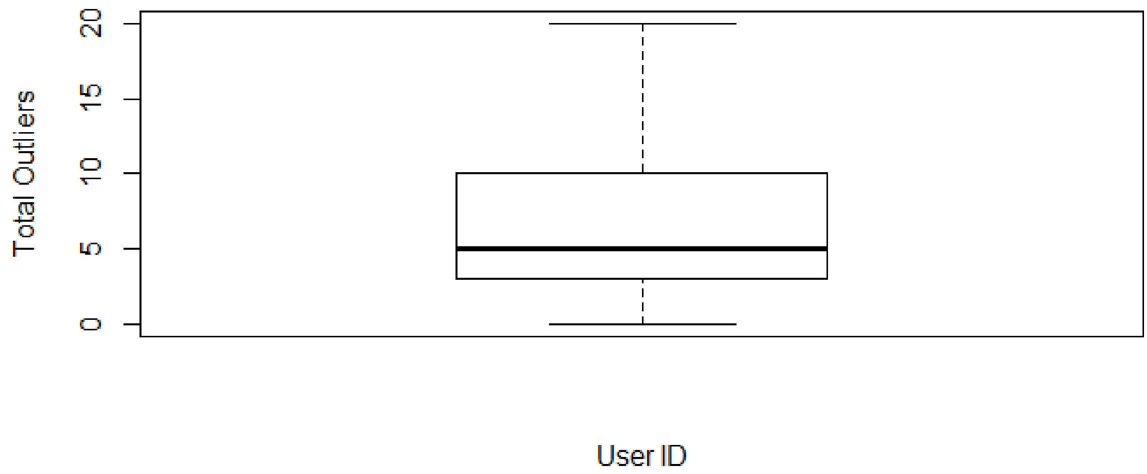


Figure 38 : Without outliers

This method however did not prove to be helpful since the threshold definition wasn’t feasible. As mentioned, multiple methods were used for the detection. The identified outlier values based on each method was sometimes different from each other.

For example, for a given range of average time, a response that falls out of the range is considered an outlier while in reality it might be reliable.

4.5 Chapter Summary

The above chapter summarizes the results which are derived from data analysis and evaluation details. Also, the unreliable responses that are identified from the analysis are mentioned in this chapter. The following chapter discusses the conclusion and future work for this research.

5 Conclusion and Future Work

5.1 Chapter Introduction

This chapter explains the outcomes of the processes that were underwent as part of this research. It also discusses the hypothesis related to the activities of this research.

The future work section elaborate on how this research can be used as a base for further developments in the improvisations of online surveys.

5.2 Conclusion

Based on the results of this research, following are the responses for the hypothesis considered.

H1. Deriving at the relevant attributes that have an effect on the online survey

This research is conducted based on behavioural data. Care was taken not to incorporate any content-based data where the answer and its relevance play the most important role. Mouse clicks, timing and scrolling were considered as key aspects in behavioural data that are used to create detailed attributes. 31 attributes were chosen for further analysis based on research study.

H2. Most influential attributes in identifying the reliability of response in online surveys

Total answer clicks, number of responded questions, checkbox changes during the survey and the status of idle are the most influential attributes in identifying the reliability of responses in online surveys. These attributes are considered to have a high influence while using the users' responses based on the self-reporting method and correct answers or either of them.

H3. Moderately influential attributes in identifying the reliability of response in the online surveys

Total middle button clicks, total clicks, average time spent between clicks, number of time scrollbar reached top, answer changes during the survey, number of time scrollbar reached bottom, page revisited status, radio button changes during the survey, average time spent on text area questions and average time spent on drop down questions are the moderately influential attributes in identifying the reliability of responses in online surveys. These attributes are considered to have a moderate influence while using the user responses based on the self-resporting method and correct answers or either of them.

H4. Least influential attributes in identifying the unreliable users in the online surveys

Average time spent per question, total left button clicks, total right button clicks, total non-answer clicks, text pasted status, pasted character count, average character per minute (cpm) for typing, the status of out of page, average time spent per page, text areas changes during the survey, drop down during the survey, text field changes during the survey, average time spent on radio button questions, average time spent on text field questions, average time spent on check box questions and average time taken between attempts on questions are the least influential attributes in identifying the reliability of the users' responses in online surveys. These attributes are considered to have a low influence while using the user response based on the self-reporting method and correct answers or either of them.

H5. Classification amongst reliable and unreliable response

As a part of the survey, the users were asked to self-report on whether their response is reliable or unreliable. 70% of data with categorical variable and the related predictor variables were used to trained the algorithms. This knowledge was used to classify the remaining data.

H6. Most suitable algorithm to predict the reliability of response in online surveys

Based on the accuracies, the algorithm that provided the highest accuracy is considered as the most suitable algorithm to predict the reliability of responses in online surveys. This study proves that the Logistic Regression would be the most suitable algorithm to identify the reliability of responses in online surveys.

