INDIVIDUAL ASSIGNMENT 2

TECHNOLOGY PARK MALAYSIA

Application of Data Mining Techniques In The Tourism Sector

CT075-3-2-DTM

DATA MANAGEMENT

APD2F2011CS(DA)

NAME: MD NOWSHAD UL ALAM

TP NUMBER: TP057903

HAND OUT DATE: 4TH DECEMBER 2020

HAND IN DATE: 16th March 2021

WEIGHTAGE: 50%

Contents

1.0 Introduction 3

2.0 The Reason/Motivation for using data mining in the Area: 4

2.1 Forecasting Tourist Expenditures 5

2.2 Analyzing Profiles of Tourists 6

2.3 Forecasting Arrivals of Tourists 7

3.0 The data mining implementation process used in the area. 8

4.0 The data mining techniques used in the area. 12

5.0 New/prospective opportunities produced by the analysis. 18

6.0 Challenges faced in this area. 20

7.0 Future direction(s) of research in this area 24

8.0 Conclusion 27

9.0 References 29

10. APPENDICES 31

1.0 Introduction

Nowadays, everyday thousands of people travel throughout the world for their business , for making vacation during holidays, some makes individual trip as their hobby and for many other reasons. A handsome amount of cash is spent on tickets, facilities, food, transportation, and entertainment. As per World Travel and Tourism Council, tourism industry address around 11% of the overall (GDP) (Werthner and Ricci, 2004). The travel industry is an information-based business where there are two kinds of flow of information. One progression of data is from the suppliers to the purchasers or travelers. This is data about goods or staff that travelers consume like tickets, room of hotel, excitements, etc. The other progression of data which follows an opposite bearing comprises of total data about vacationers to specialist organizations. In this research paper a detailed about application of data mining in tourism sector has been explain. Now to know about its (data mining) application in tourism sector first we need to know about data mining in detailed.

Data mining is a process that helps to find the pattern, correlations and as well as useful data from a large set of data by extracting information of that data. These days it can be measured newly through developed methods and new technology. It uses high artificial intelligence, machine learning it also furthermore uses statics, mathematics, and technologies. The main objective of data mining is to predict the outcome from a given data set and to make a meaningful data by which one can build relationship for example: relationship between customer and company, to reduce risk and many more. When it comes to prediction of data, data mining uses a branch of science which is well known as Statical science. The main goal of using Statistics in data mining is to work with a specific type of problems by using complex algorithms designed. In data mining

recognizing patterns can be done by using a well-structured algorithm which helps to instruct a model for a special group of problems. if we know the main reason of any problem that we are dealing with here we can also identify pattern in the data set to build a system by utilizing machine learning if it can recognize patterns. To make raw data as meaningful information sometimes we can call data mining as a process of analysis when it belongs to specific point. After having a meaningful information, then only this information is collected and stored in data servers like data warehouse or ready to use in data mining algorithms, techniques and helpful in decision making too. It also well known as Knowledge discovery of data or KDD because to check the probable events that could occur in future based on searing patterns here there is some uses of complex mathematical problems or algorithms (Sharma, 2020).

The main goal of the assignment is to do appropriate research on a specific sector and needs to come out with necessary findings such as: reason behind using data mining in this sector, techniques that should we use, benefits or opportunities creates due to this analysis, probable challenges that we may face in future etc. As main topic of research, Application of data mining in the tourism sector has been selected.

2.0 The Reason/Motivation for using data mining in the Area:

Nowadays, huge amount of people make tour throughout the world to spend their vacation, for business purpose or for other reasons. And to make tour one needs tickets, accommodation to stay, food, transportation as well as entertainment. Throughout this process, an enormous amount of money spends by the tourists. In a survey which is done by world travel and tourism council, here it says almost 10.3% of world GDP (gross domestic product) represents by only tourism sector. It nearly contributes US$8.9 trillion in world economic. Besides that, it also creates around 330 million jobs (Economic impact report, 2019). As tourism sector has a huge economic value for both private and public investors/stakeholders. In the tourism sector, obtaining and stocking tourism products can be costly and risky since its tourist could reduce anytime so accurate predictions play an important role in maintaining the tourism sector. This helps promote effective management and great economic growth .

As tourism sector involved with economic sector almost with every country, sometimes few private spot management companies as well as government organizations needs information like correlation between tourism attractions and preferences of tourists about most visited place in order to build required infrastructures for example: suitable accommodation and transportation websites or application to reduce anxiety of tourists so that more people would get interest to visit their place. Sometimes to assist or to make a way easier for tourists in making some decisions like operational, tactical, and strategic, some private and public company need detailed analysis. For instance: making a suitable timetable and good guide to help them in visiting good place as well as in communication. On account of the obligation for this investigation, in tourism sector techniques of statistic were introduced. Though analysis with the help of statistical techniques is helpful, in some case it could suffer from the disadvantages, so to avoid this sort of mistake that various hypothesis about distributions of information or data must be made before any analysis can be managed. But there is no guarantee that the outcome of the data will be valid if these assumptions are damaged. As statistical process has some limitation in this analysis, some researchers has promoted to utilize benefits of machine-learning-based data mining to apply it on tourism sector. In machine-learning, data mining process has three main uses in the tourism industry which are namely forecasting expenditures of tourists, analyzing profiles of tourists, and finally forecasting number of tourists arrivals. We will be discussing reason behind these three techniques in the below (what-when-how, 2021) .

2.1 Forecasting Tourist Expenditures

Palmar, Montano, and Sese (in press) build ANNs to forecast the tourist expenditures. It was built with different architectures in the Balearic Islands. These different monuments' performances, how the work was contrasted by comparing the MAPE, which is the mean absolute percentage error incurred. There the accuracy of forecast of ANNs was different with the architecture. Another one of the studies was conducted to predict shopping expenditures. It was conducted mainly on the tourists who visited Hong Kong, RS and it was published in the Law & AU in 2000. The explanatory variables were nine in numbers and in the dependent variable, the researchers that conducted the whole research devised fifteen of the decisions rules in the four levels of the dependent variable which could then tell the shopping expenditure with an accuracy rate of 94%.

The RS was mainly used for predicting, and the field was the tourist expenditure in dining. The research conducting area was Hong Kong and in 2002 the research was conducted by Au and Law. They used nearly ten of explanatory variables, and eighteens of decision rules which were induced. The research was then 83% correct and the research could report the dining expenditures of visitors (Sheldon, 2020).

2.2 Analyzing Profiles of Tourists

In south Africa, ANN was used to predict the profiles of the visitors of Cape Town. It was used by Bloom in 2002. In the research the survey data which was collected from approximately 694 respondents was used, mainly from the tourists who visited the Cape town in middle of the 2000-2001 that time. The method used for this research is called SOFM, and it was used for diving the tourists into three main groups,

1. People of about 39.8% which fell under the energetic and vibrant class,

2. 34.3% people who were said to be well established and were settled, and

3. 25.9% of people who are under the class of pleasure seekers.

