

# Data Collection and Analysis Techniques in Computer Science Research: A Review

Muhammad Ateeb Ather

Department of Computer Sciences, Bahria University, Lahore,  
Pakistan

03-134211-022@student.bahria.edu.pk

**Abstract.** Many researchers and scientists face difficulties in accomplishing research work successfully due to the lacking of good knowledge regarding method of scientific data collection. Several methods are utilized for collection of data in research work. Effectiveness of every research depend on the validity of collected data. Appropriate research questionnaires provide motivation and direction in structuring data gathering and data analysis approaches. This paper focuses on the data collection methods and data analysis techniques in terms of qualitative and quantitative particularly for computer scientists. Furthermore, this work is to examine numerous analysis methods, techniques and tools that could be applied to diverse type of data science applications.

**Keywords:** Data collection, data analysis, qualitative research, quantitative research.

## 1 Introduction

This research is basically a dedicated activity which is more than gathering information or report writing. It contains information gathering in a traditional systematic way, which can be examined methodically to bring to answers of research questionnaires and calculate results. Data collection is considered to be the core of any research design regardless of the area of study [69]. Every research starts with definitive questions, which must be responded. Central rule regarding data collection and analysis is not to gather and examine all types of data. Inappropriate data collection effects on the result of a study and bring towards inaccurate or wrong results. Basically data collection is considered to be the process of collecting and evaluating information in a targeted systematic manner that provides the facility to someone to respond according to research queries, hypothesis testing and results calculations.

Data gathering research module is common to all study areas including business, social and physical sciences and humanities, etc. Although methods differ by research field, main focus on guaranteeing appropriate and accurate collection must be same. The primary goal behind data collection is to maintain research integrity. Data

collection is said to be challenging task needs throughout the planning, development and analyzing process to complete job successfully. Data collection begins with defining what type of data needed (qualitative and quantitative). Which source of data required (primary and secondary).

By using different techniques data can be gathered. Decision regarding which instrument will be utilize for data collection is conducted via research question. After the gathering of appropriate and accurate data through the usage of suitable technique, next phase is the discovery of useful and appropriate information hidden inside data for further interpretation and manipulation. Analysis word defines operations, using with the intent that summarization of gathered data and forming to such an extent yielding responses to queries. In other words, it indicates, study of data to define inherent statistics. Process of applying statistical and logical methods for the extraction of accurate and related information from data is said to be data analysis [39].

During data analysis process, three things play an important role. First is organization of data, second is the classification and summarization of data, both used for the reduction of data and third is the identification and linkage of data patterns. Researchers do data analysis in bottom up and top-down mode. Data analysis includes forming of data in an appropriate manner.

Problem behind analysis of data differs from area of study to study. Different analysis techniques presented by researchers according to data type. First section describes data and information, data sources and data types. Second section describes different ways of data collection. Third section presents definition of data analysis, methods of data analysis, tools of data analysis and research techniques. Forth focuses on the data analysis techniques in terms of qualitative and quantitative perspectives. Fifth section compares quantitative and qualitative methods.

## **2 Data and Information**

Data and information both are same notions, but both are not similar thing. Primary difference among both is that information is entire and data is the portion. Information is basically the outcome of processed data. Information word drives from Latin word meaning 'conception or formation'. Information concept has diverse meanings in diverse type of contexts. Hence concept of information becomes linked to ideas of communication, education, understanding, perception, control and knowledge etc. Information related with data. Information solve out uncertainty.

Information is transferable in time, with the help of space, telecommunication, data storage and communication. Though data and information terms used interchangeably, both terms different meanings. Researchers defined concept of information in different perspectives: information as data in environ, information as resource, information as communication process part and information as knowledge representation [48].

Data word is the plural form of "Datum". In numerous means, data defined according to diverse types of people according to their specialized area. According to this study, data is stated as gathering of facts, like measurements, observation or values and statements about things [12]. Data word represents information, which gathered in some organized way and help out someone to understand the information properly. Representation of instructions, concepts and facts in formal way appropriate for

**Table 1.** Comparison among primary and secondary data.

<b>Features</b>	<b>Primary Data</b>	<b>Secondary Data</b>
Data	Real time data	Past data
Process of collection	Very involved	Easy and fast
Cost of collection	High	Low
Time consumed	More time	Less time
Type of data	Qualitative	Quantitative
Reliability	More reliable	Less reliable
Accuracy	Accurate	Lack of accuracy
Sources	Surveys, observations, experiments	Internal records, Govt, published data etc

processing, interpretation and communication via automatic means and humans is said to be data. Data not gathered irregularly, it is necessary with regards to answer queries [3].

As discussed earlier, data collection is considered necessary in every research area. After the defining of research problem, data collection task starts. Given below section elaborate different types of data and gives clear depiction of sources through which data can be collected easily.

## 2.1 Data Sources Types

Data sources can be subdivided into two categories: secondary and primary sources of data.

### 2.1.1 Primary Data

Gathering of data from information source and first-hand experience is said to be primary data. In simple words, data which is gathered for first time by scholars/authors himself. It is more realistic, accurate and objective. Its viability is higher than secondary data because it cannot be transformed via humans. Its collection time is longer and always specific to scholar requirements [9].

### 2.1.2 Primary Data Sources

In primary data collection sources, researcher need to be clearly 1 define population under study also analysis units. Primary data sources are inadequate and due to deficiency of collaboration it's hard to acquire data from primary architect. Primary data mostly gathered via negotiations or face-to-face discussions and through exchanging of emails, radio based communication, phone oriented discussions and direct surveillance. Other primary data architects are observations, survey, interviews, experiments and questionnaires [40].

### 2.1.3 Secondary Data

A way in which data gathered from architect which has been already advertised is known as secondary data. Literature review found in each research is considered to be

the foundation of secondary data. For social sciences, secondary data architects involve organizational reports and gathering of data from qualitative based research and quantitative approaches. It is said to be less effective. In some cases, when difficult to capture primary data, attaining information from secondary architect is feasible. Usage of secondary data is less costly and quicker in terms of distinguishing to primary data [9].

#### **2.1.4 Secondary Data Sources**

There are few means of gathering secondary data involve newspaper, books, research and internet articles, records, databases and profiles etc [40].

### **2.2 Data Types**

There are two types of data: qualitative and quantitative.

#### **2.2.1 Qualitative Data**

This type of data is not in the form of numbers, basically descriptive. Most of the times, such data expresses sentiments, moods and feelings etc. It is said to be the depiction of conditions, actions, connections, direct citations, case studies and records. Qualitative studies address questions 'what' 'why' and 'how'. This type of data gathering procedures play necessary part in influence estimation via supplying information beneficial to comprehend methods beyond observed outcomes and evaluate variations in people sensation of their welfare. Furthermore, it is costly and takes time. Scholars need to capture data systematically and accurately. Data gathering approaches need to perceive ethical values of research. Qualitative approaches mostly utilized in evaluation and can be organized in three groups: methods of observation, in-depth interview and document review [83].

#### **2.2.2 Quantitative Data**

Type of data which can be expressed in terms of numbers. By nature, it is in the form of structured and unstructured. Arranged or organized form of data is said to be structured data, while un-organized form of data called unstructured data. Via closed queries structured data can be generated, unstructured data can be generated via open queries [12,13]. It utilizes various measurement scales like ordinal, nominal, ratio and interval scale. Quantitative studies address questions 'how many' and 'how much'. Quantitative data collection approaches generate outcomes which can be easily summarized, compared and generalized.

#### **2.2.3 Mixed Methods**

This method gathers qualitative and quantitative data within similar study. Several researchers assume that mixed method is a new technique but researchers have been gathered qualitative and quantitative data for several eras. According to [37] when studying difficult and complex issues, one method just show a slight vision of entire depiction. Mixed methods deliver info on various stages of understanding.

Qualitative methods give deep understand ability of variables which takes to quantitative mathematical findings, when both methods grouped. Mixed methods uses both methods qualitative and quantitative either sequentially or simultaneously [79]. It

increases research validity and reliability. Researchers face problems when they use mixed methods technique is that, it is time consuming, expensive and takes effort.

### 2.3 Types of Quantitative Data

Quantitative data classified into two types: discrete and continuous data.

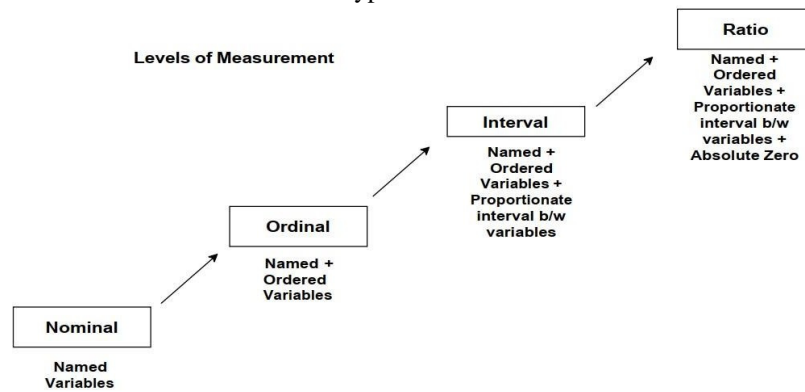


Fig. 1. Quantitative Levels of Measurement: Ordinal, Interval, Nominal and Ratio.

#### 2.3.1 Discrete Data

Type of data which comprises of calculating numbers only. It only uses separate and distinct values. It holds only finite numbers, sub-section of those numbers cannot be happened. It holds only those type of values which are in countable form, means impossible to breaking values. Graphical representation of discrete data can be done by bargraph. Its nature is countable. Examples include: no of employs in a company, no of animals in a zoo and no of cars inside parking area etc.