H7. The attribute set that has the highest accuracy of impact amongst all the algorithms in identifying the reliability of response in online surveys

Two attribute sets that showed highest accuracy while using logistic regression were chosen to identify the reliability of response. Set 1 highlights self reporting that includes number of responded questions, total answer clicks, radio button changes during the survey, checkbox changes during the survey and answer changes during the survey. Set 2 highlights correct answers that includes, total middle button clicks, the status of idle, checkbox changes during the survey, average time spent on drop down questions, number of responded questions and average time spent on text area questions.

All the attributes considered are influential to an extent. All chosen algorithms (Naïve Bayes, Logistic Regression, Decision Tree, Support Vector Machine and Random Forest) are quite reliable as it shows above 50% accuracy to predict the response. However, the accuracy of this prediction increases considerably when using Logistic Regression and highly influential attributes.

As per the expected flow of the research, which required to use the behavioural data to identify the reliability of response in the online surveys, the results obtained are favourable, proving that a survey's reliability can be improved based on the enhancement of its behavioural attributes.

This research will help to identify the reliable and unreliable responses in fore coming online surveys especially the opinion based ones. It is hard to come to a conclusion in opinion based surveys as they do not have defined answers. However, when the reponse is predicted using user behavior while filling the survey, it become more reliable and efficient. Based on the results, when the unreliable responses are eliminated, the precision of the survey will increase. This will enable them to take the right decisions based on the reliable responses.

The overall study shows that the realiability of response in online survey can be predicted using behavivoral data. This will improve the accuracy of the survey result and become more dependable.

5.3 Future Work

Most of the features have been gathered as part of the behavioural data collection with the help of the created tool. However, all the observations were not used for the analysis or conclusion. Mouse click coordinates (x,y) that were captured during the survey has not been used so far. For instance, the data were collected page wise features along with overall characteristics. Currently the overall survey characteristics were used for the analysis. This can be further enhanced, by using a page-wise detection mechanism alongside.

The users were requested to mention if their response is reliable or unreliable. Though some users had declared themselves as reliable users, they were lacking in knowledge to answer the questions in the questionnaire. This proved to have a negative impact on the overall precision. The primary mode by which the overall efficiency of this research can be improved is by providing the attendees of the survey with a questionnaire to which the answers are pre-known or the answers are easily available in ways such as an open-book survey.

5.4 Chapter Summary

To summarize, this research holds good to serve as a base to further researches in improving the reliability of responses in online surveys. The main purpose of this research is to increase the reliability of response in opinion based surveys. Results of pinion based surveys are more dependable as they not only use the answers given by the user but also analyse the behaviour of the user. More importantly, it provides a mode to eliminate unreliable responses based on various factors. This will help the people who conduct the survey as the responses will be more reliable and accurate. Future considerations of using enhanced attributes and providing a more lenient questionnaire would benefit those who yearn for better results.

6 References

[1] J. A. DeSimone and P. D. Harms. “Dirty data: the effects of screening respondents who provide low-quality data in survey research,” *Journal of Business and Psychology,* pp. 1-19, Sep 2017. Accessed on: July 20, 2018. [Online]. Available doi: 10.1007/s10869-017-9514-9

[2] C. M. Rauti. “Assessing the effects of survey instructions and physical attractiveness on careless responding in online surveys.” M.A. thesis, University of Windsor, Canada, 2017. Accessed on: July 20, 2018. [Online].
Available: https://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=8292&context=etd

[3] M. E. Wertheimer. “Identifying the types of insufficient effort responders” M.A. thesis, Middle Tennessee State University, USA, 2017. Accessed on: July 20, 2018. [Online]. Available: http://jewlscholar.mtsu.edu/xmlui/bitstream/handle/mtsu/5308/Wertheimer_mtsu_0170N_10764.pdf?sequence=1&isAllowed=y

[4] T. D. Barnes. “Detecting Insufficient Effort Responding: An Item Response Theory Approach.” M.S. thesis, Wright State University, USA, 2016. Accessed on: July 20, 2018. [Online].
Available: https://etd.ohiolink.edu/!etd.send_file?accession=wright1467038541&disposition=inline

[5] J. A. DeSimone, P. D. Harms and A J. DeSimone. “Best practice recommendations for data screening,” *Journal of Organizational Behavior,* pp. 171-181, Feb 2015. Accessed on: July 20, 2018. [Online]. Available doi: https://doi.org/10.1002/job.1962

[6] J. L. Huang, M. Liu and N. A. Bowling. “Insufficient effort responding: examining an insidious confound in survey data,” *Journal of Applied Psychology,* vol.100, pp. 828–845, Apr 2015. Accessed on: July 20, 2018. [Online]. Available doi: https://doi.org/10.1037/a0038510