To define the various segments in the market of west Australian tourists who were seniors, the SOFM was also used by Kim, Wei and Rays in the year of 2003. Three factors were used in this research, they are:

• Demographic,

• Travel motivation and

• Concerns.

The people who responded the survey were divided into four groups,

1) People who are active learners,

2) Mainly the relaxed family bodies,

3) People who are very careful in participating,

4) People who are the elementary vacationer only.

The improvement of the Tourism at any range in the give places is dependent many times on the residents who are in favor of it or not, was said by Perez & Nadal in year 2005.

A survey was also conducted in the Balearic Islands, and it was conducted on the 791 residents, the results that came from that research were divided to understand the local residents and they were clustered into four groups,

• Locals who support the developments,

• People who are prudent when it comes to development,

• People who are cautious enough and also ambivalent, and

• Protectionist's people.

The first class contains 11% of supporters, second class consists of 26%, cautious ones are of 24% of the survey, and people who fall under fourth class are of 20% of the survey (Analysts, 2014).

2.3 Forecasting Arrivals of Tourists

From many times, the ANNs have been very popular among researchers and people who conducts surveys, in the field of forecasting visitors and tourist arrivals in any defined locations. ANN is back-propagation based and it has been used to pre-tell the tourist incoming from the US- United States to Durban and South Africa, the research was based on the data available then from 19921998, said Burger, Dohnal, Kathrada and Law in 2001. The ANNs was also used then to predict the arrival numbers from Japanese tourists who were going to Hong Kong, the data were used on the 6 variables that were explanatory, the information that were collected from different variety of resources over the time from 1967 to 1996, said Law & au in 1999. Lowest MAPE was yielded by the ANNs model, and it only hit a level of 10.59% and was executed significantly, in way better than the other statistical models. A study of the same kind was conducted for forecasting tourist arrivals from Taiwan to Hong Kong using the same data from the same variables at the

same time. In that research too, the ANNs hold the best result of them all with an MAPE rate of 2.76%. a study then extended all the previous studies, considering the tourist arrivals to Hong Kong from a variety of six different countries, which were Taiwan, Japan, UK, Korea, Singapore, and USA. The ANNs method again proved that it it the supreme method over all of the statistical methods. The RS method is also have been used for the analysis of travel demand, said Goh & Law in the year of 2003. The research used the data collected between the year 1985 to 2000, and with the visitors who visited Hong Kong from ten countries, the RS model also could predict the rise or fall in the tourist numbers. The accuracy rate was about 87.2%. in the result, it was found that the volume of trade and GDPs were the most important predictor variables. In the recent times, the method of SVM has also been used for telling the tourist arrivals early to Barbados and it has yielded also a lower MAPE if it is compared with other several statistics models, said Pai and Hong in the year of 2005.

3.0 The data mining implementation process used in the area.

Several kinds of industry are having benefits of data mining to increase their revenue or to boost their business quality, including marketing, manufacturing product, chemical etc. In this way, the necessity of conventional data mining method increased. In order to understand easily or with no knowledge of the data mining context, techniques of data mining should be reliable, repeatable by an organization. As a result of this, In 1990 , a cross – industry standard process for data mining was first introduced after observing several workshops and contribution for over 300 organizations (ww.Javatpoint.com, 2011).

To apply data mining techniques first we need to understand the data and its implementation process. There are several types of implementation process which are mentioned below in details (Rungta, 2021).

1. **Business Understanding:**

This is the first phase of data implementation process in which we need to established aim of business and data mining. Here firstly of all, we have to acknowledge business and its customer purpose. We must explain their (customers) want because sometimes they don't know many things before going to trip such as: timing : which season is suitable for vacation or trip.

Second of all, we need to take funds for the latest data mining scheme, assumption and other important elements that needed for our assessment. After that using this objective of the business and present scheme we need to define or set up data mining goals.

Finally, we must make depth and detailed data mining plan and work in order to achieve both business and data mining goals.

2. **Data Understanding:**

After understanding business purpose, we need to understand data. In this phase, to examine if the records are applicable for the data mining, sanity check on data is conducted. In this phase, firstly, records of the tourists should be collected from different kinds of data sources which are available in the organization.

Secondly, the records which we may get from available sources may include multiple databases, flat filter, or data cubes. It may occur some problem like matching and schema integration during data integration process. It very difficult and complex method as records from different kinds of sources hardly match. For example, table A bears an attribute named Tourists_no though other table B contains an attribute named as tourists_id. For this reason, it is quite trick to confirm that if both of these specified entities mentioned to the same value or not. So to get rid of this problem related to mistake in the data integration process, here metadata should be used.

Thirdly, after second steps of understanding data the next step is to search for properties of obtained records. We can use some techniques to explore data which are, to answer questions of data mining that are decided in the first phase (business phase) by using the query, visualization tools and by reporting.

Finally, based on the outcome or result of the query, the data standard should be discovered. We can also understand and fill the missing records if any missing data exist in the dataset.

3. **Data preparation:**

The main objective of this phase is to make data ready for production.

In data implementation process this method takes or consume almost 90% of the project time. The main objective of data preparation process is to select, clean, transform, format and construct (if required) from different sources. So in this process we can clean our data that we collect from different sources, transform raw data as meaningful.

4. **Data Transformation:**

To make all the records that exist in dataset meaningful, data transformation operations change the data. There are some methods of data transformation which are mentioned below:

- **Smoothing:** It can be helpful while it comes to replace, remove, delete noisy records which exists in data set, smoothing can be done.
- **Aggregation:** In data, aggregation operations are applied. For instance: Number of tourists visited in few places here we can apply this operation to calculate number of tourists that visit in whole country from that sample.
- **Generalization:** In generalization process, Through the concept of hierarchies, low-level attributes is replaced by higher-level attributes. For example: City is replaced by the country.
- **Normalization:** It is a process which performed when the attributes information is scaled up o downsized. For instance: In the range from -2.0 to 2.0 post normalization, data should fall.
- **Attribute construction**: The attributes are constructed and present in the given set of attributes which are helpful for data mining.

After the outcome of this process, data set can be used in modeling as it considers as final data set.

5. **Modelling**

This is the phase in which mathematical models are used to define or to recognize data patterns.

Here, acceptable modeling process should be selected for the developed dataset based on objectives of the business. We should also build a proper scheme to test the standard and logic of the model. After testing model, we need to run the model on the appropriate dataset.

After collecting all data by the process of data implementation now we can apply data mining techniques.

6. **Evolution**

In this stage, in contrast to the business objectives, patterns identified are evaluated.