#### 2.3.2 Continuous Data

Type of data which is measurable. It can be reduced in sub-portions. This type of data cascades within continuous order. Graphical representation of continuous data can be done by histogram. Its nature is measurable. For various types of hypothesis tests, this data is use able.

Examples are weight, height, temperature and length [58].

### 2.4 Measurement Levels

Measurement can be grouped into four stages: ordinal, interval, ratio and nominal.

#### 2.4.1 Nominal Level

For identification goals, this scale is used. Another name of this scale is "categorical scale" because for easy identification, it allocates quantities to attributes. Quantities behave like labels. It can be examined via pie and bar charts. Example: sex (male/female) and status (single/married) are two groups with categories, but these groups having no ranking or natural ordering [18].

#### **2.4.2 Interval Level**

It contains principle of static space among 4 and 5 and 5 and 6, must be equal space. Calendar time is an example of interval scale. It means time among 3:00 am and 4:00 am is equal and same to the 4:00 am and 5:00 am time [22]. Most of the attitude scales are viewed as interval like multi-graphical scales, intelligence scores and semantic differential scales. It is applicable in numerous disciplines such as medicine, education and engineering etc.

#### **2.4.3 Ordinal Level**

It contains organizing or ranking of characteristics reliant on the variable being scaled. Articles within this scale ordered according to the amount of existence of variable in query. Characteristics or attributes on this scale generally organized in descending or ascending way. It calculates the ratio of existence of variable. This scale applied in advertising, surveys related to client satisfaction and market based research. For the representation of a degree, it utilizes modifier such as highly, less, very and more etc. Example: ordinal levels mostly found in questionnaires for quality ratings (good, poor, excellent and very good etc.) and agreement (agree, disagree and strongly disagree etc.) etc [18,87].

#### **2.4.4 Ratio Level**

It is said to be the peak level of measurement and it comprises all attributes of zero origin, fixed space, order and categorization. This scale of measurement permits scholar to match both the variances and comparative magnitude of quantities. Money is an example of ratio level [22,87].

### **3 Data Collection Methods**

By using several methods and from various sources, data can be gathered and attained. Following are the methods used for data collection.

#### **3.1 Observation**

It is said to be the most useable method particularly in behavioral sciences studies. For researcher's, it has been become a scientific instrument. Observation is referred to as the "logical depiction of conducts, events and objects in social setting selected for study" [50]. Scholar perceives participants conduct and record outcomes of these perceptions. This method helps out researcher to capture non-verbal signs like emotions, way of communication and note down time they devote in various tasks [60]. Furthermore, it is expensive because its time consuming and observer needs to be stay during the whole process.

#### **3.2 Focus Groups**

This method belongs to qualitative method. Marketing professional Ernest Ditcher and psychologist (1907-1991) invented the term. In this method, group of people are invited

to take part in a conversation on a specific topic. By using this method, data type which can be attained contain views, beliefs, and agreement and disagreement expressions with other members and procedures in which single or group personalities are constructed [59]. It provides an open atmosphere, with the intent that members feel free to reveal their opinions and views. To find out results, gathered data examined systematically and thoroughly. Problem with this method is that, when group mediator takes part in discussion may disturb the flow of discussion among group participants [31]. Initially, for market based research, this method used, but now it is applied in political analysis, health based campaigns and public sector marketing etc.

### **3.3 Interview**

Verbal method of asking questions is said to be interview. It's a sequence of queries a scholar asks to someone. Primarily, utilized to acquire an understanding of aims and causes for people conduct, likings, attitudes, views and beliefs on related topics. Interviews can be conducted individually or in a group [26].

Interviews can be in the form of semi-structured, structured or unstructured. In order to attain data, structured interview depends on the set of planned and standardized questionnaires while unstructured interview does not depend on the set of standardized questions. In semi-structured interview, interviewer utilizes an interview leader and it is basically a set of queries which must be covered during conversation or communication.

Furthermore, it provides free environment for asking queries in any sequence [81]. Interview method is further sub-divided into two categories: face to face and telephone interviews [13].

#### **3.3.1 Face-to-Face Interviews**

Also known as in-person interview. It is the most common and oldest way of collecting data. In this method, facial expressions and feelings can be clearly noticed. This method enhances response rate and guarantees quality of attained data. Though face-to-face interviews are effective but it is costly and takes time [63].

#### **3.3.2 Telephone Interview**

Interviewer communicates on telephone in accordance with pre-arranged queries. For this type of interviewing, closed-ended queries mostly suggested. Basically, it is short and concentrated on a gathering of focused information [64].

### **3.4 Case Study**

For researchers, case study is usually the most common and useful method [5]. It is a deep investigation process about person, condition and group etc. It contains gathering of data from different sources. Whenever deep understandability of problems required, case study method conducted. Mainly, it is applied in social science studies and address issues relevant to sociology, education and society related problems like unemployment, illiteracy and poverty etc. It is time consuming and cannot be replicated [85].

### **3.5 Survey**

Survey method is most commonly used to examine feelings, beliefs and opinions. It comprises of discovering facts in specific area of inquiry. Surveys can be directed in different ways: via telephone, face-to-face, through mail and electronically. For gathering data, surveys applied in several fields even in private and public zones. Researchers can conduct this method in different means based on the selected approach and study purpose. Survey method is time consuming.

### **3.6 Questionnaire**

Most common technique of gathering data from large group of persons. It comprises of questionnaires set hand over to communicator for solutions. It can be open-ended and closed-ended. Closed based questionnaires provide finite set of answers. In simple words, it contain some sort of scale like ordinal, nominal, ratio and interval and multiple choice type questions. In an openended questionnaire, data gathered can be concentrated on group methodologies furthermore group attributes. In which questions never based upon tick mark but there is blank space where respondent write answer. It is helpful to simplify and compute people conducts and attitudes [12,13].

### **3.7 Oral History**

Process of recording and storing data which is gathered from first hand person from their past memories and experiences. Interviewer asked questions and respondent responses are documented, which stored for future usage. Researcher uses audio-video tapes for interviews. The main difference among interview and oral history is that, interviews concentrated on a specific topic while oral history based on particular topic and it is less organized. Knowledge gained by this method is unique, because it shares the interviewee thoughts, beliefs, views and understanding.

### **3.8 Document Review**

It is a process of gathering data via reviewing current documents. Documents could be external or internal to a firm or a program. Documents can be in hard form and may be consists of newsletter, program logs, meeting time and reports etc. it is a best source of gathering historical data. Relatively, it is less costly. It takes time to gather, examine and review several documents. Also, data gathered via this method can be inappropriate or incomplete [74].

### **3.9 Ethnography**

It's a way of study of people and cultures in a systematic manner. In social sciences discipline, ethnography assumed to be the main method of data collecting. Basically provide deep understanding of rituals, behaviors and likings of different cultures and societies. The main goal of this method is to attain complete understandability of people's beliefs and activities, as well as the location type in which they reside, by

means of perceiving and questionnaires. Researcher directly contributes in data collection process [8].

### **3.10 Experiment**

A way of data collection which provide the understandability of cause and effect associations of variables. In other words, affected variable is said to be dependent variable while the variable which will manipulate the dependent variable is known as independent variable. This method applied in various research domains like sociology, agriculture and medical etc. Problem of using experiment method is that it is costly and takes time.

## **4 Data Analysis**

Analysis term denote processes through which data organized and summarized in such a way, that leads answer to research questionnaires. For further manipulations, data analysis organizes data in a suitable manner. Data analysis problem differs according to research study. Data analysis play a vital role in research domains and it is the process of placing statistics and facts to resolve research issues. Moreover, necessary phase of research is the exploration of data, which is getting from data analysis and helpful to create implications and extract outcomes [77]. Data analysis is a method to get order, structure and provide meaning to pile of gathered data. Although this method takes time, disorganized and ambiguous but also called captivating and innovative method [51].

Data analysis is the way of cleansing, reviewing, altering and forming data along with the aim of finding valuable information. It supports businesses to make decisions. In the present era, data analysis applied in various domains like science, business and education etc. Also, provide means of effectively performing businesses by proper decision making. To ensure integrity of data is the main component of data analysis.

### **4.1 Data Analysis Types**

There are several types of analysis provided along with diverse kind of goals.

#### **4.1.1 Descriptive Analysis**

The main goal of this analysis is to summarize and define data set. For conducting statistical analysis, Descriptive analysis considered to be the main step. It provide means of data distribution, help out to identify outliers and errors and generate correlations among variables. Best way to conduct this analysis is that, first need to decide about variable types after that select approach which will suitable according to types of variable. If descriptive analysis structured systematically, then it takes no time and easy to use [75].

#### **4.1.2 Exploratory Data Analysis**

Exploratory data analysis (EDA) is a critical procedure of conducting initial examinations on data in order to identify data patterns, discover variances and hypothesis testing through the assistance of graphical illustrations [42]. The goal behind

EDA is to discover new aspects of data. Also helpful to explore data and formulate hypothesis which further lead to experiments and collection of new data. For data visualization, some graphical methods are histogram, box and scatter plot etc.

#### **4.1.3 Inferential Analysis**

From data samples, inferential analysis permits to make inferences about large population. It comprises of two parts: hypothesis testing and estimating parameters. Hypothesis testing tells about where researcher or someone make usage of sample data to answer research questionnaires. In estimating parameters part, it takes statistics from data samples and utilize it to answer about parameters of population.

