Analysis and Prediction: Hidden Patterns of Foreign Visits

Weerarathna R. A. D. A. S	Fernando W. P. U	Rangadhari M. I. T	Jasandan K.	Vithana N.
Sri Lanka Institute of Information Technology Sri Lanka	Sri Lanka Institute of Information Technology Sri Lanka	Sri Lanka Institute of Information Technology Sri Lanka	Sri Lanka Institute of Information Technology Sri Lanka	Sri Lanka Institute of Information Technology Sri Lanka

ABSTRACT

Tourism was formally recognized as a fast-growing industry and thereafter, subsequently placed emphasis on the development of tourism in Sri Lanka. At present tourism is one of the main economical revenue resource which makes the Sri Lankan economy stronger. The key focus of tourism in Sri Lanka is to do the attraction of huge number of tourists throughout the year. In the meanwhile, tourist information such as nationality, country, age, visiting seasons, details of travel agents have been counted for this research successfully. However, earlier there were no any certain advantages taken upon the information collected. Since the "**Predictor**" system is delighted to meet best performance achievement by analysing historical data regarding arrivals of tourists possible to use for the data mining algorithms. Data mining is a largely automated methodology that has been used to find hidden patterns. This has different types of algorithms which will be benefited at the right time predicting future arrivals and turn up to improve the standards of tourism.

General Terms

Decision Tree algorithm, Clustering algorithm, Time series algorithm, ARIMA algorithm

Keywords

Data mining, hotel , support , systems , decision, tourism, ARIMA , algorithm

1. INTRODUCTION

Date Mining consists of finding interesting trends or patterns in large datasets to guide decisions about future activities. Data mining tools should be able to identify patterns in data using minimal user input. The patterns provide useful and unexpected insight to data analyst, which can be further, investigated subsequently using Decision Support tools. However, manual data analysis has not achieved to target the methods for efficient computer-based analysis. To fulfil this need, data mining was born. Decision support systems are defined as interactive computer-based systems intended to help decision makers utilize data and models in order to identify problems, solve problems and make decisions [1].

In the modern hotel industry, decisions making acts a very foremost part which will be benefited to grow the scale and complexity. Decision making must be predicted efficiently to maintain the high performance. Conversely any anomaly within these decisions making can quickly degrade the business profit efficiently. At present, it involves high risk in predicting the decision and offers precisely.

This research is based on data mining techniques. Data mining has attracted great deal of attention in the information industry and in society in recent years, due to the wide availability of huge amount of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for application ranging from hotel industry analysis.

Data mining concepts and techniques can be used for uncovering interesting data patterns. There are so many techniques that are currently in use to improve the hidden pattern of foreigner visits. Management cannot identify hidden patterns of foreign visits easily. Therefore, decisions making is not more effective. As a solution for this, the project team has designed a framework that can be used to analysis datasets and predict the hotel activities. Identify hidden patterns are very useful to management to plan the future year as much as profitable. However, many of hotel datasets have not been analysed to gain any efficient advantages. "**Predictor**" system has been presented for analyses the historical data regarding arrivals of tourists which has been used by the data mining algorithms. This is the best solution for overcoming problems faced by management of the hotel. Applying data mining with tourism sector may help decision makers to know tourists closely.

Analyse the data with data mining to predict the highest reservation seasons, total number of guests, most visited nationalities and the age groups

With historical data, the project team has predicted the highest reservations that has been made by visitors. By considering the collected data can be given promotions to the people who visiting in Sri Lanka. Thereafter the project team can be given an idea to the hotel management how to manage the sustainable development based upon the analysed results.

Analyse the best algorithm to be used to mining data sets

As the conducted research, a set of steps for some successful experiments need to be developed. At this stage, project team selected the appropriate Data Mining techniques and applied to some certain extent order and developed a methodology/algorithm to gain knowledge from the selected data sets. Data Mining is a broad term applied to several kinds of exploratory analysis. There are many different Data Mining techniques and algorithms such as Time Series algorithm, Decision tree algorithm, Clustering algorithm.