1. In contrast to the business goals, Results created by the data mining model ought to be evaluated.
2. Here gaining the knowledge about business understanding is known as an iterative process. . In fact, during the time of understanding , it could be possible to raise new requirements in due to data mining.
3. To move the model in the next phase that is deployment, a decision is taken from go or no-go decision.

4. **Deployment**

In the deployment stage, the records which we got by data mining discoveries are transport to everyday business operations.

- for non-technical stakeholders, the information or data found during data mining process ought to be made straightforward.

- A deployment plan with detailed information, for delivery, support, and checking of data mining discoveries is made.

- A last task report is made with exercises learned as well as with key experiences at the time of project. This assists with improving the organization's business strategy.

By following the above process of implementation techniques in data mining we can collect information, clean and transform it as meaningful information and make ready for working. No after implementation mining techniques can be apply in the data set.

4.0 The data mining techniques used in the area.

Data mining techniques is the process by which we can predict data from the past information.

The data mining techniques that are used in this area are mentioned below:

1. Classification:

To restore significant and applicable information about data and metadata, classification techniques can be applied. It is also known as problem identifying process by which we can know to which a new observation belongs to. For example: we can classify tourists based on their sex if we gather the information like number of tourists that visited in a place in every year.

Types of classification techniques that are used in this area are:

1. **Logistic Regression**

Logistic regression is generally a supervised classification technique in data mining. In problem of classification, the output variable which is known as Y can only accept discrete values for the given input which is known as X. It is also known as regression model as it develops a model of regression in order to predict the outcome of the probability which a given data entry belongs to

category numbered as 1. It models the data through the use of sigmoid function and assumed that follows of data occur by linear function.

$$g(z) = \frac{1}{1+e^{-z}}$$

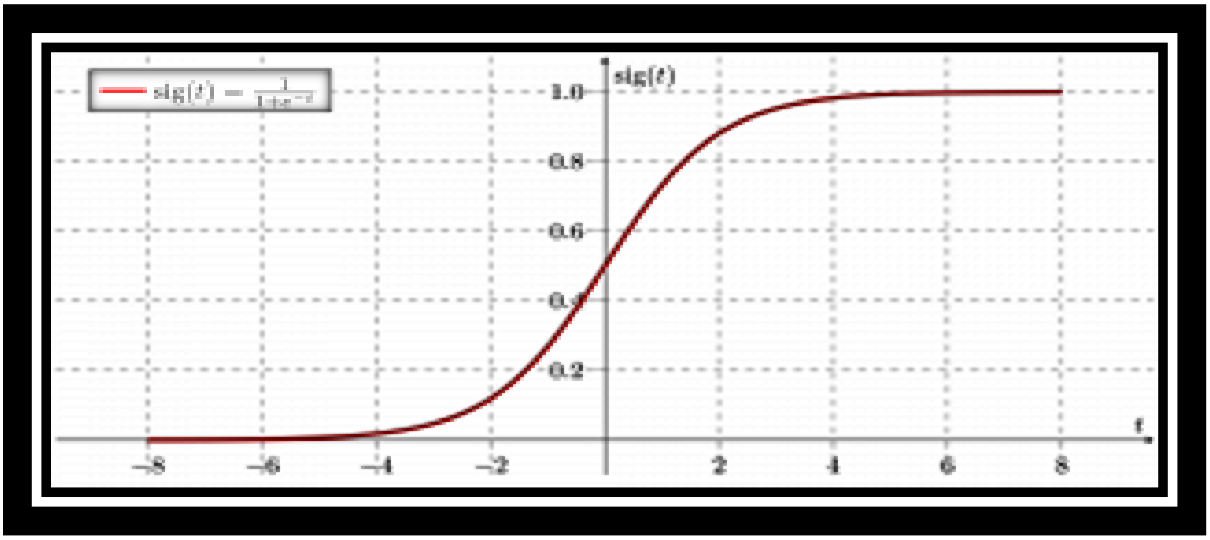


Figure – 1(visualization of logistic regression)

When decision threshold is brought into the function only than logistic regression consider as classification technique.

By following this process, organization can observe flow of tourism. For example: in which place tourist go to make trip, which season they (traveler) has more interest in making trip etc.

2. Decision Tree

To develop model of classification in data mining, decision tree mining is used. In the shape of structure like tree it develops classification model. It is categorized as a supervised learning technique. We can use this mining technique in both categorical and numerical data. Generally, to develop classification and regression models we can see the use of decision tree. This method is used to build data models which will predict the class label or values which needed in decision-making process/ technique. We can visualize the outcome of a dataset that is decision step by step by using this process which makes the work easy to understand (Software Testing Help, 2021).

With the help of decision tree process, organization which works for tourism sector can visualize their statistic or plan from the beginning so that they will have minimum risk in future and they can also guide their staff according to their plan.

3. Support vector Machines (SVM)

Support vector Machine(SVM) is known as one popular supervised learning algorithms in machine learning that is used to in regression problems and classification. The main objective of this process is to develop the best line which can disconnect n-dimensional space into classes so that in future we can easily new data point in correct category. In this technique, hyperplane is known as best decision boundary. Extreme points or vector that are chosen by SVM, helps in building hyperplane which (extreme points) is also known as support vector. Here in the below, there are two different types of data in the diagram which are classified by using hyperplane (JavaTpoint, 2011).

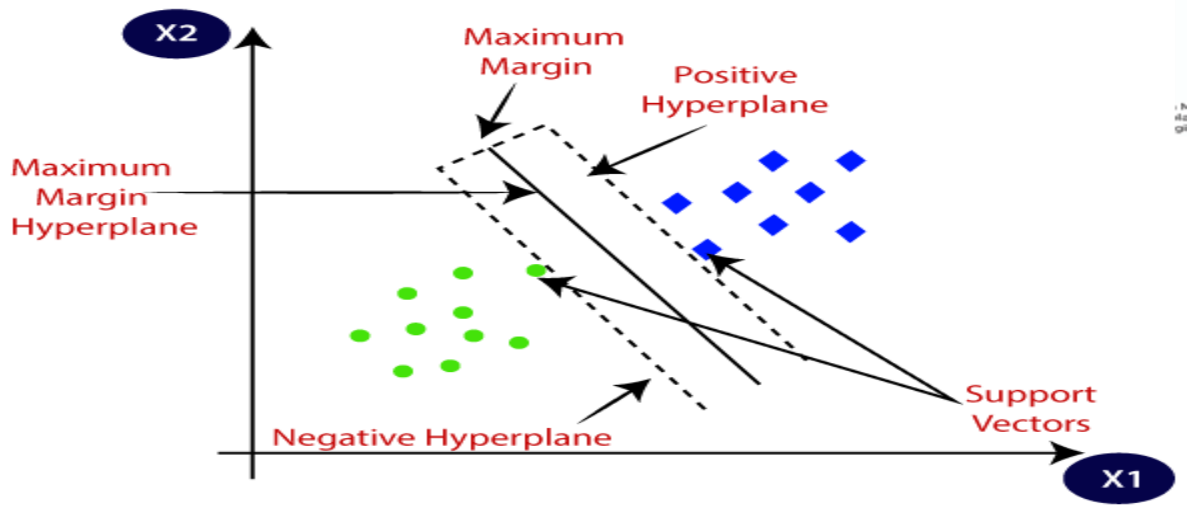


Figure – 2(visualization of Support Vector Machines)