[7] A. S. McKay. “Improving data quality with four short sentences: how an honor code can make the difference during data collection.” M.A. thesis, California State University, San Bernardino, 2014. Accessed on: July 20, 2018. [Online].
Available: https://core.ac.uk/download/pdf/55330253.pdf

[8] D. J. Leiner. “Too fast, too straight, too weird: post hoc identification of meaningless data in internet surveys,” *SSRN Electronic Journal,* Nov 2013. Accessed on: July 20, 2018. [Online]. Available doi: 10.2139/ssrn.2361661

[9] A. W. Meade and S. B. Craig. “Identifying careless responses in survey data,” *Psychological Methods*, vol.17, pp. 437–455, Apr 2012. Accessed on: July 20, 2018. [Online]. Available doi: 10.1037/a0028085

[10] F. Austet et al., “Seriousness checks are useful to improve data validity in online research,” *Behavior Research Methods,* Oct 2013. Accessed on: July 20, 2018. [Online]. Available doi: 10.3758/s13428-012-0265-2

[11] A. Meade & B. Craig. “Identifying careless responses in survey data,” *Psychological Methods,* Apr 2012. Accessed on: July 20, 2018. [Online]. Available doi: 10.1037/a0028085

[12] S. Stieger and U. Reips. “What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study,” *Computers in Human Behavior*, Nov 2010.
Accessed on: July 20, 2018. [Online]. Available doi: 10.1016/j.chb.2010.05.013

[13] Knowledge Sourcing Intelligence LLP, *Rearch and Markets*, Aug 2017. Accessed on: July 20, 2018. [Online].
Available: https://www.researchandmarkets.com/research/64kglh/global_online

[14] J. Ye, Handbook of Research on Electronic Surveys and Measurements, 2007, pp. 83–89. Accessed on: July 20, 2018. [Online]. Available: http://www.irma-international.org/viewtitle/20219/

[15] M. Alice, *R-bloggers*, The University of Melbourne, Sep 13, 2015. Accessed on: July 20, 2018. [Online]. Available: https://www.r-bloggers.com/how-to-perform-a-logistic-regression-in-r/

[16] M. K. Frederik and S. B. Pond. “Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys” *Computers in Human Behavior*, July 2015. Accessed on: July 20, 2018. [Online]. Available doi: 10.1016/j.chb.2015.01.070

Appendix

Appendix A : Questionnaire used for the data collection

Page 1

Please mention your name.

Are you instructed to fill this questionnaire in a serious manner?

- Yes
- No

Which of the following age group do you belong to?

- Below 20 Years
- 21 - 30 Years
- 31 - 40 Years
- 41 - 50 Years
- Above 60 Years

What is the highest level of educational qualification you currently hold?

- Doctorate
- Master's Degree
- Bachelor's Degree
- Professional Qualifications
- Advanced Level

Which is a commonly used nickname for the beautiful island nation of Sri Lanka?

- Ruby of the Pacific
- Pearl of the Indian Ocean
- The Diamond of the Nile
- Jewel of the Nile

In which Sri Lankan city, also the capital of the last independent kingdom in 1592, is the Esala Perahera festival held?

Please choose...

Sri Lanka is known for exporting tea, coffee, rubber and cinnamon, but only one of them is native to the island. Which one?

- ○ Coffee
- ○ Cinnamon
- ○ Tea
- ○ Rubber

Identify two nearest countries to Sri Lanka.

- ☐ India
- ☐ Canada
- ☐ South Africa
- ☐ Bangladesh

In which Sri Lankan province is the largest international airport located?

What are the colors in the Sri Lankan flag?

- ☐ Dark Red
- ☐ Green
- ☐ Black
- ☐ Yellow
- ☐ Other:

Next

Page 2

Sri Lankans have traditionally gotten electricity from which form of energy?

- Wind energy
- Solar energy
- Nuclear energy
- Hydro-powered energy

The R. Premadasa Stadium (RPS) in Colombo is named after a former Sri Lankan _______.

On which Sri Lankan rock palace fortress, with an entrance between the paws of a lion, can you find a mirror wall with graffiti and fresco paintings of young maidens on sheer rock?

- Polonnaruwa
- Sigiriya
- Anuradhapura
- Mihintale

How many monsoon seasons are there in Sri Lanka?

- 1
- 2
- 3
- 4

What is the Sri Lankan international dialing code?

Elephants are one of the native animals of Sri Lanka. Though the population is declining, space has been set aside in National Parks and Nature Reserves for their habitat. Three of these are names of parks where elephants can be seen, which are they?

- [ ] Theravada
- [ ] Wilpattu
- [ ] Udawalawe
- [ ] Yala

Which is the institution that deals with Drug Addiction in Sri Lanka?

Please choose...

Cricket is the most popular sport in Sri Lanka. Whom did they beat in 1996 to win the World Cup?

- India
- Australia
- Pakistan
- South Africa

Please indicate your agreement with the following statements.