The used dataset is from single hotel based in Colombo area and the predicted results will be based on the sample data set provided. Historical data regarding nationality, country, age, travel agent details, arrival and departure dates of 3 years were available on the data set. Thus, predictions were based on

these captured areas. The accuracy of predicted data mainly depends on the data set.

The project team build a standalone Application for hotel management to make decisions and make plans for future much easy and efficiently. "**Predictor**" is a system that automates Business Intelligence (BI) process. Data mining is commonly termed for a set of techniques that help to analyse and recognize significant facts, relationships, trends, patterns, exceptions and anomalies that might otherwise go unnoticed that exists in large amounts of data.

2. BACKGROUND

There are many researches done in the field of data mining.

In 2015, N. Rathee and S. Choudhary gave an introduction of Data Mining System and development of different data mining systems. They choose which platform was optimum for the development of Data Mining Systems. They had used the juncture of artificial intelligence, statistics, database systems, machine learning and business intelligence. According to them focus on the open source technology may carry out more benefits [2].

"Data Mining for Security Purpose & Its Solitude Suggestions" is a data mining systems that allow mining for safety applications. This technology facilitates the idea of privacy and then talked about the developments particularly those on privacy preserving on data mining. In order to achieve goals of every organization and business, need to fully exploit the data by extracting all the useful information and put effect suitable policies from it. [3]. Data mining process concern large-volume, complicated, expanding data sets with collective, autonomous sources. In data mining and analysis, there were some popular open source tools [4]

In 2012 introduced a framework "Business Intelligence and analytics: From big data to big impact" which classifies the evolution, applications, and turn up research in the areas of Business Intelligence and analytics which includes business-centric practices and methodologies that can be applied to various high-impact applications such as, security, e-government, medical, market intelligence and e-commerce [5].

"Applying Data Mining to Analyze Travel arrangement in Searching Travel Target Choices" was determined traveler's interest extracted from search behavior when the traveler searches for tourism destination. This technology facilitates remote interaction and social communication with people. When the aggregated data about the tourists is conferred in the appropriate way, analyzed by the proper algorithm, and put into the right hands, it could be adapted into worthwhile way of knowledge for making significant decisions by tourism service providers to boost revenue and income. Data mining can be a very favorable mechanism for evaluate tourism-related data. [6]

S. Muthuselvan,Dr. K. Soma Sundaram gave an introduction about the sequential pattern mining and also briefly expressed the major categories of the sequential pattern mining. This paper carries out an organized survey of the sequential pattern mining algorithms. The information produced is very convenient for business determination making and very useful in understanding market trends [7]

In 2015 J. A. Patel, has developed a research paper to identify data mining techniques and also tools required for its implementation. There are many algorithms recommended that crack to address the manner of data mining. In the

participation of today's business world has caused the data input to boost at a tremendous rate thus making data mining a very important part of industries. An importance of varied companies has been attracted more and more due to the implementation of continuous Data mining processes [8]

The paper provides a comprehensive review of hotel management system. The results of this study undertaken relationships in a hotel environment. Work in the area of Hotel Management involves ensuring that all operations, including fitting, food and drink and other hotel services run smoothly. With growing competition in the industry it's important for a hotel to understand the needs of their guest [9]

Finding of this research is a stepping stone in identifying seasonal patterns of foreign visiting. By the analysed data of previous years and with a prediction for the future years will be letting guests to know all the related quality services from the hotel. The frame work also can be used to discover any information about the tourists visits. By the way there were some frame work using for the prediction purposes. Among of them, newly made up **"Predictor"** system have it out with especially for the focus of identifying hidden patterns of foreign visitors

3. METHODOLOGY

Data mining denotes a rather difficult and specific field. A complicated and fixed approach is necessary for the use of data mining in order to help organizations use the data mining. For this project, Prototypes are expected and there will be very critical risk assessment when going through the project because the team does not have much experience within the domain. The spiral model emphasizes risk analysis. Remove all potential risks through careful analysis and, if necessary, by constructing a prototype