For example: In some countries to get more facilities or to avoid spending money some of the foreigners shows them as local guy as sometimes they have same body structure like local people,

so if we want a model that can exactly detect whether he/she is local or foreigners and such a model that can be built by SVM algorithm. Here first we will train our model with lots of language, accent, body structure etc. which are needed to detect the country so that it can learn about different kinds of people around world. So as extreme point creates boundary between people around the world and it will choose extreme cases. Based on the support vectors, SVM will detect which country they belong to.

2. Clustering:

Clustering is technique in machine learning which puts together unlabeled dataset in groups. You could define it method of collection the information focuses into various groups, comprising of comparable information focuses. The items with the potential likenesses stay in a group that has less or no similitudes with another group. This is done by locating similar patterns in the unlabeled data, things like the size color and shape. It does not require any supervision as this is an unsupervised learning machine. Unsupervised learning does not require supervision as it find the hidden stuff by itself (Tawde, 2019).

When done applying this technique, each cluster is given a cluster id, this way system could simplify big and difficult data. This technique is mostly used in statistical data analysis.

Companies like Netflix use this technique in the recommendation to give recommendation based on your previous search.

There are 5 types of clustering method which can also be divided into hard clustering and soft clustering.

1. Partitioning clustering

Partitioning clustering is one of the 5 clustering method which divides dataset into non-progressive groups, it divides it to k group sets and use the K to describe the number of already defined groups. One good example of this could be the k means clustering algorithm (edureka, 2014).

An organization wants to open travel agency across a country. To open this agency, they may need some information like:

1. They need to analyse the area from where more tourists come to make trip in this country frequently.
2. They also need to know how many hotels that are always used by traveller in his area.
3. Then they need to figure out the price and standard of all hotel and from where they will receive their customer etc.

After solving all these challenges along with a lot of analysis and mathematics, they can now open their agency. In this way clustering works by providing meaningful and easy process of sorting out in real life problems.

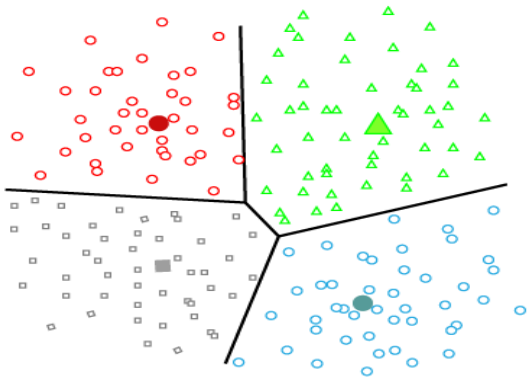


Figure – 3(visualization of Partitioning clustering)

2. Density based clustering method.

There is also Density based clustering method which connects very dense areas to clusters. And discretionarily shaped distribution is made when the dense region is connected. The algorithm does this by identifying different clusters in the data and then connect very dense areas to cluster.

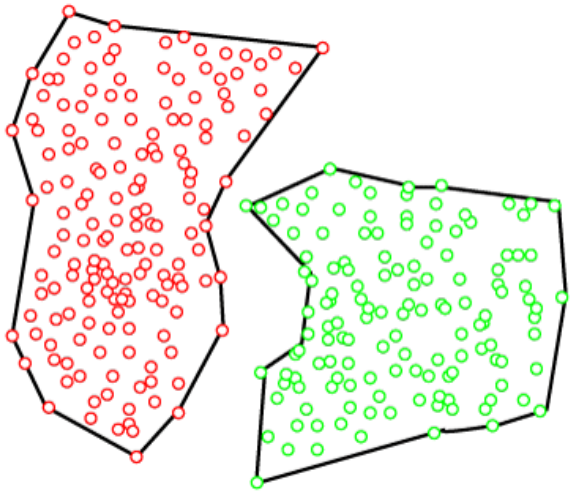


Figure – 4(visualization of Density based clustering method)

3. Distribution model-based clustering

Distribution model-based clustering is also another method and in this method it divided the data using probability, it does the grouping by making assumptions of distribution mostly the gaussian distribution.

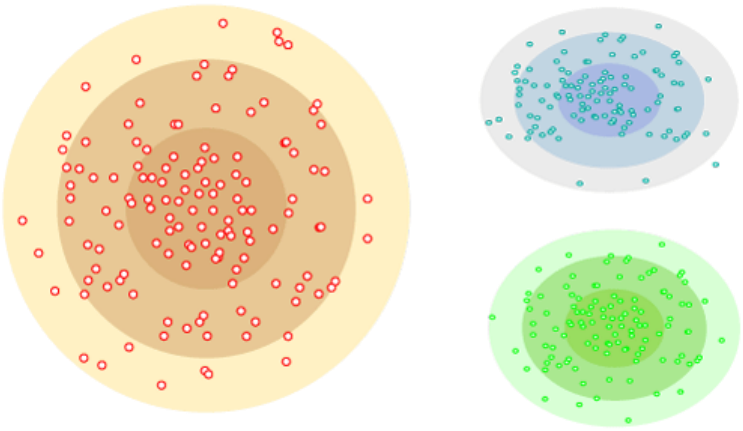


Figure – 5(visualization of Distribution model-based clustering)

4. hierarchical clustering

There is also hierarchical clustering and with this we can replace partitioned clustering because it is an alternative. the dataset is partitioned into groups to make a tree-like construction, which is likewise called a dendrogram. The perceptions or quite a few groups can be chosen by cutting the tree at the right level. The most well-known example of this is agglomerative hierarchical algorithm.

5. Fuzzy clustering

Fuzzy clustering is a soft method which one object could belong to more than one cluster. Each dataset has a bunch of participation coefficients, which rely upon the level of enrollment to be in a group.

By these methods we can separate different kinds of visited place by tourism which will help us to detect which place would be profitable to invest and if we find any place where we have our staff, but tourism has no attraction to visit or hardly any tourist travel there we will just move all the staff to other place where we have more demand so that in that method we can save our time as well as money. We can also study customer (tourist) to segment the customers based on their preference and choice.