	True	False
Bambarakanda is the tallest waterfall in srilanka.		
The lotus is the national flower of Sri Lanka.		
"Sri Lanka Matha" is the national anthem of Sri Lanka.		
Adam's Peak is the tallest mountain in Sri Lanka.		
Mahaweli is the largest river in Sri Lanka.		

Please indicate the level of your agreement with the following statements.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Nuwara Eliya is one of the coldest places in Sri Lanka.					
Sri Lanka is one the tea exporters in the world.					
The coastal region of Sri Lanka was not effected by Tsunami.					
The tourism industry is developing in Sri Lanka.					
Sri Lankan transportation system has been developed over the last 10 years.					

Previous

Next

Page 3

In December 2004 Sri Lanka suffered severe damage and loss of life from which force of nature?

- Cyclone
- Fire
- Monsoon
- Tsunami

Which is the popular reserved wildlife sanctuary in Sri Lanka?

Please choose...

In which Sri Lankan city is the natural harbour located?

Which country provided assistance to construct the Hambantota harbour?

- Iran
- Japan
- Russia
- China

Which animal orphanage is located in Pinnawela?

Please choose...

Which animal orphanage is located in Pinnawela?

Please choose...

Another place of interest in Sri Lanka is The Sri Dalada Maligawa or Temple of the Sacred Tooth. This beautiful complex houses a tooth from which highly venerated person?

Who wrote the National Anthem of Sri Lanka?

- ○ Mahagama Sekara
- ○ Karunarathna Abeysekera
- ○ Donlton Alwis
- ○ Ananda Samarakone

What are the current television stations based in Sri Lanka?

- ☐ ITN Channel
- ☐ Rupavahini
- ☐ STAR Plus
- ☐ Hiru TV
- ☐ Zee TV

Please indicate your agreement with the following statements.

	True	False
Every month, the Full Moon Day is known as "Poya Day".	○	○
All Poya Days are national holidays in Sri Lanka.	○	○
Alcohol is served on Poya Days.	○	○
Hikkaduwa is located in the hill country of Sri Lanka.	○	○
The narrow stretch of water between Sri Lanka and India is called the Palk Strait.	○	○

Please indicate the level of your agreement with the following statements.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Our country gains a lot of revenue through tourism.	○	○	○	○	○
The number of vehicles have increased over the last 2 years.	○	○	○	○	○
The cost of living has increased in Sri Lanka.	○	○	○	○	○
There is an extreme lack of electriciy and water supply in Sri Lanka.	○	○	○	○	○
The acadamic fee of private schools in Sri Lanka are low.	○	○	○	○	○

Previous

Submit

Appendix B : Code for the tool creation (JavaScript)

Create the storage objects

```
if(!jsonString) {
  var mainStorageObj = [{
  userDetails: {
  surveyCode : survey_id,
  surveyTitle : survey_title,
  browserName : browser_name,
  versionNo : browser_version,
  osName : os_name,
  osVersion : os_version,
  mobileType : mob,
  tabType : tab,
  ScreenWidth : widthSC,
  ScreenHeight : heightSC,
  ip : userip
  },
  pageDetails:[]
  }];
} else {
  var mainStorageObj = JSON.parse(jsonString);
}
 var localObj = {
  pageTitle: page_title,
  pageStartTime : null,
  pageEndTime : null,
  pageSpentTime : null,
  clickedElements : [],
  questionIds : [],
  answerSelections : [],
  pageIdleTimes : [],
  pageAwayTimes : [],
  scrollElements : [],
};
```

Calculate the pasted text details

```
var text_all;
var text_all_pasted;
var pasted;
var paste_count=0;
var text_copy_paste;
var character_count=0;
var pasted_elements=[];;
var textlength;
function text_diff(first, second) {
        var start = 0;
        while (start < first.length && first[start] == second[start]) {
            ++start;
        }
        var end = 0;
        while (first.length - end > start && first[first.length - end - 1]
        == second[second - end - 1]) {
            ++end;
        }
        end = second.length - end;
        return second.substr(start, end - start);
}
$('textarea, input[type="text"').bind('paste', function(event) {
    var self = $(this);
    var orig = self.val();
    press_time=moment().format('YYYY-MM-DD HH:mm:ss');
    setTimeout(function() {
        pasted = text_diff(orig, $(self).val());
        pasted_elements.push(pasted);
        textlength = pasted.length;
        character_count=character_count+textlength;
        paste_count=paste_count+1;
    });
});
```