#### **4.1.4 Predictive Analysis**

For making future predictions, predictive analysis uses background or present data to discover patterns. Main objective is the usage of machine learning approaches and statistical algo's to assess future results which depends on past data. Predictions accuracy focused on the input variables. Organizations and companies used predictive analysis to enhance operations and productivity skills [46].

#### **4.1.5 Confirmatory Data Analysis**

Confirmatory data analysis (CDA) emphasizes on approving and negating current hypothesis. CDA, in which research examine evidence by utilizing conventional statistical instruments like dependence, importance and inference. It contains many things such as regression and variance analysis, hypothesis testing and to generate approximations along with quantified scale of precision. It is used to fail or approve theory of measurement.

#### **4.1.6 Casual Analysis**

Based on the cause and effect associations between variables. Cause denotes a reason. Goal of casual analysis is the identification of the origin of a problem rather than discovering symptoms. It is helpful to reveal facts which takes toward certain condition or state. With the support of different methods, casual analysis can be formulated.

#### **4.1.7 Mechanistic Analysis**

The goal of MA is the understandability of exact variations in variables that bring towards to the variations in other variables. Useful in engineering and physical sciences, needed high fidelity Measurement error is said to be the noise in data. Mechanistic analysis demands maximum effort. Applications which comes under this analysis type are randomized trial set of data.

### **4.2 Data Analysis Methods**

There are different analysis methods according to business and technology perspectives, but two main methods are: qualitative and quantitative.

#### **4.2.1 Qualitative Analysis**

Qualitative analysis method focused on the questions like 'how', 'what' and 'why'. Such type of queries are addressable by qualitative means like attitude rating and

questionnaires etc. Qualitative analysis mostly conducted in texts and descriptions means, also contain audio and video illustrations.

#### **4.2.2 Quantitative Analysis**

Quantitative analysis method utilizes statistical and mathematical modeling. It is measureable in number forms. Quantitative analysis make usage of statistics to find out answers of research queries in terms of numbers and define the purpose of numbers in respond to research related questions. Some other methods are given below.

#### **4.2.3 Text Analysis**

Text analysis is a method of data analysis to gain machine understandable facts and to identify data patterns from large sets of data by the usage of data mining techniques and via databases. The goal behind this method is the generation of structured data from unstructured content. This method is said to be information extraction, text mining and text analytics. Vagueness of human dialect is considered to be the major challenge of text analysis. There are variety of text analysis software tools. Some of them are: apache OpenNLP, visual text, genism, distributed machine learning toolkit, google cloud natural language API, KNIME text processing and general architecture for text engineering-GATE etc.

#### **4.2.4 Statistical Analysis**

Statistics comes under the mathematics branch and deal with data gathering, data interpretation, data analysis and data demonstration whereas biostatistics belongs to statistics branch where statistical approaches applied on biomedical data that lead towards ultimate solution. Statistical analysis play a vital role and considered necessary for those types of researches which contains statistics as a part of research methodology. It demonstrates 'what happen?' via usage historical data. It contains data gathering, data analyzing, data interpreting and presenting of data. Statistical analysis examines data set or data sample. It is applicable in situations such as collection of research expositions, plotting studies and surveys and statistical modeling [72]. It is suitable for BI related organizations because these organizations work with huge volumes of data. Statistical Analysis can be classified into two types: descriptive and inferential analysis.

##### *Inferential Analysis*

In inferential analysis, analysts examines sample from whole data. Via choosing distinct samples, analysts identify diverse type of conclusions from similar type of data. Also, it shows associations among various variables and make estimations.

##### *Descriptive Analysis*

In descriptive analysis, analysts examines whole data. For continuous type of data, it demonstrates deviation and mean while for categorical data type, illustrates frequency and percentage.

#### **4.2.5 Diagnostic Analysis**

It demonstrates 'why did it happen?'. Diagnostic analysis give in-depth analysis to answer queries. To discover behavior patterns of data, diagnostic analysis is feasible. It

is also known as root cause analysis and it contains procedures such as data mining, data discovery, drill down and correlations. In discovery procedure, analysts discover sources of data, then those data sources help out to understand outcomes. Drill down procedure concentrates on the certain data aspect. Data mining procedure basically extracts useful information from huge size of data sets. Correlation process denotes a degree in which variable pairs connected linearly [38].

#### **4.2.6 Predictive Analysis**

To identify future trends, PA uses historical data and gives that data to machine learning model. To predict what will happen in the future, model applied to existing data. Several companies and organizations choose predictive analysis to attain benefits. It contains methods like data mining, artificial intelligence, data modeling, machine and deep learning. Common uses include fraud detection, risk reduction and operations refinement etc. [38, 46].

#### **4.2.7 Prescriptive Analysis**

Prescriptive analysis is said to be a mixture of business rules and data. Data used for prescriptive analysis can be in external and internal forms while business rules are boundaries, best practices and preferences. In this analysis, analysts define which necessary action needs to be taken to resolve present issue. Many organizations prefer prescriptive analysis because it expands data performance. Analysts examine data and take necessary decisions to look at current issues and situations [38].

### **4.3 Data Analysis Tools**

Data analysis tools help users in processing, analyzing and manipulation of data, also examine associations between sets of data. Furthermore, helpful to discover trends and data patterns for making interpretations and draw conclusions. Selection of right data analysis tool is a challenging task because usage of same tools never fulfills necessary requirements in different domains. First of all, need to understand what type of data will be used for analysis purpose? After that, must be clearly define, in which form data will be presented because some studies concentrate on qualitative data and some work with quantitative data. So, different analysis tools are available for diverse type of data. Few of them are given below.

#### **4.3.1 Quirkos**

It helps users in sorting, managing and understanding text data. Users utilize Quirkos to code or tag related data units and match themes over hundreds of diverse sources. Furthermore, Quirkos supports qualitative researchers in analyzing and understanding data like articles, surveys and interview scripts. Some features of this program are: visual interface shapes coding easy and fast; execute comparative interpretation that highlights variances among respondent clusters and construct multifaceted queries to drill down within data; coding with Quirkos is simple and directly import pdf, word files and text. Moreover, Quirkos has built-in functionality of working with several languages and every script type. Quirkos is a good mode of work with small data. In Quirkos, everything is attractive, interesting, visual and colorful.

#### **4.3.2 Nvivo**

For qualitative and mixed method research, Nvivo provide support. It's a qualitative data analysis software. Nvivo is helpful for organizing, analyzing and discovering insights in qualitative and unstructured data like open-ended surveys, social media, interviews, web content and articles. Without Nvivo, when work with qualitative data, work will take time and difficult to navigate. Some features are: examine and import video, web data, images, online surveys and emails; coding, word frequency, text search, matrix coding and coding comparison queries; coding review; relationship coding; code and organize several sources of data into one main file; add notes and interpretations; data visualization and assign attributes to data. User interface and text analysis presented in different languages such as French, Spanish, English, Chinese, Japanese and English etc. For windows and MAC OS, Nvivo is suitable. It save time and speedily retrieve, store and organize data [6]. Researcher presented five necessary tasks through which Nvivo simplify qualitative data analysis. Tasks are: query and manage data, reporting, manage ideas and visual based modeling [33].

#### **4.3.3 WebQDA**

In distributed and collaborative environ, WebQDA software provide support for qualitative data analysis. Though in qualitative analysis, there are some software packages which address unstructured (image, audio etc.) and nonnumeric data and some others packages can be utilized various researchers in a distributed and collaborative work environment.

WebQDA software directed researchers in diverse situations that require analysis of qualitative data, asynchronously and synchronously, collaboratively and separately. Features of webQDA are: simple in usage; provide security of data; feasible for several research kinds; web-based software and well-suited for all OS's. It follows other programs design that accessible in market and creates difference via providing real-time, online cooperative work and service which assists process of research. With the help of WebQDA, documents editing, linking, viewing and organizing can easily be done [17].

#### **4.3.4 HyperRESEARCH**

For qualitative analysis, it is considered to be the powerful tool. It is reliable and flexible. Tool is simple to learn and use, cost-effective and cross-platform. Without usage of additional powerful features, researcher can easily learn how to retrieve data and data coding. Cross-platform abilities allow consumers to utilize it anyplace or on every system (computer). It provides innovative features which give control and whole accessibility of data. Also, provides flexibility in respect of how someone can get access of data. HyperRESEARCH case oriented architecture gives adaptability to auto-code several sources to several cases, applying prescribed codes to different phrases and keywords via single command. Code map construct visual illustration of codes and their associations and behave just as a robust code selection tool. With the help of statistical software, code frequencies generated and then export for future analysis. Moreover, memos can be written for adding advanced information regarding codes. It automatically generates single file out of study file. It is in-expensive and interface is easy [32].

#### **4.3.5 Transana**

It provides advanced tools for qualitative data (audio, image and video) analysis. It offers two modes of demonstrating analytic importance of client's data units, coding and categorization. In powerful means, users can use these analytic systems just as users work to generate comprehension of their data which is qualitative. Transana's text and graphical based reports are immensely customizable and flexible and allow users to discover analytic links inside data. Some features of Transana are: in single analysis, it incorporate audio, video, text and image based data; handle multifaceted media data in powerful modes; by the usage of multi-user version and cooperate with friends in real time across distance. Within Transana, users can examine data and can import text based documents. Primary aim of Transana is to simplify management and analysis of digital based data. It produces different types of reports which is useful in delivering complete vision of data and could be exported for future analysis into other tools (spreadsheet etc.) [53].