5.0 New/prospective opportunities produced by the analysis.

As we know data mining is the procedure of calculating a database of big amount of pre-existing database. By this analyzation new information can be found using which intelligence can be gained. The customers mainly created the tourism industry, and as time passes the industry receives a lot of data from the customers, analytics plays the main role for analyzing this large quantity of data and to result in trends and meaningful important data patterns. This analysis can help to understand the relationship between tourists and hotels, or the data tourists provide.

Through this investigation the needs of customers can be understood, and the problems related to tourism can be solved quickly. From delving into data mining, there are some partial usefulness too, that is the mining can be limited by some features but if various strategies are used then it is possible to go further to find information. If the data mining is focused on the existing tourist behavior-based data, transactions, demographics, geological location, it can help to extract the customer needs and market products to customers more efficiently. In the tourism, data mining helps to refine the tourism strategy, makes communication more relevant and efficient, and at the bottom line the results become more positive. Because of the obligation for this investigation, in tourism sector techniques of statistic were introduced (Hausold, 2018).

Opportunities that will create by this analysis are:

- By conducting data mining in the tourism, a relevant and sound travel experience can be given to tourists. Beside that through data mining techniques travel agency can understand their customer "their needs and habits". And from that traveller can also select their place. Thus, it helps to maintain band's loyalty too.
- The straightaway displayed advantage of it is, the tourists praise the service they get and staying incentive is increased, also they are more likely to report in for the same service while future.
- The business profit is increased. For example: If a company can give good opportunity to its customer, they(customer) will refer to other customer in this way business will expand. Side by side data about the development of individual tourist, their purchasing preference, their desires. This information, if effectively treated, addresses a huge benefit for organizations in the travel industry area which presently have the chance to characterize and advance their systems to build their deals.
- Old data can be analyzed defined tools and can be use in workplace in own favor. So that it will be easy to predict the outcome of future of that sector.
- Is beneficial for the improvement for the business. Example of strategy: A best example of strategy usage for data mining can be taken is the prize optimization. Where the service provider offers the hotel bookings, or guide fees according to the analysis result of the opponent service provider's offered prices and the seasonal data and behavioral structure of the opponent can also be found.

- Travel companies can develop a better way of making trip guide process if they know interest of the customer side by side, they can remove destination (omitting from their plan) which has less demand in case of travelling. Also, it could help tourism with providing efficient travel plan/route to the tourism therefore they could travel more place in sort time and efficiently (s, 2017).
- Data mining can be helpful in making decision method and can improve how the tourism sector runs. Users in this sector can make informed decisions based on analysis and number-driven data. They can every identify their regular customer at every stage in the travel planning process. Side by side company can also improve and maintain their efficiency and quality in the service (revfine.com, 2021).
- Company which are involved in tourism industry can also use data mining to compile and analysis records about their main competitor to get a clear knowledge and understand about what they(other company) are offering to their customer and can follow this in this company as well.
- To find all types of hidden detail information we usually use data mining and adding data mining may helps an organization to improve their website. In the same way, it (data mining) also gives information which may use the technology of data mining. Thus, it helps an institution to increase website optimization.
- The online action of the traveller gives full scale information that gives us a superior image of their conduct. Hence, examines based on studies or meetings with specialists, lacking the important objectivity to find accurate solutions on the genuine situation of the travel industry currently belong in the past.

Here other than that we also have so many opportunities including fraud detection among tourists ,about health of traveler etc. (Rajkumar, 2014).

6.0 Challenges faced in this area.

Data mining is quite possibly the most generally utilized techniques to get information from various sources and coordinate them for better uses. A lot of challenges can be seed during the

time of actual implementing of data mining. With fast advancement in the field of data mining, organizations are required to keep up to date with every one of the new turns of events.

Complex algorithms structure the reason for data mining as they take into consideration information division to distinguish different patterns and examples, recognize varieties, and anticipate the probabilities of different occasions occurring. The raw data may come in both simple and computerized design and is intrinsically founded on the wellspring of the information. Organizations need to monitor the most recent data mining patterns and stay refreshed to do well in the business and defeat challenging competition.

These days data mining and information disclosure are developing a significant innovation for business and scientists in numerous areas. data Mining is forming into set up and confided in discipline, numerous yet forthcoming difficulties must be solved. The challenges that may face in this area are listed below:

1. Security and social Challenges:

As decision-making procedures are usually done by data collection-sharing, so it requires extensive security. Private data about people and sensitive information which are gather for customers profiles and to understand the client's behavior patterns. Illegal admittance to data and the classified idea of data turning into important issue for which Several kinds of problems may arrive with data mining in general apply to the tourism sector from them, one of these is the difficulty of sharing information throughout the sources. Although there is boundless information all over the cities, its frequently spread out across many organizations which they do not combine data clusters.

On the top of that, organizations repeatedly do not have a standard way of gathering information which stop it from being used to its full potential. Luckily, the UN's WTO is as of now chipping away at making a standard measurable system to manage this problem.

So, here security problem may raise in doing data mining in tourism sector as through data mining we collect numerous information of the tourism.

2. User interface:

The information found is discovered by utilizing data mining tools is helpful just only if it is fascinating or more all reasonable by the client. From good visualization interpretation of information, mining results can be facilitated, and betters comprehend their requirements. To acquire good visualization much research is carried out for enormous informational collections that show and control mined knowledge.

1. Mining based on Level of Abstraction: Data Mining method should be collaborative since it gives permission to its users to focus on pattern finding, introducing, and optimizing solicitations for data mining depend on returned output.
2. Integration of Background Knowledge: Organization may use previous data of the tourism to express discovered patterns to coordinates the exploration method and to communicate found patterns.

3. Mining Methodology Challenges:

These types of difficulties are connected to data mining draws near and their impediments. Problem that occurs due to mining approaches are given below:

1. Mining approaches in adaptability
2. Variety of information accessible
3. Capacity of the domain
4. Problem in maintaining and controlling of noise in records/ data etc.

Various methodologies may execute differently dependent on the data consideration for example: few algorithms need noise-free data while sometimes most data sets bear exceptions, meaningless

or incomplete data that lead to complication in the analysis method and in fees cases compromise the precision of the outcomes.

As tourism sector is bound with enormous information about people, places, hotels etc. it maybe one of the major problems that may create during research.

4. Challenges with complex data:

Real-world data can be heterogeneous, and it could be multimedia data including pictures, audio and video, complex information, worldly data ,spatial information, time series, natural languages test and so on. It is hard to deal with these different sorts of information. New tool and procedures are creating to extract relevant information.

1. Complex data type: The database which are consider as complex data can incorporate complex data components, object with graphical data, spatial information, and transient data. Mining every one of these sorts of data is not pragmatic (practical) to be done in one device.
2. Mining from Varied Sources: As we know different kinds of data are gathered from different sources on network. As we stored from different sources, sources of the data may be of various types depending on how they are stored for instance structured, semi-structured or unstructured.