Calculate the idle time (if user is idle more than the defined number of seconds)

```
var count_idle_time=0;
var IDLE_TIMEOUT = 60;
var _idleSecondsCounter = 0;
var cit=0;
var wsi;
document.onclick = function() {
    _idleSecondsCounter = 0;
};
document.onmousemove = function() {
    _idleSecondsCounter = 0;
};
document.onkeypress = function() {
    _idleSecondsCounter = 0;
};
function CheckIdleTime() {
    _idleSecondsCounter++;
    if(count_idle_time >= 60 && _idleSecondsCounter==1){
    count_idle_time=count_idle_time+1;
    idle_end_time=moment().format('YYYY-MM-DD HH:mm:ss');
    var cit = count_idle_time;
    var iet = moment(idle_end_time);
    var timespec=iet.subtract(cit, 'seconds');
    var idle_start_time=moment(timespec).format('YYYY-MM-DD HH:mm:ss');
    var pageIdleObj = {
    idleTime : count_idle_time,
    idleTimeStart : idle_start_time,
    idleTimeEnd : idle_end_time,
    }
    localObj.pageIdleTimes.push(pageIdleObj);
    count_idle_time=0;
    }
    if (_idleSecondsCounter >= IDLE_TIMEOUT) {
        count_idle_time=_idleSecondsCounter;
    }
}
```

Calculate time where user was out of page

```
var d = 0;
var e;
var start_no=0;
var starttime;
var endtime;
var away_time_diff=0;
var get_current_date_time;
var get_time_in_seconds;
var away_time_difference=0;
$(window).focus(function() {
    endtime=moment().format('YYYY-MM-DD HH:mm:ss');
    get_focus_date_time = new Date();
    get_focus_time_in_seconds = Date.parse(get_focus_date_time)/1000;
    stopCount();
    cit=window.setInterval(CheckIdleTime, 1000);
});
$(window).blur(function() {
    starttime=moment().format('YYYY-MM-DD HH:mm:ss');
    get_blur_date_time = new Date();
    get_blur_time_in_seconds = Date.parse(get_blur_date_time)/1000;
    startCount();
    clearInterval(cit);
});

function startCount()
{
    d = d+1;
    e = setTimeout(function(){startCount()},1000);
}
function stopCount()
{
    clearTimeout(e);
    if(d>0){
        var a = moment(starttime);
        var b = moment(endtime);
        away_time_diff=b.diff(a, 'seconds');
        var outofpageObj = {
        awayTimeStart : starttime,
        awayTimeEnd : endtime,
        awayTime : away_time_diff,
        }
       localObj.pageAwayTimes.push(outofpageObj);
    }
    d=0;
}
```

Calculate the scrollbar details (reached top and bottom)

```
function getDocHeight() {
    var D = document;
    return Math.max(
        D.body.scrollHeight, D.documentElement.scrollHeight,
        D.body.offsetHeight, D.documentElement.offsetHeight,
        D.body.clientHeight, D.documentElement.clientHeight
    )
}
function amountscrolled(){
    var reached;
    var winheight= window.innerHeight || (document.documentElement
    || document.body).clientHeight
    var docheight = getDocHeight()
    var scrollTop = window.pageYOffset || (document.documentElement
    || document.body.parentNode || document.body).scrollTop
    var trackLength = docheight - winheight
    var pctScrolled = Math.floor(scrollTop/trackLength * 100)
    if(pctScrolled==0){
    reached="top";
    } else if (pctScrolled==100){
    reached="bottom";
    }
    if(pctScrolled==0 || pctScrolled==100){
        var scrollObj = {
        reachedType : reached,
        timeScroll : moment().format('YYYY-MM-DD HH:mm:ss'),
        }
       localObj.scrollElements.push(scrollObj);
    }
}
window.addEventListener("scroll", function(){
    amountscrolled()
}, false)
```

Calculate the typing speed on textarea and textfield

```
function checkSpeed(){
    iTime = new Date().getTime();
    if (iLastTime != 0) {
        iKeys++;
        iTotal += iTime - iLastTime;
        iWords = $('textarea,input[type="text"]').val().split(/\s/).length;
        cpmval=Math.round(iKeys / iTotal * 60000, 2);
        wpm=Math.round(iWords / iTotal * 60000, 2);
    }
    iLastTime = iTime;
}
$('input[type="text"]').keyup(function(e) {
    press_time=moment().format('YYYY-MM-DD HH:mm:ss');
    checkSpeed();
});

$('textarea').keyup(function(e) {
    press_time=moment().format('YYYY-MM-DD HH:mm:ss');
    checkSpeed();
});
```

Calculate the clicked element details

```
$('body').bind('mousedown', function (el) {
    var button_type_r;
    if (el.which == 1) {
    button_type_r="left";
    } else if (el.which == 2){
    button_type_r="middle";
    } else if (el.which == 3){
    button_type_r="right";
    } else {
    button_type_r="unknown";
    }
    xvalue=el.clientX;
    yvalue=el.clientY;
    time_int=c-current_timer;
    current_timer=c;

    var date_now = new Date();
    date_now_milli=Date.parse(date_now)/1000;
    if(previous_clicked_time==0){
        previous_clicked_time=load_milli;
    }
    diff=date_now_milli-previous_clicked_time;
    previous_clicked_time=date_now_milli;
    var clickObj = {
        element : el.target.localName,
        button_type : button_type_r,
        cordinates : {
          x : xvalue,
          y : yvalue,
        },
        time_between_clicks : diff,
        clicked_time : moment().format('YYYY-MM-DD HH:mm:ss'),
      }

  localObj.clickedElements.push(clickObj);
});
```