#### **4.3.6 SPSS**

Statistical package for social sciences (SPSS) software execute quantitative analysis. This software write and read data from other spreadsheets, statistical packages and databases. SPSS software only execute statistical operations. It is mostly applied in social science areas like psychology. In psychology field, techniques like t- test and cross-tabulation etc. presented within 'analyze' menu. SPSS software GUI has two sort of views like data and variable view. Data view displays spreadsheet sight of variables and cases (columns and rows). Variable view allows user to modify it via data kinds and comprises of various headings such as type, label, name, decimals, align, measures, width and missing. SPSS limitation is that not applicable for analyzing huge set of data. In nursing and medicine fields, researchers work with huge data sets, so SAS software used in both areas rather than SPSS for analyzing clinical data. This software covers huge range of statistical techniques like data summarization (standard deviation and calculate means), determine associations between variables (regression and correlation) and graph outcomes (line graph and bar charts) [62].

#### **4.3.7 SQL**

SQL stands for 'Structured query language'. It's a programming language. Suitable for data management which resides inside relational databases. To handle data that is structured, SQL is feasible and for analysts behave like a database instrument. Applicable in data science domain. The aim behind is, mostly data kept into relational databases, when someone require to attain and need to expose its value, SQL helps out to perform such operations. For decision making, data analysts make use of SQL to modify, access, read and analyze data which is kept inside databases. Diverse types of SQL based RDBM systems presented. Few of them are: PostgreSQL, Oracle, MySQL and MS SQL etc. There is a wide range of data visualization tools for SQL. Some include: metabase, tableau, chartio, google data studio, andlooker etc [24].

#### **4.3.8 SAS**

SAS is said to be the data analysis software. It provides variety of predicting techniques like: event modeling, scenario planning, what-if analysis and hierarchical

reconciliation. SAS choose variables which utilize for modeling method and creates forecasts for future. Statistical analysis system is a software package which designed to perform different operations such as mining of data, data modification, business intelligence and data management etc. It repossess data from several sources and conduct statistical analysis on that data. To the feasibility of non- technical consumers, SAS provides graphic interface which is said to be point and click interface. SAS programs consist of PROC and DATA phases. PROC phase used for data analysis while DATA phase performs manipulation and data retrieval operations. Every phase involve chain of statements. DATA phase subdivided into two phases: execution and compilation. Compilation phase discover syntax errors and execute declarative statements while phase of execution emphasizes on the linear execution of executable statements. Sets of data arranged in table forms. Rows are said to be observations while columns are known as variables. PROC phase comprises of PROC statements. PROC statements uses named procedures. Procedures basically execute analysis task and reporting on sets of data to generate graphics, statistics and analyses. PROC based statements suitable for showing results, data sorting and for other tasks. Over three hundred named procedures exist. Every named procedure include significant form of statistical work and programming. SAS data available in different layouts like PDF, RTF, Excel and HTML etc. SAS software package include beyond two hundred modules. Few SAS modules are: quality control, data mining, statistical analysis, graphics and presentation and applications facility etc [66].

#### **4.3.9 Python**

Python is a general purpose and high level programming language. Python syntax is easy and simple. For data cleaning, data modeling and for creating analysis algo's, python language used in data analysis process. Main feature of this language is that, it is user friendly. Moreover, it is portable. Users only write code in python language and run it into different platforms without any changeability. Programmers do not require to memorize system architecture and neither handle memory. Wide range of components, libraries and packages make python language as a useable language and many companies like Netflix, Spotify, Reddit and Dropbox use python to perform operations. Furthermore, with machine learning and text mining features, python play a necessary role in advanced analysis procedures. Some popular visualization packages are: Plotly, Seaborn and Matplotlib. Python interface also utilized for other analytics based systems [54].

#### **4.3.10 R**

R is said to be computing environment which emphasizes on the graphics based data representation and statistics. Moreover, it is an open source programming language interface. It comprises of more than fifteen thousand source packages, most of them used for data manipulation, data visualization, data loading and data modeling. R environment permits analysts with programming proficiency to perform and generate any kind of analysis on data. It is the high level data analysis tool, stated to as language and statisticians designed it. It can run on several platforms such as Linux, Macintosh OS, UNIX and windows etc. It is suitable to perform different types of statistical base analysis like factor cluster, regression and conjoint analysis etc. R language is easy for beginners who have no skills of high level programming language. Via usage of single

command, R execute complex mathematics problems. In statistical area, R language with widerange of graphics libraries like Plotly, shiny, Knitr, XGBoost and ggplot make R different to others. Now a days, R is widely used in companies and business industries like Facebook, google, Airbnb and twitter etc [35].

#### **4.3.11 MATLAB**

Matlab (matrix-laboratory) is a programing language as well as analytical platform. All aroundthe world, large range of scientists and engineers make use of Matlab for designing and analyzing systems. For expressing computational mathematics, matrix oriented Matlab is preferable.Integrated graphics make this language easier for data visualization. Desktop environs providediscovery, experiments and explore operations. InMatlab, analysts can perform analysis on huge setof data. Primary feature of Matlab is that, its codecan combined with other languages that leads to employ applications and algo's in business, web and manufacturing systems. Without knowing about data layout and source, Matlab analyze massive amount of complex data in finer ways. Matlab analyze different kinds of data. Some data types are: signals, time-series, images, geospatialand dates and times [41].

#### **4.3.12 Java**

It is an object-oriented programming language.Over three billion devices execute java. It helps programmers to write once, run anywhere (WORA), it means java code is executable on variety of platforms like windows, mac and Linuxetc. Java syntax is same as C++ and C languages.It is suitable in data science domain and dataanalysis processes like import and export of data, cleaning of data, natural language processing, visualization of data and statistical analysis. For the necessity of data engineering, numerous social applications based on java which includes twitter,Facebook and LinkedIn etc. It contains various libraries and instruments accessible for machine learning and data science domains [14].

#### **4.3.13 Datapine**

It is most common business intelligencesoftware. It provides advanced analysis features for whom that require accurate and speedy onlinedata analysis solutions. By using this tool, analystscan easily combine diverse types of data sources, can run advanced analysis operations and can generate interactive dashboards. With the help of data analytics software, no one need to manually analyze huge data sets. Datapine consists of advanced level of predictive analytics features. Some features of datapine are: data alerts, drag- drop interface, self-service analytics, modern dashboards and ad-hoc queries etc.

#### **4.3.14 Data Visualization Tools**

Data visualization play a vital role in analysis process. It is the graphics based demonstration of data. Visualization tools basically provide a feasible mean to understand and to see data patterns and trends. Researchers and analysts, need to design and develop visual illustration of data. When analysts work with large volumes of data sets, procedure of making visualization makes analyst task simpler and easier. A lot of data visualization tools are available. Few of themare: charts, graphs, tables, maps and world cloud etc.

### *Charts*

For organizing huge range of data, charts are suitable. Charts are pictures and tables. People utilize charts for interpreting present data and for making future predictions. It is said to be the graphical data representation. In chart data represented via symbols, like lines in line chart, slices in pie chart and bars in bar chart. It can also represent functions, numeric data and give information. For understanding huge data volumes and associations among data parts, usually charts utilized. Charts applied in different domains and can be made via the help of computer utilizing chart base applications and by hand. Different types of charts available for data representation. Few of them are: line, bar, pie, column and histogram etc.

### *Graphs*

Graph is a kind of chart used to illustrate association among diverse data sets via devising vertical and horizontal axis. Graphs are subset of charts. Graphs utilized for raw data and show visual representation of variations and trends within data over time interval. Graphs signify mathematical relations or interconnections among different data sets. There are different types of graphs such as bar and line graphs etc. and useable for various purposes.

### *Word Cloud*

A word cloud is a cluster or group of words portrayed in various sizes. Also said to be text or tag clouds. It is a word visualization which shows most useable words inside text from tiny to huge, in accordance with how frequently each seems. To perform word cloud, first phase is the extraction of words. It's a good concept to remove punctuation and plurals, correct spelling errors and automatically associate words which are same. Various text based analysis systems are available that require word clouds. For instance opinion mining, investigate and patent analysis [70, 82]. For the visualization of text data, word clouds are feasible and easy in utilization.

## **4.4 Research Techniques**

Research term initiated from old French word "recherchier". It means search and search again. In common dialect, research denotes search for knowledge. It can be described as a systematic and scientific search for appropriate information on a particular area. Research is basically a skill of scientific exploration.

[40] state research as a structured determination to attain new knowledge. Few people deliberate research as a movement, movement from identified to unidentified. Actually it is a journey of discovery.

It is a scientific method of answering a research query, resolving an issue and creating new knowledge by methodical collection, analysis and organization of information with crucial aim of making research beneficial in decision making. Systematic research within any domain of inquiry contains three operation such as data analysis, collection and analysis of data.

In computer science research different data collection and analysis techniques are available. Few research techniques are: qualitative research, non-empirical research, qualitative research and engineering research techniques [21].

#### **4.4.1 Quantitative Research Techniques**

Quantitative research is the way of analyzing and gathering numerical based data. It could be utilized to identify averages and patterns, test causal association, make predictions and simply outcomes to broader populations. It's generally applied in social and natural sciences such as economics, biology, marketing, chemistry, sociology and psychology etc. Quantitative research techniques can be used for forecasting, laboratory experimentation and field experimentation research.

#### **4.4.2 Qualitative Research Techniques**

Process of analyzing and gathering non-numerical based data to comprehend experiences, concepts and views. This type of research can be utilized to collect in-depth visions into a problem and produce new concepts for research. Qualitative research is generally applied in subjects of humanities and social sciences like education, history, anthropology and health sciences etc. Qualitative research techniques are applicable in action, descriptive, ethnographic and focus group research.