5. Challenges in performance

The data mining system performance relies upon the algorithms and techniques are utilizing. The algorithms and techniques planned are not sufficient lead to influence the performance of the data mining process.

1. **Efficiency and Scalability of the algorithms.**
   To extract records from enormous amounts of data in the data set, the data mining algorithm should be proficient and scalable. So, if an organization cannot manage to collect proficient and scalable data about traveller around the world, about hotels, places etc. they may face critical situation.

2. **Improvement of mining Algorithms.**

   Factors like huge size of the database, the whole data flow and the trouble of data mining approaches motivate the creation of parallel and disseminated data mining algorithms (geeksforgeeks, 2020).

Here above are some challenges that organization which involve with tourism sector may face during the research. During research , we may face many other challenges which are maybe ignorable.

7.0 Future direction(s) of research in this area

From the previous discussion, we come to know that different kinds of techniques of data mining have been utilized successfully in recent time in order to analyze data of tourism sector. When it comes to tourism, we cannot find much research or effort in order to develop tools or software for data mining by researchers. As we know from the above discussion there are some techniques including networks, case-based reasoning and decision trees which have been used a little in this sector. This is because due to the shortage of available prepackaged tools or software that uses this process. In future we expect, uses of these process/ techniques in tourism data mining. The other methodology that we may see later-on is that of ensemble data mining in which more than one process or strategy is utilized for investigating the data. At the point when ANN is utilized for forecasting tourist expenditures in a specific country, the presentation may not be excellent because of overtraining. In any cases, if clustering is done before the Ann forecast, the examination can be more significant as the data is gathered and preprocessed to an extent. There might be freedom to join ANN with RS or SVM in a similar process with clustering.

Side by side, in tourism industry, organizations which have been delayed in adapting the process of information mining are presently catching up with the others. Extricating significant data through the method of data mining is generally used to settle on basic business choices. In the coming decade, we can expect information mining to become as universal as a portion of the more pervasive advancements utilized today. In tourism sector, few development or future trends that may happened because of data mining are :

1. Media Data Mining

This is probably the most recent strategy which is making up for lost time as a result of the developing capacity to catch helpful data accurately. It includes the extraction of information from various types of sources of multimedia like sound, text, hypertext, video, pictures, and so forth and the information is changed over into a numerical representation in various format. This strategy can be utilized in clustering and classification, performing similitude checks, and furthermore to recognize associations. With the help of this method, we can collect information accurately and we can include more data of the tourism by which business will expand.

2. Ubiquitous Data Mining

This strategy includes the mining of data from cell phones to get data about people. Despite having a few difficulties in this kind like intricacy, protection, cost, and so forth this strategy has a ton of freedoms to be tremendous in different industry particularly in contemplating human-computer process. By this process we can also gather information of the tourists so that we can know something about their habits like what they want to order form restaurant, how they pass their leisure time.

3. Distributed Data Mining

This is a type of data mining process which includes the mining of huge amount of data stored in in various organization areas or at various company. So that we can know about what other organization are offering to their traveler and how they are increasing their popularity among their customers. Again, if one can know these sorts of information, they can follow this and can increase their revenue in business. Profoundly modern algorithms are utilized to get data from various areas and give appropriate bits of knowledge and reports dependent on them.

4. Time Series and Sequence Data Mining

The essential utilization of this sort of data mining is investigation of repetitive and occasional trends. This method is additionally useful in analyzing even random occasions which happen outside the ordinary arrangement of occasions.

So, in future, this method will help us to detect in which season or months tourist want to make trip throughout the world (Kraus, 2018).

Here , the process of getting information and techniques which may use to get accurate information about tourism in future that are mentioned above. Other than these techniques and methodology, there are many other opportunities, techniques and features that may add in future.

8.0 Conclusion

As we know, recently tourism industry plays an important role in terms of economy all over the world. Because of this many researchers are now doing research on it. Here in this research paper, only few main data mining application areas have been discussed so far. In future, more areas on this (tourism) sector likely to come for instance tourism recommendation analysis, tourism promotion response analysis and so on. For developing these activities data mining plays an important role.

Here after all above discussion, this research paper gives a deeper knowledge about the application of data mining in the tourism sector . Here we can find many reasons for implementing data mining in tourism sector which are likely forecasting tourist expenditures, analyzing profile of tourists, and forecasting arrival of tourists and so many. To know about the application of data mining in this sector, firstly we need to know its implementation process for that here we can see a good explanation on that. After that, the research paper also explains about data mining techniques that have been used so far in this industry which are including - Logistic Regression, Decision Tree, Support vector Machines (SVM), Clustering. And by using this technique we will have some opportunities including – opportunities in giving a relevant and sound travel experience to the customer, expanding of business, increasing profit in business, saving money and time , and so on which has been given in detailed in the above. As we know everything in this world has some side effect also and data mining is not different from them so after opportunities, next we can see something about facing challenges has been discussed which may face during research including Security and social Challenges, Mining Methodology Challenges, Challenges with complex data etc. Lastly , after challenges here we see about future direction of the research paper in which we can see about some techniques which may use in future and their opportunities has been discussed.

Therefore, data mining had play an important role in a tourism sector by identifying or predicting factors that affect the tourism sector and many more which can help the government as well as private organization in making suitable decisions to boost the country's economy, GDP and also it will help to create many job opportunities for the unemployed which can make economically stable one country a head from other.

After completing this research, it can be seen that a little amount of knowledge that related to data mining and the application of data mining in tourism sector has been gained. From the research it

is clear that in future, application of data mining in this sector is very important in order to make business of any organization stable that related to this sector and to stay strong along with other competitors.

9.0 References

Analysts, D., 2014. *Destination Analysts.* [Online]
Available at: https://www.destinationanalysts.com/research/visitor-profiles/
[Accessed Monday March 2021].

Economic impact report, 2019. *Economic impact report.* [Online]
Available at: https://wttc.org/Research/Economic-Impact#:~:text=In%202019%2C%20Travel%20%26%20Tourism's%20direct,10%20jobs%20around%20the%20world
[Accessed Monday March 2021].

edureka, 2014. *edureka.* [Online]
Available at: https://www.edureka.co/blog/k-means-clustering/
[Accessed Friday March 2021].

geeksforgeeks, 2020. *Challenge in data mining.* [Online]
Available at: https://www.geeksforgeeks.org/challenges-of-data-mining/
[Accessed Friday March 2021].

Hausold, A., 2018. *Tourism review.* [Online]
Available at: https://www.tourism-review.com/big-data-apps-are-used-more-often-news10713
[Accessed Monday february 2021].

JavaTpoint, 2011. *JavaTpoint.* [Online]
Available at: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm
[Accessed Monday March 2021].