Add the details to browser’s local storage

```
$('#movesubmitbtn').click(function(e) {
    localObj.pageEndTime = moment().format('YYYY-MM-DD HH:mm:ss');
    var date_end_now = new Date();
    end_date_now_milli=Date.parse(date_end_now)/1000;
    page_diff=end_date_now_milli-load_milli;
    localObj.pageSpentTime = page_diff;
    mainStorageObj[0].pageDetails.push(localObj);
    localStorage.setItem("analytics", JSON.stringify(mainStorageObj));
});
```

Pass the details to server through AJAX call from local storage and delete from local storage

```
$.ajax({
  url:"http://knreviews.com/limesurvey/script.php",
  type:"POST",
  cache: false,
  async: false,
  data:{myfile:'test',json:jsonString},
  dataType:"json",
  success:function(data){
      if(data.status==1){
          delete window.localStorage["analytics"];
          localStorage.removeItem("analytics");
      }
  },
  error:function(data) {
  }
  });
```

Appendix C : Tool Validation

Tool Features	**Method or Software Used for the Validation**	**Validation Status**
Survey Code	Lime Survey Admin Panel	Pass
Survey Title	Lime Survey Admin Panel	Pass
Browser Name and Version	Third Party Script (ssl.mousestats.com and hotjar.com)	Pass
OS Name and OS Version	Third Party Script (ssl.mousestats.com and hotjar.com)	Pass
Mobile and Tab (Yes/No)	Third Party Script (hotjar.com)	Pass
Screen Width and Height	Third Party Script (ssl.mousestats.com) and Auto Clicker Software	Pass
IP	Third Party Script (ssl.mousestats.com)	Pass
Page Code	Browser Inspect Element	Pass
Page Title	Lime Survey Admin Panel	Pass
Page Started/Loaded Time	Screen Recording - Debut Video Capture Software	Pass
Page Ended Time	Screen Recording - Debut Video Capture Software	Pass
Spent Time on Page	Screen Recording - Debut Video Capture Software	Pass
Clicked Element on Page	Screen Recording - Debut Video Capture Software	Pass
Clicked Button Type	Screen Recording - Debut Video Capture Software	Pass
Clicked Page Coordination	Auto Clicker software and Mouse Recorder Premium	Pass
Clicked Screen Coordination	Auto Clicker software and Mouse Recorder Premium	Pass
Time Taken Between Two Clicks	Auto Clicker software and Mouse Click Info Software	Pass

Clicked Time	Mouse Click Info Software	Pass
Question IDs on Page	Lime Survey Admin Panel and Browser Inspect Element	Pass
Answered Question ID	Lime Survey Admin Panel	Pass
Question Type	Screen Recording - Debut Video Capture Software	Pass
Answering Element ID	Screen Recording and CSSviewer Addon Chrome	Pass
Answer Value	Screen Recording - Debut Video Capture Software	Pass
Answering Time Started	Screen Recording - Debut Video Capture Software	Pass
Answering Time Ended	Screen Recording - Debut Video Capture Software	Pass
Pasted values/Text On the Page	Screen Recording - Debut Video Capture Software	Pass
No of Characters Pasted	Screen Recording - Debut Video Capture Software	Pass
Pasted Characters Count	Screen Recording - Debut Video Capture Software	Pass
Character per Minute for Typing	-	-
Page Idle Start Time	Record User Idle Time Software	Pass
Page Idle End time	Record User Idle Time Software	Pass
Page Idle Time Duration	Record User Idle Time Software and Mouse Recorder Premium	Pass
Page Away Start Time	Screen Recording - Debut Video Capture Software	Pass
Page Away End Time	Screen Recording - Debut Video Capture Software	Pass
Scrollbar Reached Type (Top/Bottom)	Screen Recording - Debut Video Capture Software	Pass
Scrollbar Reached Time (Top/Bottom)	Screen Recording - Debut Video Capture Software	Pass

Appendix D : Tool Installation

Download the files from http://knreviews.com/downloads/tool

Download the installation guide from http://knreviews.com/installation-guide/tool/installation-guide-for-tool.pdf.

Install and configure.