Before data is being mined several steps have to follow

- Data integration: First collected and integrated all the data from different sources

- Data selection: Then selected the data that can be used for data mining

- Data cleaning: Most of the time, data gathered are not clean and may contain errors, missing values, noisy or inconsistent data. So, they have to remove. This process was done manually

- Data transform: This step is done for get more efficient results and to gain results in understandable manner. Mainly normalization was done for the data

- Data mining: In this step the algorithms were used for mine the database. Data set was analyzed using Naïve Bayes, Linear Regression, Decision Tree, and Frequent Pattern Growth and ARIMA algorithms

- Pattern Evaluation and Knowledge Presentation: First some of the results that are not related were removed fine tune other results well to understand by the users of the system

Using ARIMA algorithm can forecast number of customers and the guest patterns. The algorithm was coded manually. In here Time Series model was use as the data mining model.

Analyses the previous years and identify most visited guests, guest ages, nationalities and travel argent details for hotel management decisions. View guest patterns which guest prefers to used most. This was created by using Decision Tree

model and the Clustering model. Basically, this gives an idea about past guest patterns in chart format. Predict Seasonal promotion guest patterns in order to be prepared for upcoming seasons. Identifying those patterns can analysis the number of

guest patterns go high which are in End of the year and the beginning of the year. The results are getting through a mining model in Microsoft Business Intelligence using DMX queries.
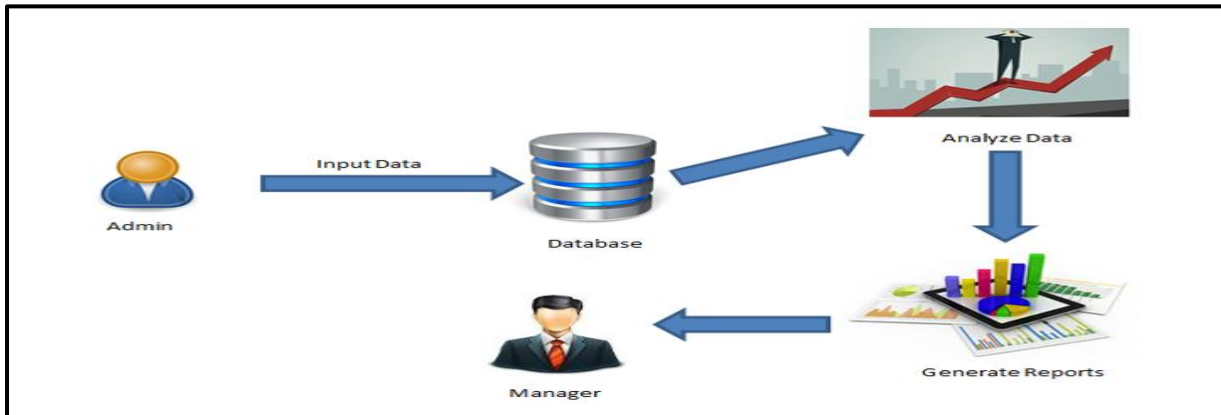