#### **4.4.3 Non-empirical Research Techniques**

It's a method which contains theory development as opposed to utilizing experimentation and observation. This type of research finds solutions to problems via usage of present knowledge as its source. Although, it never means that new inventions and concepts cannot be found inside established and current knowledge pool. It plays significant part in scientific community by providing a place to start procedure of creating scientific evidences in new research areas. Non-empirical techniques involves theorem proof, review of existing research, conceptual research and future research.

#### **4.4.4 Engineering Research Techniques**

Engineering not only concerned with knowledge of natural phenomena but concerned with how knowledge assist human requirements. Such variables like user compatibility, safety, cost and adaptability to several external operating environs and situations must be taken into account in development, designing, maintenance and operational support of services and products that engineers produce. Hence engineering contains incorporation of knowledge, methods, experiences and techniques from various areas.

### **4.5 Data Analysis Techniques**

In previous sections, discussed different analysis methods. This section describes different techniques of data analysis in terms of qualitative and quantitative. Here we will see a brief description of different techniques which mainly focused on qualitative and quantitative data analysis.

#### **4.5.1 Qualitative Data Analysis**

Qualitative Data Analysis (QDA) is considered to be the procedure of cataloging, depiction and linkage of phenomena along with researcher's thoughts and ideas. Phenomena beneath study must be defined accurately. QDA is said to be an iterative procedure, specify collection, analysis and processing of data operations are entwined and not succeeding procedure [61]. QA converts data into findings. It comprises of

reduction of huge amount of raw data, discover useful data patterns and develop a framework. In QDA, three steps need to be follow: description, organizing and interpretation of data [65].[80] QDA process consists of six stages: data discovery and data defining, data gathering and data storage, data sampling and reduction of data, data coding and structuring of data, testing and the last is to write up research. Researchers must be proficient in data interpretation and explanation, so, there is a need of designing and developing a conceptual framework and data need to be classified. Although for quantitative data analysis, there are some rules which need to be follow but for qualitative analysis, there are no clear rules. Some techniques for qualitative data analysis introduced, these techniques helpful for researchers from different perspectives [7].

#### *Theoretical Propositions*

Theory is a proper, testable description of events, contains the details about how things associate with each other. Moreover, researcher describe the term preposition. "Propositions are said to be statements deal with associations between theories". Proposition describes logical connection between thoughts via declaring a universal link among theories [86]. In scientific method, propositions play necessary part. Proposition is same as hypothesis, but the main aim of proposition is, recommends a connection among two theories but that connection never tested with the help of experiment. So, proposition fully based on previous research. Proposition never based on testable data, so there is more difficulty to negate in scientific environment. In researcher's point of view, for example, if stated reaction is a result of an act, then this study comes under theoretical propositions. Researcher define, if certain proposition is wrong and if variations exists among conflicting concepts. The same way, records made for breaking, theories made for testing [86].

#### *Triangulations*

It means, researchers gain data from several sources and apply different types of techniques [7]. From doing this, researchers attain knowledge by using different techniques [30]. In triangulation, researcher utilizes one measurement method then matches findings of that method with that one, utilizing other method. If other method proves findings of first method, then triangulations reinforce finding reliability. If it never happen, then it means usage of one measurement method is not considered to be reliable [11]. According to [23], triangulation technique is expensive in terms of effort. Researcher [20] presented four kinds of triangulation. Theoretical, data, methodological and investigator triangulation. In methodological triangulation, via utilizing two methods data gathered. In investigator triangulation, data collected and analyzed via the help of two researchers. In data triangulation, through diverse type of sampling approaches, data collected. In theoretical triangulation, with the help of two theoretical based position data elucidated.

#### *Grounded Theory*

Theory which obtained from data, gathered systematically and examined via research procedure. Theoretical development is the primary goal of grounded theory. In grounded theory technique, collection of data, data analysis and ultimate theory has direct association with each other. Researchers presented two main features of grounded

**Table 2.** Coding variations between three content analysis approaches.

Content analysis type	Study begins with	Timing of defining keywords or codes	Keywords or codes source
<b>Summative content analysis</b>	Keywords	Keywords recognized during and before analysis of data	Keywords extracted from literature review and researchers Interest
<b>Conventional content analysis</b>	Observation	Codes defined during analysis of data	Codes extracted from data
<b>Directed content analysis</b>	Theory	Codes defined during and before analysis of data	Codes extracted from related research findings and Theory

theory. One of them is that, this technique is recursive. Recursive means, analysis and collection processes work in a circle and other one is that it's concerned with the theory development out of data [7]. All around the world, grounded theory is considered to be the most popular research design. It begins with broad and clear research questionnaire. Then questionnaire discovers area for study. Researchers presented a list of different methods of analysis and gathering of data: data classification and initial coding, producing simultaneous data or data analysis and data gathering, intermediate coding, memos writing, theoretical integration, sampling, sensitivity and saturation, selection of primary class and comparative analysis through the usage of abductive and inductive logic [57]. Data is obtainable from different ways such as videos, observations, books and documents etc. to ensure analysis validity, all type of data could be coded [16]. When in a study apply grounded theory and its approaches scholars can easily describe procedure and scheme linked with phenomenon [57]. Researchers define eleven rules which must be followed throughout analysis and data gathering for ensuring grounded theory application reliability. Meanwhile, data analysis and data collection are interconnected procedures, first of all data gathered, then it takes time in analysis due to future interviews and observations influences. Concepts are important part of analysis process, because in the whole procedure of analysis, researchers require data conceptualization to work with that data. Moreover, there is a need to develop categories and must be linked with each other. Difficult for researchers to memorize all hypothesis and categories, so 'writing memos' play a vital role in grounded theory to monitor all aspects of study. Systematic and sequential way of analysis and data gathering as biggest privilege of grounded theory [18]. Grounded theory method takes time [7].

*Content Analysis*

Content analysis [44] is a research based technique which is used to make proper inferences from texts. It performs well when to deal with data like open-ended surveys, interview based data and user feedback etc. Data which need to be analyzed could be in any form like recorded form or in written form.[71] Content analysis is a dominant technique of data depilation. The main goal of this technique is that, it is replicable and systematic approach which compresses text words into smaller groups depending upon

explicit coding rules. It deals with huge data volumes [7]. For different type of unstructured data, this technique is suitable [36]. Presented three approaches of content analysis. Main difference between approaches are schemes of coding and coding origins.

#### *Framework Analysis*

Framework analysis technique playing an important role in health domain. It gives evident and systematic phases to analysis procedure, in such a way, researchers have clear perception about phases through which outcomes can be acquired from data. Framework analysis consists of five phases. These are: familiarization, charting, indexing, thematic framework identification and interpretation and mapping. In familiarization phase, partial and complete understanding of data included. Discovery of thematic framework, it is primary coding framework. Designed from emerging and theoretical issues from the phase of familiarization. In indexing phase, apply framework to data, with the help of textual and numerical codes for the discovery of particular data chunks. In the phase of charting, get headings from framework for creating data charts, in such way someone can simply read complete sets of data. In the phase of interpretation and mapping, analysts find relationships, explanations, concepts and patterns of data supported with the help of plots and visual demonstrations. Analysts design plans, generate topologies, define concepts, provide clarifications and discover links inside data at this stage [68]. Researchers provide variety of data display concepts which helpful to explore data in framework analysis terms [47,56].

#### *Discourse Analysis*

Discourse is considered to be the complete system through which individuals communicate, called 'language'. It contains both type of communication (verbal and non-verbal, written). Discourse analysis is all about examining language inside its social perspective. This technique is different from other qualitative analysis techniques that effort to comprehend meaning of societal actuality for artists in that it strives to reveal the mode inside which that reality generated [25]. It's an analytic technique instead of theory. It has many disciplinary roots such as anthropology, philosophy, sociology, communication studies, literature and linguistics etc [29]. Researcher identified three types of discourse analysis. Critical, formal linguistic and empirical [34]. On any data set such as (written and spoken) this analysis can be conducted. It contains focus groups, interviews, speeches and other form of texts.

#### *Ethnographic Analysis*

It contains discovery of categories associated with population, human life specially education, health related issues, family and atmosphere [55].

This analysis technique utilizes an iterative procedure [77]. In an iterative process cultural concepts rise are translated, converted and characterized in written form. Nurses applied ethnographic techniques in health domain for discovering and recording dissimilarities in how diverse type of culture and social communities accept and comprehend disease and health.

#### *Constant Comparative Analysis*

Constant comparative analysis technique takes one data piece (interview, theme and statement, etc.) and match that data piece with others that can be same or diverse, for

**Table 3.** Comparison among qualitative analysis techniques.

	<b>Ethnographic Analysis</b>	<b>Grounded Theory</b>	<b>Framework Analysis</b>	<b>Discourse Analysis</b>	<b>Content Analysis</b>
<b>Research question and Aim</b>	Involvement in natural setting to attain insider capability e.g. uptake of medication, service estimations	From empirical data, generate theory e.g. stigma in mental health; views and perceptions of mental in diverse tribal groups	Mostly utilized for problem based approaches in health based services research e.g. what are the training needs for primary care staff?	Capture variations of text or Public Discourse e.g. political theory understanding. Social changeabil- Ity	Capture the meaning at Descriptive level e.g. why do carers attain facilities for Their families with anorexia?
<b>Methods and sampling</b>	Observational studies	Unstructured question-naire Theoretical sampling	Convenience sampling Interviews with semi-structured question-naire	Theoretical/Purposeful sampling Speeches, newspapers, documents	Purposeful sampling Documents e.g. newspapers
<b>Analysis</b>	Data driven; but no fixed commitment to developing new theory	Iterative and constant comparative approach Data driven	Deductive approach Theory driven	Complete, in-depth analysis of discourses, written text, conversations, speeches	Deductive approach
<b>Researchers position</b>	Researchers skill and neutral position vital	Researchers position or potential bias is managed	Researcher neutral position, limited interpretation	Expected high level of abstraction or interpretation	Researcher neutral position

developing conceptualization of feasible connections among several data chunks. According to [27], this technique allocates codes which reflect methodological associations. Constant comparative method consists of four phases: compare incidents appropriate to each group, theory delimiting, theory writing and integrating groups and properties.