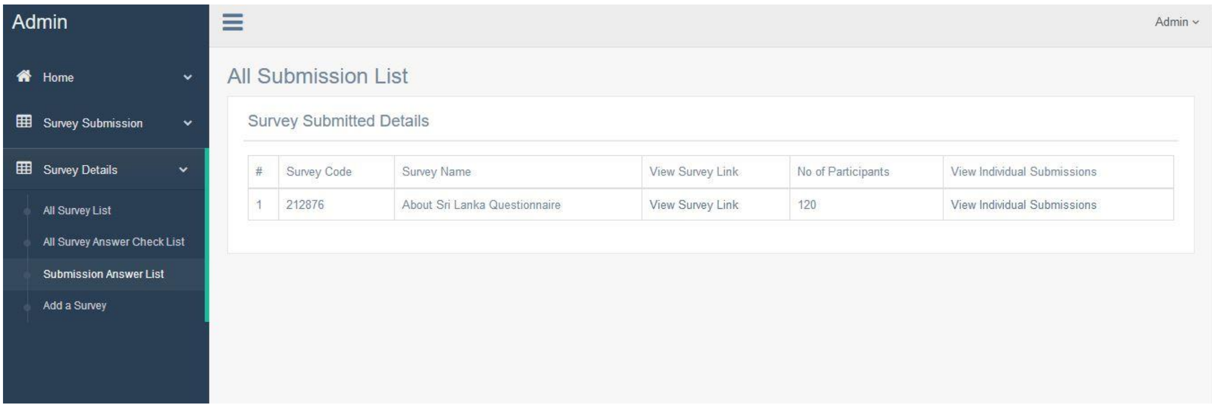
Appendix E : System Installation for Dataset Download

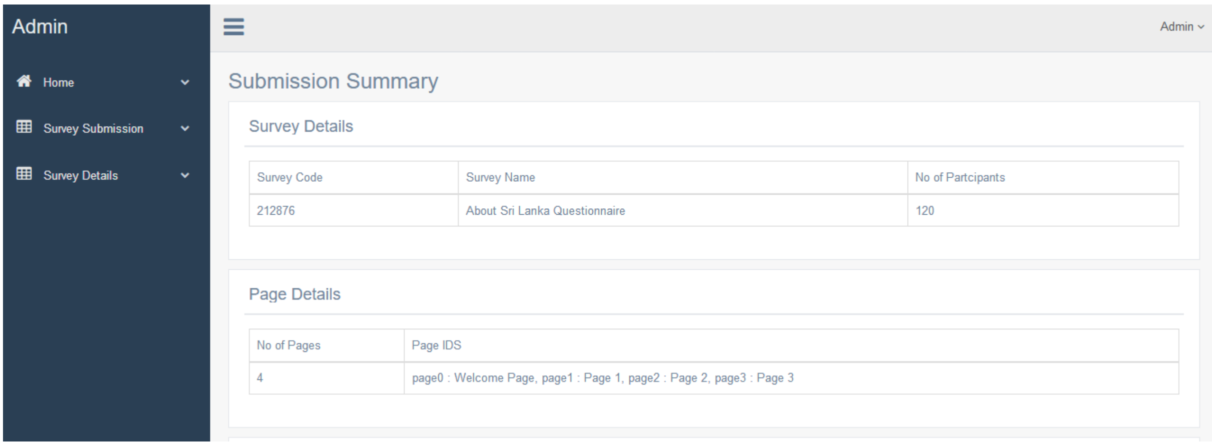
Download the files from http://knreviews.com/downloads/system

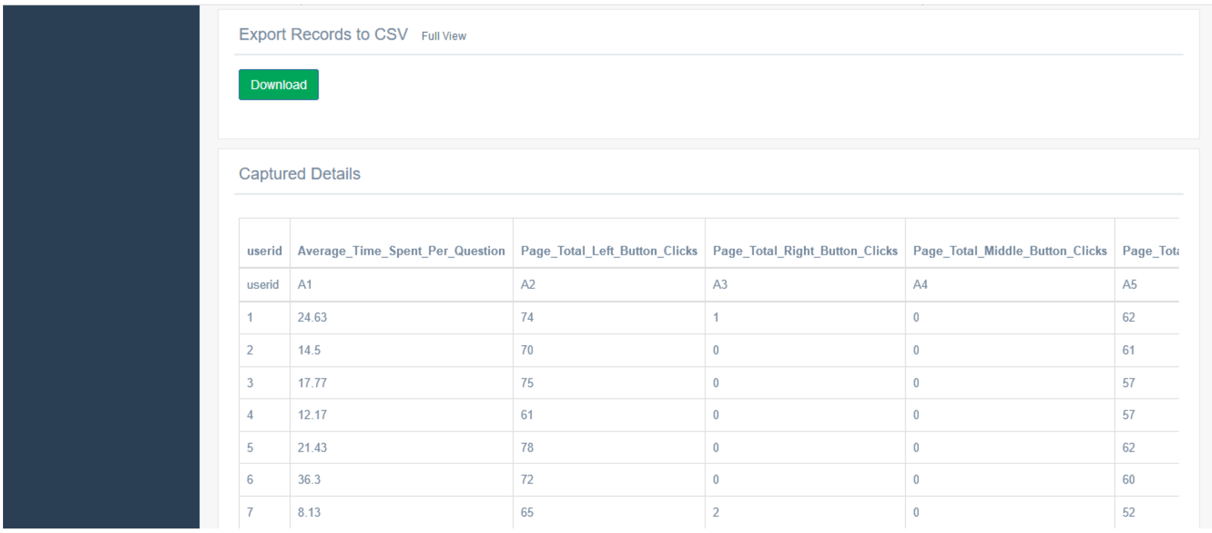
Download the installation guide from http://knreviews.com/installation-guide/system/installation-guide-for-system.pdf.