Kraus, N., 2018. *Tourism review.* [Online]
Available at: https://www.tourism-review.com/big-data-technology-used-in-the-tourism-sector-news10293
[Accessed Monday March 2021].

Rajkumar, 2014. *Big data.* [Online]
Available at: https://bigdata-madesimple.com/14-useful-applications-of-data-mining/
[Accessed Monday March 2020].

revfine.com, 2021. *revfine.* [Online]
Available at: https://www.revfine.com/big-data-travel-industry/
[Accessed Monday February 2020].

Rungta, 2021. *Guru99.* [Online]
Available at: https://www.guru99.com/data-mining-tutorial.html#11
[Accessed Monday February 2021].

s, B., 2017. *data mining in tourism.* [Online]
Available at: http://misclassblog.com/big-data-2/data-mining-in-tourism/
[Accessed Monday March 2021].

Sharma, R., 2020. *upGrad.* [Online]
Available at: https://www.upgrad.com/blog/data-mining-techniques/#What_is_Data_Mining
[Accessed Monday March 2021].

Sheldon, 2020. *Sage Journal.* [Online]
Available at: https://journals.sagepub.com/doi/abs/10.1177/00472875303200103?journalCode=jtrb
[Accessed Monday March 2021].

Software Testing Help, 2021. *Software Testing Help.* [Online]
Available at: https://www.softwaretestinghelp.com/decision-tree-algorithm-examples-data-mining/
[Accessed Monday March 2021].

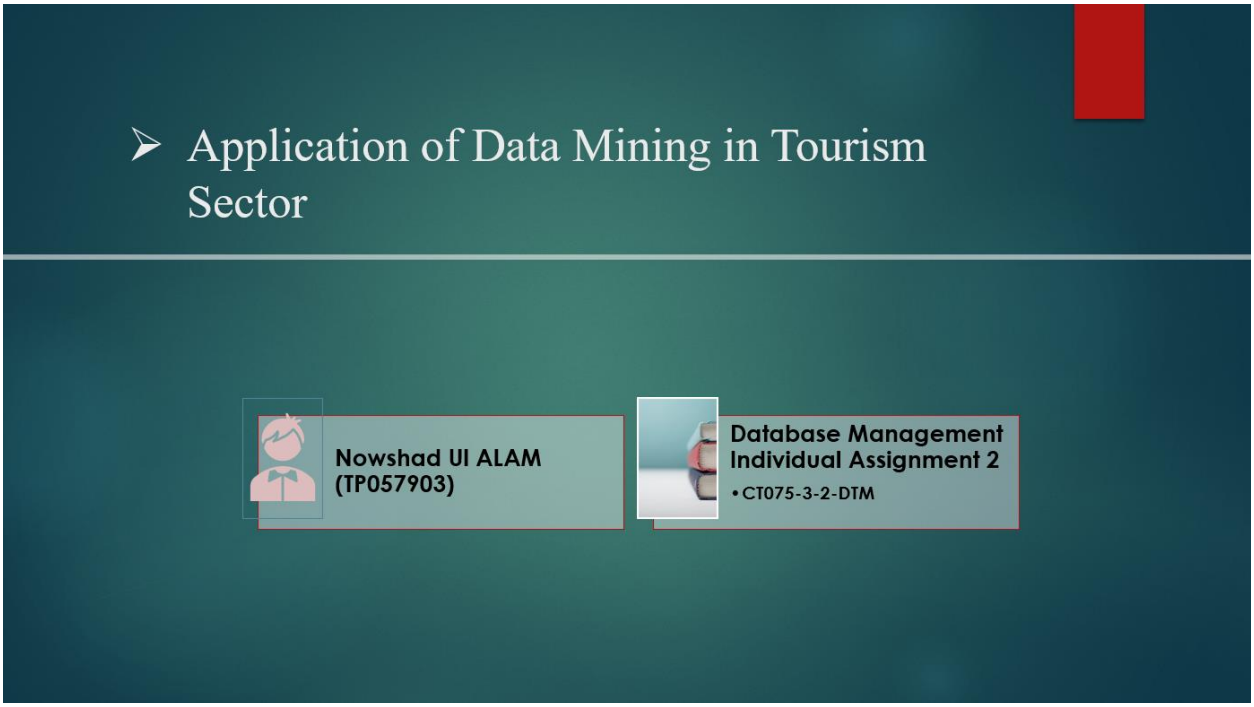
Tawde, S., 2019. *EDUCBA.* [Online]
Available at: https://www.educba.com/what-is-clustering-in-data-mining/
[Accessed Friday March 2021].

what-when-how, 2021. *Data Mining in Tourism (information science).* [Online]
Available at: http://what-when-how.com/information-science-and-technology/data-mining-in-tourism-information-science/#:~:text=The%20three%20main%20uses%20of,forecasting%20number%20of%20tourist%20arrivals.
[Accessed Monday February 2021].

Ricci, F. (2004). *E-Commerce and Tourism | December 2004 | Communications of the ACM*. [online] Acm.org. Available at: https://cacm.acm.org/magazines/2004/12/6347-e-commerce-and-tourism/fulltext [Accessed 14 Mar. 2021].

ww.Javatpoint.com, 2011. *Javatpoint.* [Online]
Available at: https://www.javatpoint.com/implementation-process-of-data-mining#:~:text=Data%20Mining%20Implementation%20Process&text=Data%20mining%20is%20described%20as,)%2C%20Machine%20learning%20and%20statistics.
[Accessed Monday March 2021].