*Narrative Analysis*

It utilizes texts like conversations, stories, photos and interviews etc. Narrative analysis behave like a device in domain of arts, education, cognitive science and sociology etc. it concerned with stories content and stories structure presented in verbal and written form [4, 19]. Narrative analysis technique utilized for content analysis which collected from several means like surveys, field observations and interviews etc. Mostly stories and views shared through people concentrated on the discovery of answers to research queries [77].

*Sentiment Analysis*

It is said to be the systematic analysis of expressions. Specially, it emphasizes on examining conducts and concepts on area of interest via usage of machine learning

approaches. In data mining domain, sentiment analysis could be explained from two ways: operational and functional. Functional way emphasizes on the practical usage of technique. For example, sentiment analysis is a strategy which classifies body of documented information for defining conducts and emotions regarding certain problem. This SA definition, describes the way of working. Operational way emphasizes on the method operations just as subarea of computational linguistics. [45] Sentiment analysis is same as opinion mining which emphasizes on categorizing and extracting texts with the help of computer programs and machine learning. In simple words, it is a data mining approach which utilizes computational linguistics, text analytics and NLP for the identification and extraction of content of heed from textual data. Sentiment analysis is the most useful and applicable qualitative technique. The goal behind it, classification and interpretation of emotions delivered inside documented data. Sentiment analysis integrates several activities to generate knowledge from textual data, in processing terms. This processing classified into five phases: collection of data, text preparation, sentiment classification and detection and output demonstration. In data collection phase, sentiment analysis gain benefit from user created content on internet. Data can be in different forms, like emotions and behaviors etc. So, for extracting and classifying data, NLP and text analysis operations utilized. After data extraction, data will be arranged for analysis purpose. Text preparation phases contains data cleaning which is extracted from data collection phase before performing analysis. In text preparation, elimination and identification of non-textual content performed and removed irrelevant content from textual dataset. In sentiment detection phase, with the help of computational tasks, reviews, beliefs and concepts extracted from textual dataset. Sentiment detection performed at diverse levels such as phrases, complete sentences and documents and single term with the help of different techniques like: negation, lemmas and opinion words etc. Sentiment classification phase, classifies every particular sentence into classification sets. These sets signified on two utmost points on a scale (positive and negative, like and dislike good and bad). Classification consists of various points same to star rating utilized via retailers, hotels and restaurants. Fifth phase show output. Several ways available for output presentation, but the most common of them is the usage of graphical based displays. Output presentation format based on research requirements.

#### **4.5.2 Quantitative Data Analysis**

Quantitative data analysis is measurable in number forms. Quantitative data analysis techniques concentrate on numerical, mathematical and statistical analysis of large sets of data. It contains statistical data manipulation with the help of computational algorithms and techniques. Quantitative analysis techniques mostly utilized for the explanation of certain phenomena and for making predictions. [43] Researchers provide a list of similarities among quantitative and qualitative data analysis techniques. According to researchers, both methods are same in four ways. These include: try to escape from errors and wrong conclusions, discovery of aspects and patterns which are same or diverse, reveal study design in a useful way and make usage of reasoning to reach on some conclusion. Different quantitative data analysis techniques are available. Some of them are as follows.

### *Regression Analysis*

For describing biological phenomenon, regression term invented in 19th century by Francis Galton. It is quantitative data analysis technique that utilized when researchers require modeling and analysis of variables and relationship among variables. Regression analysis basically examine association among dependent (known as outcome variable) and independent variables (said to be features, predictors and covariates). The goal behind is to evaluate how one or more variables effect on dependent variable for the identification of patterns and trends. It is feasible to make predictions regarding future trends. This statistical technique also utilized to comprehend which independent variables are connected to dependent variables and help out to discover relationships. Regression analysis could be utilize to understand casual relations among variables but this may actuate to wrong associations. Several kinds of regression models are available but selection of right model based upon data type. Before performing analysis, every regression technique consists of few assumptions which must be fulfil. These techniques vary regarding the kind of independent and dependent variables.

**Regression Analysis Types:** In medical domain, usually three kinds of regression analysis used. These are linear, cox and logistic regression. Linear regression is the easiest model to use because its demonstration is easy. A technique in which dependent variable is said to be continuous. Connection among dependent and independent variables supposed to be linear. For predictive analysis, linear regression is the most useable technique. Linear regression is sub-divided into two types: simple and multiple. When worked with the help of one dependent and one independent variable then it is said to be simple linear regression. On the other side, when used one dependent and more than one independent variable known as multiple linear regression. It is feasible when dependent variable is said to be binary in nature but if it is non-binary then multinomial or polynomial logistic regression can be used. Logistic regression not require linear connection among variables. Special type of regression and feasible for 'time-to-event' data. For instance, time from first heart attack to second [28].

### *Monte Carlo Simulation*

It is also said to be Monte Carlo method. It is a computerized method suitable for producing models of viable outputs and their probability dispersions. Basically it considers variety of viable outputs and evaluates how specific output would be comprehended. This method is feasible for data analysts for conducting advanced level of risk analysis which allow data analyst to forecast about what can happen in future and take decisions. It is the commonly used method for estimating influence of random variables on a particular outcome variable which makes this method suitable for risk based analysis [67].

### *Factor Analysis*

A technique which reduces huge range of variables to fewer range of aspects. Its working depends on that several distinct, evident variables relate with one another because all connected with fundamental set up. It is feasible for two reasons. One is it reduces huge sets of data into fewer one and second is it expose concealed patterns. [15] FA utilizes mathematical processes for identifying patterns from variables set. [84] Primary aim of factor analysis is the data summarization in such away that patterns and

associations could easily be understood and interpreted. It is applicable in numerous domains like medicine, geography, social and behavioral sciences and economics. It permits researchers to explore thoughts which could not be perceived and quantified easily like joy, wealth and satisfaction etc.

**Types of Factor Analysis** It can be divided into two types: exploratory and confirmatory.

**Exploratory Factor Analysis** Main goals behind exploratory factor analysis are: determine the quantity of common aspects effecting variables set and to determine the power of association among every aspect and every perceived measure. It is suitable to expose the underlying arrangement of relatively huge variables set. It attempts to reveal complex patterns via discovering sets of data and predictions testing [15]. Some computer programs like SPSS 'statistical package for social sciences' and SAS 'statistical analysis system' permit for the usage of exploratory factor analysis. Although every program offers diverse type of techniques and options but some phases are same like rotation, extraction, interpretation and selection [73]. Via usage of two methods (Q and R type factor analysis) exploratory factor analysis can be done. In Q type analysis, factors evaluated from single respondent while in R type analysis, factors evaluated with the help of correlation matrix.

**Confirmatory Factor Analysis** It tries to approve hypothesis and for the representation of factors and variables utilizes path analysis diagrams. It works just as a tool which is applied for the approval and rejection of theory of measurement. Two sorts of model need to be explored in confirmatory factor analysis: structural equation and measurement model. For several areas of behavioral and social sciences, CFA considered compulsory tool of analysis. CFA comes under the family of structural modeling approaches which permit for examining casual associations between observed and hidden variables. Primary objective of CFA is that it helps researchers to observe slot among observation and theory. SEM behave like an umbrella and it comprises of several techniques for data analysis. Some statistical analysis examples found comes under the SEM umbrella like multi-group confirmatory factor analysis, confirmatory factor analysis, for longitudinal analysis-latent models and regression with the help of latent variable outputs. Confirmatory factor analysis is an important phase to perform SEM model types [10]

### *Cohort Analysis*

Term cohort defines cluster of people. Cohort study is referred to as research paradigm focusing on a specific cluster with same characteristics and perceives over time interval. Cohort study can be classified into two types: prospective and retrospective cohort study. Prospective cohort studies are said to be scheduled earlier and accomplish over future division of time. Retrospective cohort studies examine data that available and attempt to discover risk factors for specific situations. Elucidations are finite because scholars cannot collect lost data and never go back [52]. It is a kind of behavioral analytics which accesses data from dataset and instead of observing all users just as single component, for analysis it splits them into associated sets. These cohorts or associated sets basically share mutual experiences and attributes under a fix time period. Cohort is basically a set of individuals who share mutual features. For instance, students who joined university in 2020, denoted to as cohort 2020. For analyzing client conduct, corporations apply cohort analysis. Without this, companies face problems to

understand life process that every client experience over a certain time interval. Companies perform cohort analysis for understanding patterns and trends of clients over time. There are some phases which helps to execute cohort analysis properly. These phases are: define what query you need to be respond, determine metrics which will be helpful to response against query, determine particular cohorts which relate with each other, perform cohort analysis and the last phase is results testing[52]. There are three kind of cohorts: time, size and segment oriented cohort. During a given time interval, time oriented cohorts are those when client registered for a service. Cohort analysis depicts client conduct based on the time when clients begun to use company service and product. Segment oriented cohorts are said to be those clients who bought and paid for a particular product and service in past. It clusters clients by product kind and service level they registered for. Client needs might be different according to sign up for basic and advanced level services. Size oriented cohort denote numerous sizes of clients who buy company service and product. Clients might be enterprise level, small and middle sized businesses.