Figure 1: "Predictor" High Level Diagram

4. RESULTS

First the project team have transformed data in required format by removing noise and any other inconsistencies. All the results
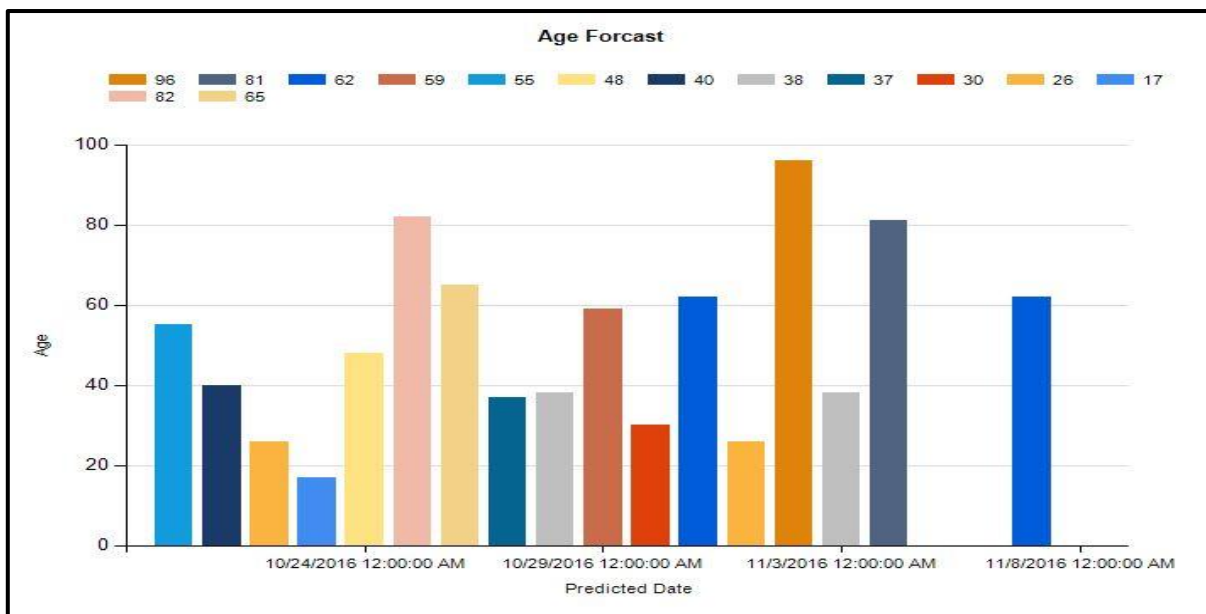

Figure 2: Age Group Forecasting Results

Identifying those patterns can analysis the number of guest patterns go high which are in End of the year and the beginning of the year. The results are getting through a mining model in Microsoft Business Intelligence using DMX queries.

Reports has been generated by using most popular guest patterns apart from the "**Predictor**" system which would be an essential requirement for the hotel management activities and its associate within the track of revenues for the decision making as well.

from these models were refined as well. Time series forecasting is the use of a model to predict future values based on previously observed values.

Figure 2 shows the predicting age groups for the next upcoming dates. Hotel management can get a clear idea about their guest and can get ready to manage the hotel environment accordingly.
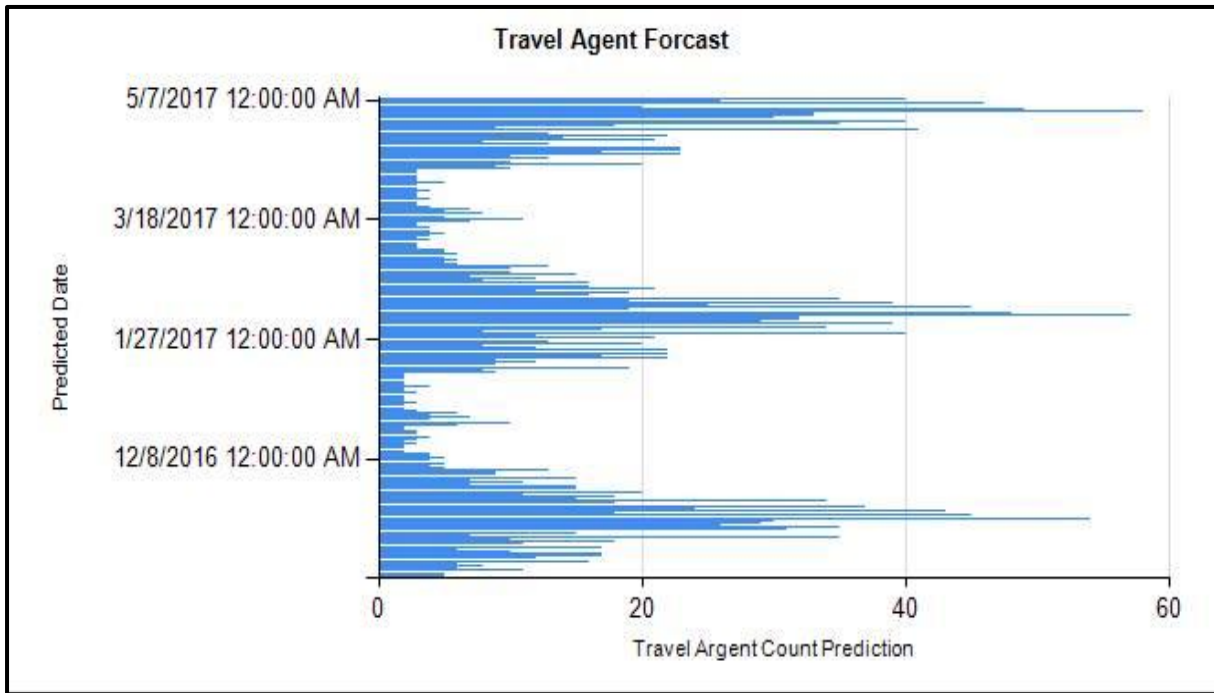
Figure 3 : Travel Agent Forecasting Results