Install and configure.

Appendix F : Dataset Preparation System







Appendix G : Classification Using R

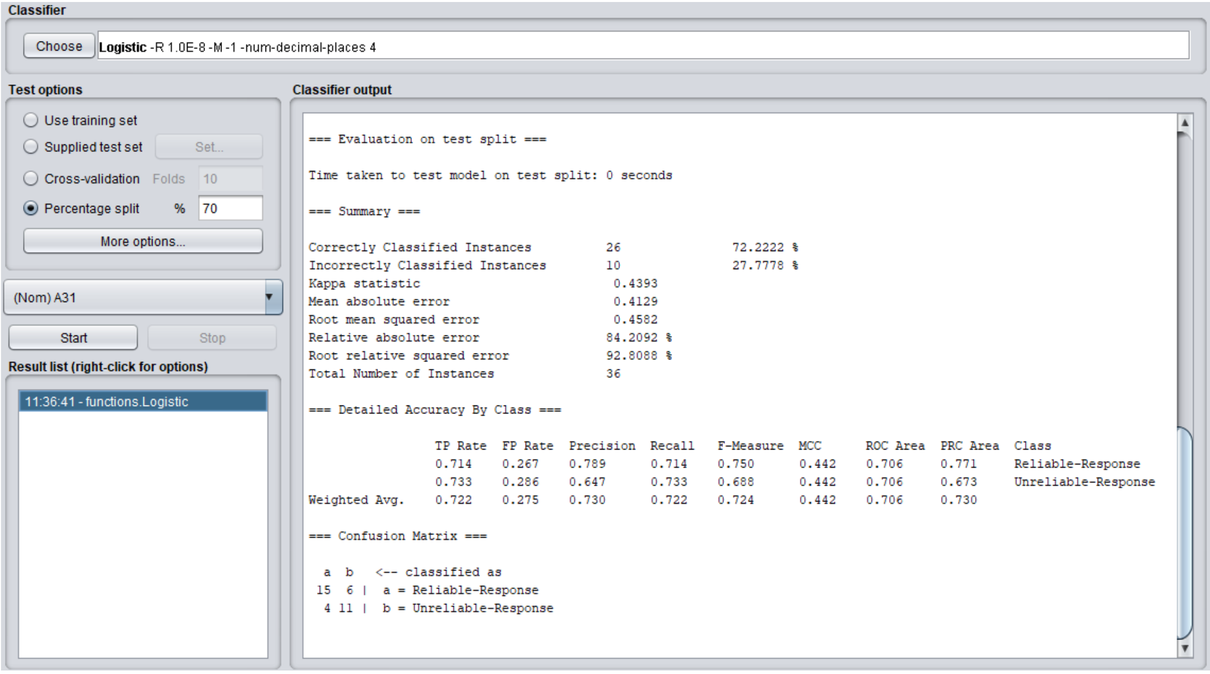
Forward Feature Selection with Logistic Regression

```
DataSet = read.table("G:/system/data-set-final.csv",header=T,sep=",")
set.seed(120)
dt = sort(sample(nrow(DataSet), nrow(DataSet)*.7))
train=DataSet[dt,]
test=DataSet[-dt,]
A31<- factor(DataSet$A31)
dim(train)
dim(test)
fullModel=glm(A31~ A1+A2+A3+A4+A5+A6+A7+A8+A9+A10+A11+A12+A13+A14+A15
              +A16+A17+A18+A19+A20+A21+A22+A23+A24+A25+A26+A27
              +A28+A29+A30, data = train, family = binomial)
nullModel=glm(A31~1,data = train, family = binomial)
nullModel
library(MASS)
forward = stepAIC(nullModel, direction="forward", scope = "fullModel")
formula(forward)
summary(forward)
#Predict the test data
p2 <- predict(forward, newdata=test, type = 'response')
pred2 <- ifelse(p2>0.5, 1, 0)
tab2 <- table(Actual = test$A31, Prediction = pred2)
```

Appendix G : Logistic Regression Using Weka

Identifing reliable and unreliable responses based on self-reporting

Attributes : A4,A16,A23,A25,A30



Identifing reliable and unreliable responses based on correct-answers

Attributes : A30,A5,A22,A23,A18