#### *Time Series Analysis*

It is a statistical based technique utilized for the identification of cycles and trends over time. Timeseries data is an arrangement of data points that measure identical variable at diverse time intervals. Time series is said to be the gathering of observations made consecutively via time. Time series contains single variable records is said to be univariate while time series holds more than one variable records known as multivariate. Time series could be in two forms: continuous and discrete. Discrete series involves observations which measured at separate points of time whereas continuous series contains observations which measured at every instance of time [1]. Several forces and causes which influence observation values within time series are modules or components of time series. Four components are: seasonal, random, trend and cyclic modules. Trend illustrates the common tendency of data to reduce or enhance throughout extensive time period. It is noticeable that propensities may enhance, reduce or stable in diverse time units. Cyclic changes defines medium term variations in time series, affected via situations which replicate within cycles. Cycle duration extends over extensive time interval, commonly more than one year. For instance, business cycle involve four steps: depression, decline, recovery and prosperity. Seasonal changes are variations in time series happened within a year throughout season. Primary aspects causing seasonal changes are: weather conditions, customs and climate etc. For producers, and businessmen, seasonal variation play a vital role for developing future plans properly. Irregular variations initiated via uncertain effects, which never reprise in a certain pattern and also not regular. So random variations happened due to flood, war, earthquake and strike etc. No statistical technique available to measure random variations [1].

#### *Cluster Analysis*

Several models of data mining based on some notion of resemblance among information pieces encrypted in concerned data. Numerous names suggested for methods of clustering, relying on the area of application in data science. For instance, 'numerical taxonomy' term applied in biology[76], 'Q analysis' term often applied in psychology, 'unsupervised pattern recognition' term is applicable in artificial

**Table 4.** Comparison among qualitative and quantitative data collection methods.

<b>Qualitative Data Collection Methods</b>	<b>Quantitative Data Collection Methods</b>
Unstructured or semi-structured Techniques	Fixed and more structured techniques
Not test oriented or instrument	Test oriented or instrument
Not measureable	Measurable
Not utilized for statistical based Tests	Utilized for statistical based tests
Small sample size	Large sample size
Data generally ordinal or Nominal	Data generally interval or rational
Text based to gather Information	Number based to gather information
Interview forms contain open-ended questionnaires	Interview forms contain close-ended questionnaires
Lack of strong scientific control	Applies strong scientific control

intelligence research and ‘segmentation’ term used market researchers. ‘Cluster analysis’ term applied as a step in the process of knowledge discovery [78]. Cluster analysis discover structures inside set of data. The objective behind is the sorting of diverse datapoints into clusters internally similar. In simple words, data points inside a group are same with one another and different to data points inside another group. Clustering is useful to understand how data dispersed in a certain set of data or for other algorithms just as preprocessing phase. For statistical analysis of data, cluster analysis is the most popular technique and highly useable in various domains like information retrieval, bioinformatics, pattern recognition, computer graphics, machine learning, data compression and data mining. Clustering methods is sub-divided into two sets: partitional and hierarchal clustering methods [49]. Partitional methods concurrently split group of data points into sub-groups while hierarchal cluster methods find consecutive clusters by utilizing previously conventional clusters. Different types of clusters are available. These are: centroid, connectivity, distribution and density clustering. Clustering has huge quantity of applications spread over several fields. Some popular clustering applications are: image and market segmentation, anomaly detection, social network analysis, medical imaging and search result grouping. Clustering is said to be an unsupervised machine learning technique but it can be utilized for improving accuracy of supervised machine learning algorithms.

## 5 Conclusion

Evaluation in every research work focuses on information collection and then analyzing different scenarios. Success and quality of research work is based on the knowledge of data collection application and methods. Data type is significant in statistical analysis and appropriate knowledge of is compulsory for analyzing sets of data with suitable selection of methods. Qualitative and quantitative information falls in a continuum and differs according to the kind of data, collection method, tools and approaches of data analysis. This paper presented diverse kinds of data collection methods which can be applied in any evaluation along with complete depiction of numerous techniques and

tools employed by researchers for data collection. Starting with the explanation of secondary and primary data, quantitative and qualitative data and every data collection method is defined in detail. Various methods which have been described with focus on groups, observation, case study, experiment and interview etc. Moreover we showed holistic view of different types of qualitative and quantitative tools which can be used for analyzing data. Numerous statistical analysis software packages are available like SPSS and SAS etc. With the help of these packages, it becomes simple and easy to obtain output and interpret results. Researchers must interpret diagrams, tables and statistics accurately. The main purpose of this paper is to differentiate and elaborate the different quantitative and qualitative techniques in computing discipline.

## References

1. Adhikari, R., Agrawal, R.K.: An Introductory Study on Time Series Modeling and Forecasting. arXiv preprint (2013). doi: 10.48550/ar.
2. Oluwatosin Ajayi, V.: Primary Sources of Data and Secondary Sources of Data. Benue State University (2017)
3. Antonius, R.: Interpreting Quantitative data with SPSS. Sage (2003)
4. Bamberg, M.: Narrative Analysis. APA Handbook of Research Methods in Psychology, 3 (2010). doi: 10.1037/13620-006.
5. Baskarada, S.: Qualitative Case Study Guidelines. The Qualitative Report, 19(40), pp. 1–25 (2014). doi: 10.46743/2160-3715/2014.1008.
6. Bazeley, B.: Qualitative Data Analysis with Nvivo, pp 6–15 (2007)
7. Bell, E., Bryman, A., Harley, B.: Business Research Methods. Oxford University Press (2018)
8. Brewer, J.: Ethnography. McGraw-Hill Education (2000)
9. An Introductory Brief.: Qualitative and Quantitative Research Techniques for Humanitarian Needs Assessment. Physical Review, 47, pp. 777–780 (2012)
10. Brown, T.A., Moore, M.T.: Confirmatory Factor Analysis. Handbook of Structural Equation Modeling, pp. 361–379 (2012)
11. Bryman, A.: Integrating Quantitative and Qualitative Research: How is it Done?. Qualitative Research, 6(1), pp. 97–113 (2006). doi: 10.1177/1468794106058877.
12. Burchfield, R.W.: The New Fowler's Modern English Usage. Clarendon Press Oxford (1996)
13. Checkland, P., Holwell, S.: Information, Systems, and Information Systems. John Wiley & Sons Chichester (1998)
14. Chekanov, S.V.: Scientific Data Analysis Using Jython Scripting and Java. Springer Science & Business Media (2010). doi: 10.1007/978-1-84996-287-2\_1.
15. Child, D.: The Essentials of Factor Analysis. A&C Black (2006)
16. Corbin, J.M., Strauss, A.: Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. Qualitative sociology, 13(1), pp. 3–21 (1990). doi: 10.1007/BF00988593.
17. Costa, A.P., de Souza, F., Moreira, A., Neri de Souza, D.: Webqda—Qualitative Data Analysis Software: Usability Assessment. In: 2016 11th Iberian Conference on Information Systems and Technologies, pp. 1–6 (2016). doi: 10.1109/CISTI.2016.7521477.
18. Dalati, S.: Measurement and Measurement Scales. In: Modernizing the Academic Teaching and Research Environment, pp 79–96 (2018)
19. Demuth, C., Mey, G.: Qualitative Methodology in Developmental Psychology. International Encyclopedia of Social and Behavioral Sciences (2015). doi: 10.1016/B978-0-08-097086-8.23156-5.
20. Denzin, N.K.: Triangulation 2.0. Journal of Mixed Methods Research, 6(2), pp. 80–88 (2012). doi: 10.1177/1558689812437186.