Figure 3 shows the predicting Travel Agents for the next year. This helps to give promotions to the identified Travel Agents, who sent the more guests.

Providing attention for guests was enhanced the guest interactions for the hotel.

Also, Guests can arrange their transportation much more smoothly with the knowledge and support of the travel agents.

Figure 4 shows a sample code that used for analysis of age groups.

```
SELECT
  t.[Nationality],
  t.[Age],
  Cluster()
From
  [Myclust_aug20]
PREDICTION JOIN
  OPENQUERY([Predictor],
    'SELECT
      [Nationality],
      [Age]
    FROM
      [dbo].[myclust]
    ') AS t
ON
  [Myclust_aug20].[Nationality] = t.[Nationality] AND
  [Myclust_aug20].[Age] = t.[Age]
where t.[Nationality] = @nation
```

Figure 4 : Sample DMX query of analysis Age Group by Nationality

Figure 5 gives an analysis of past years comparison results. Through this management can get an idea about their up and downs in hotel environment
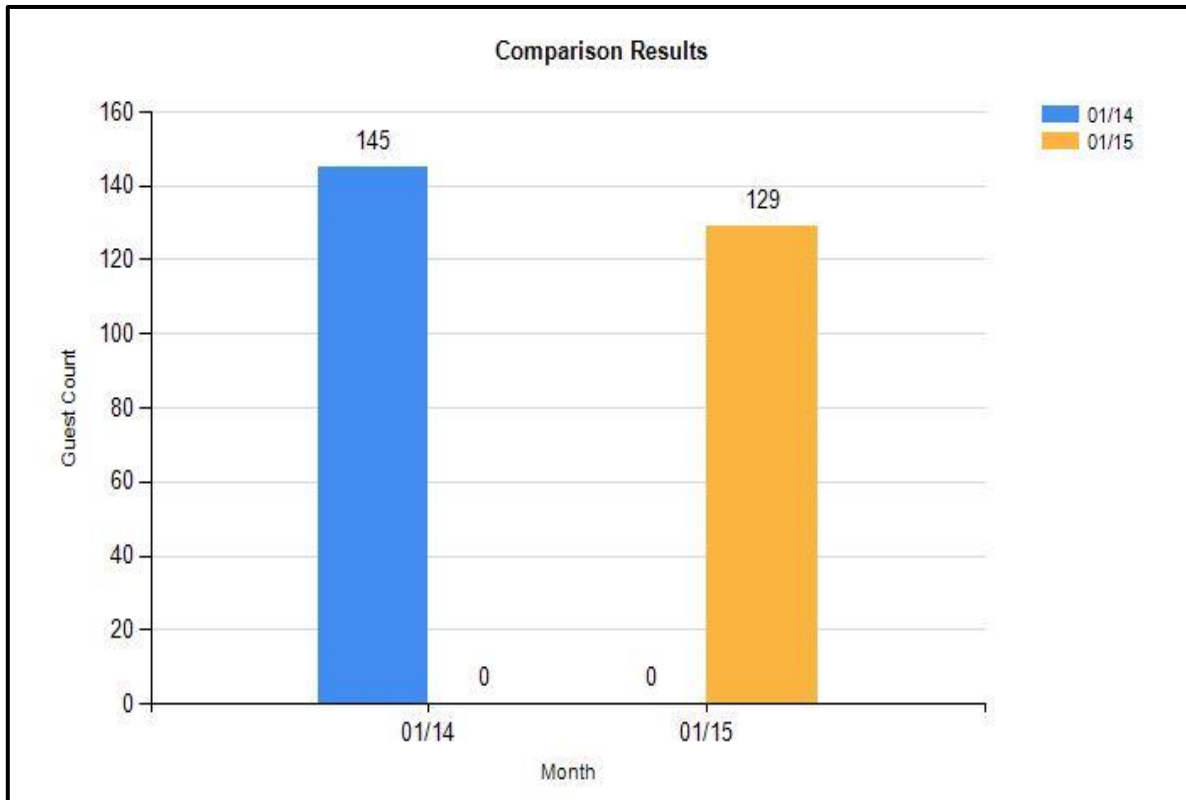
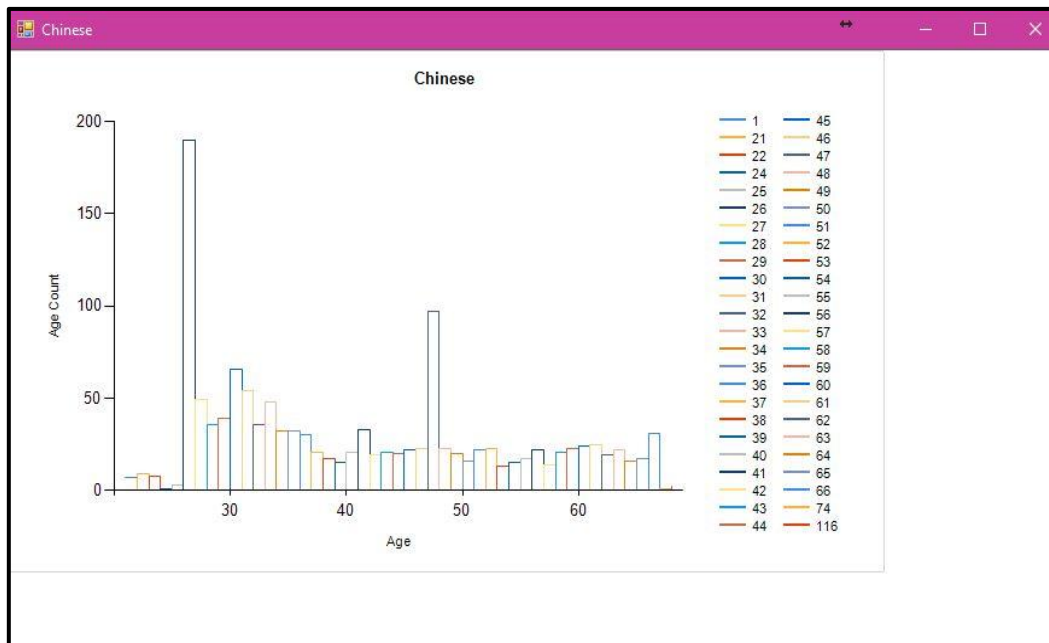**Figure 5: Past Years Guest Count Comparison Results**

Figure 6: Age Group Analysis Results

From Figure 6 gives an idea about the interesting age group patterns of customers for past three years.

This identifies interesting nationality patterns of customers for selected age groups.