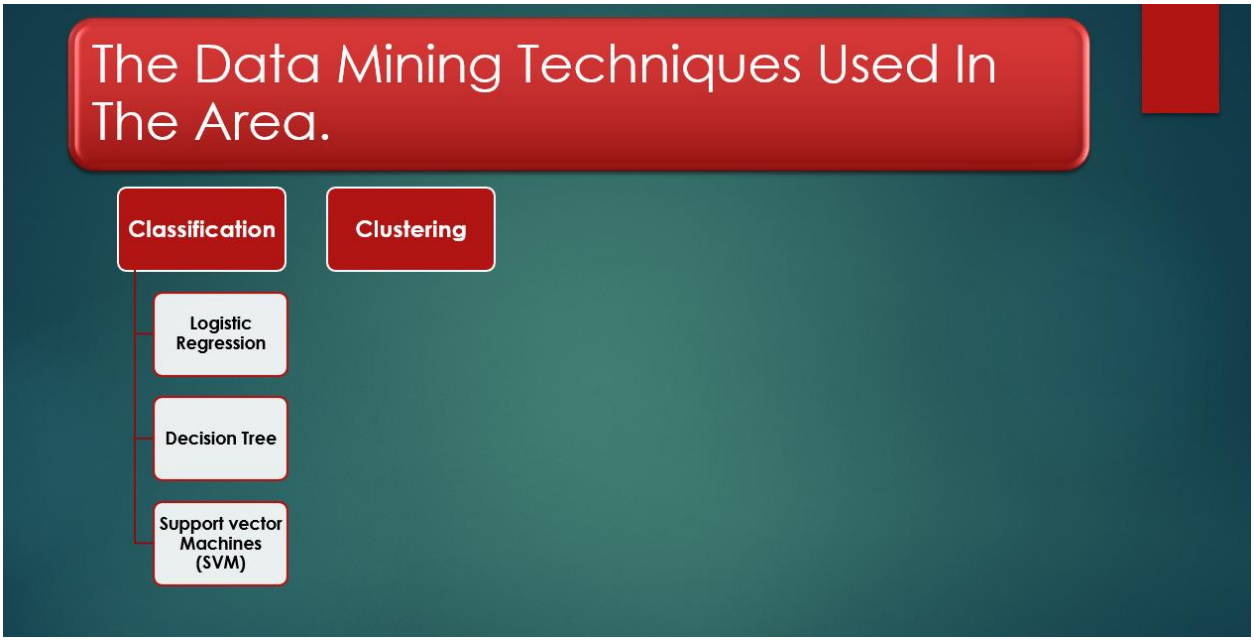
10. APPENDICES



Slide 1

Slide 2

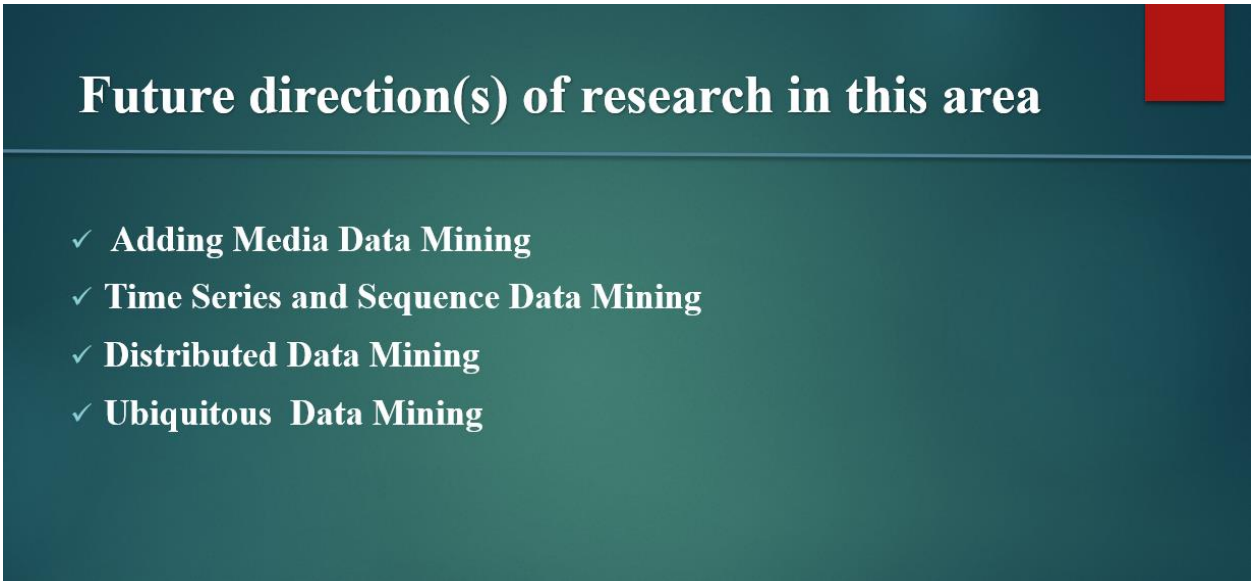
Slide 3



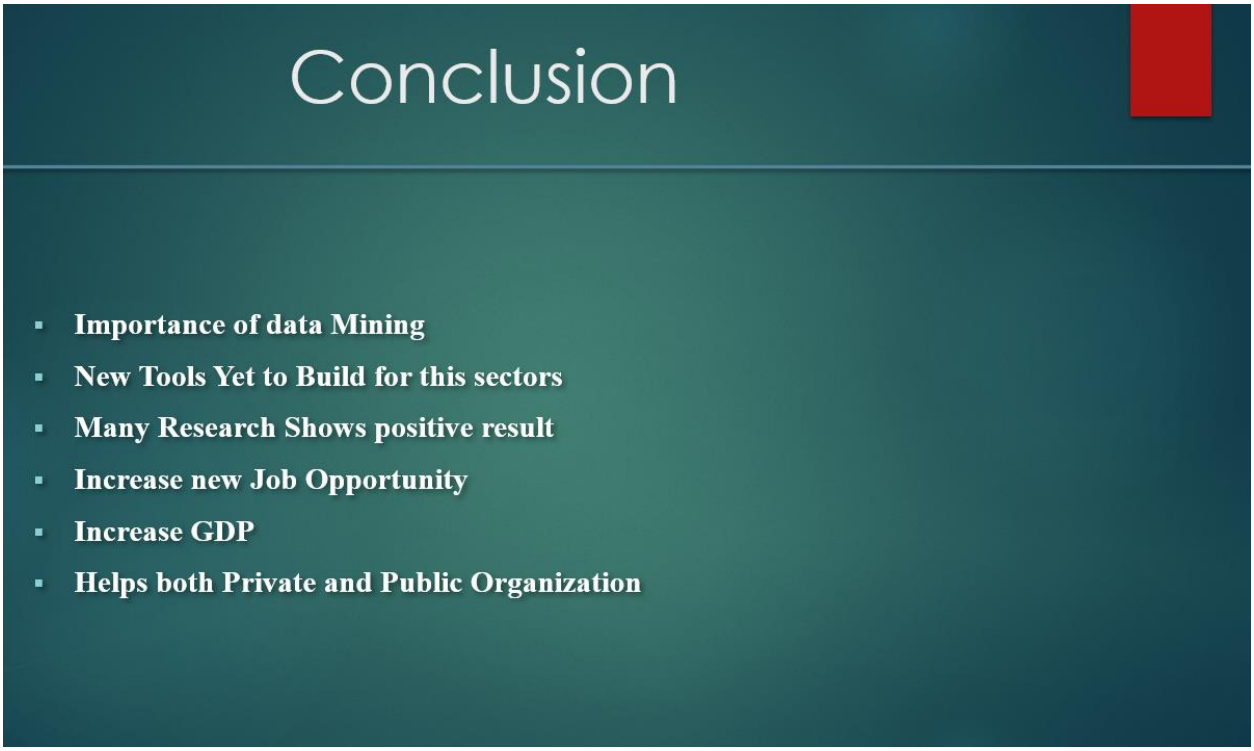
Slide 4

Slide 5

Slide 6



Slide 7



Slide 8