21. Dignos, A.: *Research Methods* (2019)
22. Field, A.: *Discovering Statistics Using SPSS* (2009)
23. Flick, U.: *Mixing Methods, Triangulation, and Integrated Research*. *Qualitative Inquiry and Global Crises*, 132(1), pp. 1–79 (2011). doi: 10.4324/9781315421612-7.
24. Fotache, M., Strimbei, C.: *SQL and Data Analysis. Some Implications for Data Analysis and Higher Education*. *Procedia Economics and Finance*, 20, pp. 243–251 (2015). doi: 10.1016/S2212-5671(15)00071-4.
25. Geertz, C.: *The Interpretation of Cultures*. Basic Books (1973)
26. Gill, P., Stewart, K., Treasure, E., Chadwick, B.: *Methods of Data Collection in Qualitative Research: Interviews and Focus Groups*. *British Dental Journal*, 204(6), pp. 291–295 (2008). doi: 10.1038/bdj.2008.192.
27. Glaser, B.G., Strauss, A.L.: *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Mill Valley, CA: Sociology Press (1967)
28. Gogtay, N.J., Deshpande, S.P., Thatte, U.M.: *Principles of Regression Analysis*. *Journal of the Association of Physicians of India*, 65, pp. 48–52 (2017)
29. Grant, D., Michelson, G., Oswick, C., Wailes, N.: *Guest Editorial: Discourse and Organizational Change*. *Journal of Organizational Change Management* (2005). doi: 10.1108/09534810510579814.
30. Graue, C.: *Qualitative Data Analysis*. *International Journal of Sales, Retailing & Marketing*, 4(9), pp. 5–14 (2015)
31. Greenbaum, T.L.: *The Handbook for Focus Group Research*. Sage Publications (1998), doi: 10.4135/9781412986151.
32. Hesse-Biber, S., Dupuis, P., Kinder, T.S.: *Hyperresearch: A Computer Program for the Analysis of Qualitative Data with an Emphasis on Hypothesis Testing and Multimedia Analysis*. *Qualitative Sociology*, 14(4), pp. 289–306 (1991)
33. Yahmady, A., Hilal, H., Al Abri, S.S.: *Using Nvivo for Data Analysis in Qualitative Research*. *International Interdisciplinary Journal of Education*, 2(2), pp. 181–186 (2013)
34. Hodges, B.D., Kuper, A., Reeves, S.: *Discourse Analysis*. *Bmj*, 337, pp. 879 (2008). doi: 10.1136/bmj.a879.
35. Hothorn, T., Everitt, B.S.: *A Handbook of Statistical Analyses Using R*. CRC press (2014). doi: 10.1201/b17081.
36. Hsieh, H.F., Shannon, S.E.: *Three Approaches to Qualitative Content Analysis*. *Qualitative Health Research*, 15(9), pp. 1277–1288 (2005). doi: 10.1177/1049732305276687.
37. Hurmerinta-Peltomäki, L., Nummela, N.: *Mixed Methods in International Business Research: A Value-Added Perspective*. *Management International Review*, 46(4), pp. 439–459, (2006). doi: 10.1007/s11575-006-0100-z.
38. Husamaldin, L., Saeed, N.: *Big Data Analytics Correlation Taxonomy*. *Information*, 11(1), pp. 17 (2020). doi: 10.3390/info11010017.
39. Ibrahim, M.: *The Art of Data Analysis*. *Journal of Allied Health Sciences Pakistan*, 1(1), pp. 98–104 (2015)
40. Kabir, S.M.S.: *Basic Guidelines for Research*. Chittagong: Book Zone Publication (2016)
41. Knight, A.: *Basics of Matlab and Beyond*. Crc Press (2019)
42. Komorowski, M., Marshall, D.C., Saliccioli, J.D., Crutain, Y.: *Exploratory Data Analysis. In Secondary Analysis of Electronic Health Records*. Springer, pp. 185–203 (2016)
43. Kreuger, L., Neuman, W.L.: *Social Work Research Methods: Qualitative and Quantitative Approaches: With Research Navigator*. Pearson/allyn and Bacon (2006)
44. Krippendorff, K.: *Content Analysis: An Introduction to Its Methodology*. Sage Publications (2018)
45. Kumar, A., Teeja, M.S.: *Sentiment Analysis: A Perspective on its Past, Present and Future*. *International Journal of Intelligent Systems and Applications*, 4(10), pp. 1 (2012)
46. Kumar, V., Garg, M.: *Predictive Analytics: A Review of Trends and Techniques*. *Int. J. Comput*, 182, pp. 31–37 (2018)

47. Lacey, A., Luff, D.: *Qualitative Data Analysis*. Trent Focus Sheffield (2001)
48. Madden, A.D.: *A Definition of Information*. Aslib Proceedings: New Information Perspectives. Emerald Group Publishing Limited, pp. 343–349 (2000)
49. Madhulatha, T.S.: *An Overview on Clustering Methods*. Arxiv Preprint Arxiv (2012). doi: 10.48550/arXiv.1205.1117
50. Marshall, C.: *Roschman, g. Designing Qualitative Research*. Sage Publications California (1989)
51. Marshall, C., Rossman, G.B.: *Designing Qualitative Research*. Sage Publications (2014)
52. Mason, W.M., Wolfinger, N.H.: *Cohort Analysis*. International Encyclopedia of the Social & Behavioral Sciences (2001)
53. Mavrikis, M., Geraniou, E.: *Using Qualitative Data Analysis Software to Analyse Students' Computer-mediated Interactions: The Case of Migen and Transana*. International Journal of Social Research Methodology, 14(3), pp. 245–252 (2011)
54. McKinney, W.: *Python for Data Analysis: Data Wrangling with Pandas, Numpy, and Ipython*. O'reilly Media (2012)
55. Merriam, S.B.: *Qualitative Research and Case Study Applications in Education*. Revised and Expanded from *Case Study Research in Education*. Eric (1998)
56. Miles, M.B., Huberman, A.M.: *Qualitative Data Analysis: An Expanded Sourcebook*. Sage (1994)
57. Mills, J., Birks, M.: *Qualitative Methodology: A Practical Guide*. Sage (2014)
58. Mishra, P., Pandey, C., Singh, U., Gupta, A.: *Scales of Measurement and Presentation of Statistical Data*. Annals of Cardiac Anaesthesia, 21(4), pp. 419 (2018)
59. Morgan, D.L., Krueger, R.A.: *The Focus Group Guidebook*. Sage (1998)
60. Musante, K., DeWalt, B.R.: *Participant Observation: A Guide for Fieldworkers*. Rowman Altamira (2010)
61. Nieuwenhuis, J.: *Analysing Qualitative Data*. First Steps in Research, 2, pp. 104–130 (2007)
62. Ong, M.H.A., Puteh, F.: *Quantitative Data Analysis: Choosing Between Spss, Pls, and Amos in Social Science Research*. International Interdisciplinary Journal of Scientific Research, 3(1), pp. 14–25 (2017)
63. Osang, J., Udoimuk, A., Etta, E., Ushie, F., Offiong, N.: *Methods of Gathering Data for Research Purpose and Applications Using Ijser Acceptance Rate of Monthly Paper Publication*. Iosr Journal of Computerengineering Iosrjce, 15(2), pp. 59–65 (2013)
64. Ott, R.L., Longnecker, M.T.: *An Introduction to Statistical Methods and Data Analysis*. Nelson Education (2015)
65. Patton, M.Q.: *Two Decades of Developments in Qualitative Inquiry: A Personal, Experiential Perspective*. Qualitative Social Work, 1(3), pp. 261–283 (2002)
66. Peng, C.J.: *Data Analysis Using SAS*. Sage Publications (2008)
67. Platon, V., Constantinescu, A.: *Monte Carlo Method in Risk Analysis for Investment Projects*. Procedia Economics and Finance, 15(14), pp. 393–400 (2014)
68. Ritchie, J., Spencer, L.: *Qualitative Data Analysis for Applied Policy Research*. The Qualitative Researcher's Companion, 573, pp. 305–29 (2002)
69. Sapsford, R., Abbott, P.: *Ethics, Politics and Research*. R. Sapsford & v. Ju Eds. Data Collection and Data Analysis, pp. 317–340 (1996)
70. Stasko, J., Goörg, C., Liu, Z.: *Jigsaw: Supporting Investigative Analysis Through Interactive Visualization*. Information Visualization, 7(2), pp. 118–132 (2008)
71. Stemler, S.: *Practical Assessment, Research and Evaluation* (2002)
72. Sundaram, K.R., Dwivedi, S.N., Sreenivas, V.: *Medical Statistics: Principles & Methods*. Anshan (2010)
73. Tabachnick, B.G., Fidell, L.S.: *Using multivariate statistics*. Allyn and Bacon. Needham Heights Ma (2001)
74. al, E.R.T.E.: *Data Collection Methods for Evaluation: Document Review* (2018)
75. Thompson, C.B.: *Descriptive Data Analysis*. Air Medical Journal, 28(2), pp. 56–59 (2009)

76. Thorel, M., Krichevsky, M., Lévy-Frébault, V. V.: Numerical taxonomy of mycobactin-dependent mycobacteria, emended description of *Mycobacterium avium*, and description of *Mycobacterium avium* subsp. *avium* subsp. nov., *Mycobacterium avium* subsp. *paratuberculosis* subsp. nov., and *Mycobacterium avium* subsp. *silvaticum* subsp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 40(3), pp. 254–260, (1990).
77. Thorne, S.: Data analysis in qualitative research. *Evidence-Based Nursing*, 3(3), pp. 68–70, (2000).
78. Thrun, M. C.: Approaches to cluster analysis. In: *Projection-Based Clustering Through Self-Organization and Swarm Intelligence*. Springer, pp. 21–31, (2018).
79. Venkatesh, V., Brown, S. A., Bala, H.: Bridging the qualitative–quantitative divide: guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, pp. 21–54, (2013).
80. Watling, R., James, V., Briggs, A.: Qualitative data analysis: using NVivo. In: *Research Methods in Educational Leadership & Management*, pp. 381–396, (2012).
81. Whitehead, D., Annells, M.: Sampling data and data collection in qualitative research. In: *Nursing and Midwifery Research: Methods and Appraisal for Evidence-Based Practice*, pp. 105–121, (2007).
82. Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., Qu, H.: Opinionseer: interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), pp. 1109–1118, (2010).
83. Yilmaz, K.: Comparison of quantitative and qualitative research traditions: epistemological, theoretical, and methodological differences. *European Journal of Education*, 48(2), pp. 311–325, (2013).
84. Yong, A. G., Pearce, S.: A beginner’s guide to factor analysis: focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2), pp. 79–94, (2013).
85. Zainal, Z.: Case study as a research method. *Jurnal Kemanusiaan*, 5(1), (2007).
86. Zikmund, W. G., Babin, B., Carr, J., Griffin, M.: *Business Research Methods*, 7th ed. Thompson Learning, CA, (2003).
87. Zikmund, W. G., Carr, J. C., Griffin, M.: *Business Research Methods*. Cengage Learning, (2013).