In the course of this information may provide an idea about total foreign visits with filtered details

Clustering Algorithm was used to perform the results in age group analysis.

5. DISCUSSIONS

Deploying an automated solution is not an easy task in tourism. Nowadays it is important to deploy data mining into tourism to increase profit margin. The research has been accomplished by developing the PC based desktop application, Hybrid mobile applications and Web based application. "**Predictor**" data mining system will be a vast advantage to hotel management systems as it increases accuracy, efficiency and effectiveness in hotel activities.

Key issue faced during the research is some of data was not accurate, data cleansing and filtering. The process ended up time consuming and was mostly done on a trial and error basis.

Processing a large number of record set for all 3 years from 2013 to 2015 from excel sheet at once was challenging to be performed on a local computer.

Findings of the research signify the importance of further research of the dataset. Low availability of documented information and unformatted data were key challenge encountered during the research.

6. CONCLUSION

This system is knowledge based predicting system for any dataset in any size. This system was tested on real time data. A leading famous hotel database has been used as the case study. The project team built quality product which any person can understand and operate the analysis. Most of the data mining software's provide short predictable data only. This product provides reasonable and explainable report to the users who uses this product.

Findings of this data mining research is a stepping stone in knowledge discovery foreign visiting patterns. To increase the hotel revenue "**Predictor**" data mining system has developed and which will be utilized of providing efficiency for best achievement of the hotel goal. This information can be organized and integrated by analyzing supportive information of guests. The historical data were used for the data analysis based on the algorithms results such as most visited countries, age groups, details of travel agents and seasons etc.

By considering more closely to the hidden pattern of foreign visits could be identified the future seasonal arrivals which will be a major factor of increasing the hotel acquirements. Then be tapped to guide relevant decisions. However, identifying important variables and relationships of data mining research could be brought more economical benefits of the hotel industry.

The project team also intended to do the modifications with the data mining deployment methodology. "**Predictor**" data mining application would be a great step into hotel management systems Island-wide and worldwide. This research can be developed successfully to meet the efficient decision making for hotel industry in the future.

7. REFERENCES

[1] R. P. Chakraborty, "Integration of Data Mining Systems using Sequence Process," International Research Journal of Engineering and Technology (IRJET) , vol. 02, no. 01, pp. 204-211, Apr. 2015.

[2] N. Rathee and S. Choudhary, "Study of different Data Mining system & Platform," International Research Journal of Engineering and Technology, vol. II, no. 1, Apr. 2015.

[3] S. Khan, D. A. Sharma, A. S. Zamani, and A. Akhtar, "Data Mining for Security Purpose & Its Solitude Suggestions," INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, vol. 1, no. 7, Aug. 2012.

[4] L. Wang and G. Wang, "Data Mining Applications in Big Data," Computer Engineering and Applications, vol. 4, no. 3, Sep. 2015.

[5] H. Chen, R. H. L. Chiang, and V. C. Storey, "BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT," MIS Quarterly, vol.36, no. 4, Dec. 2012.

[6] P. Juwattanasamran, S. Supattranuwong, and S. Sinthupinyo, "Applying Data Mining to Analyze Travel Pattern in Searching Travel Destination Choices," The International Journal Of Engineering And Science (IJES), vol. 2, no. 4, pp. 38–44, Apr. 2013.

[7] S. Muthuselvan and D. K. S. Sundaram, "A Survey of Sequence Patterns in Data Mining Techniques," International Journal of Applied Engineering Research, vol. 10, no. 1, pp. 1807–1815, 2015

[8] J. A. Patel, "Classification Algorithms and Comparison in Data Mining," International Journal of Innovations & Advancement in Computer Science, vol. 4, no. Special, May 2015.

[9] B. M. Noone, S. E. Kimes, and L. M. Renaghan, "Integrating customer relationship management and revenue management: A hotel perspective," Journal of Revenue and Pricing Management, vol. 2, no. 1, pp. 7–21, Apr. 2003.