International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Business Intelligence using Data Mining Techniques for Salon Industry

Padma R , Josephene Sheeba S , Thirumurthi Raja A

Abstract: ***Data Mining has many application areas. One such application area is in Salon Industry. This paper presents the aims to harnessing the power of information to support marketing and management decisions, in Salon Industry. A group of salons data was collected, analyzed and its impact on business growth and decisions. This case study used a collage of the existing datamining techniques to aid the Salon industry to augment the business expansion, tactical decision making, and strategic marketing activity using the mined knowledge to understand customers, products and effective use of resources. This work gives an idea of the vast opportunity for growth in the beauty sector using data mining, to impact the managers and investors.***

Keywords: ***Data Mining, Predict, Salons Service Sector, Marketing Decisions***

I. INTRODUCTION

The beauty industry has had a tremendous increase in the last decade. Beauty products – 60%, salons – 35%, beauty treatments – 5%. New players look at this potential market. The Beauty Industry In India Is Currently nailed At USD 950 Million Is forecasted To Reach At USD 2.68 Billion By The Year 2020 [21]. Due to the enormous opportunities in this sector, the growth trends and revenue generation patterns needs to be analyzed by inventors. The data mining algorithms are used to analyze these trends among salons and customers, for predicting the revenue. The data mining techniques help investors, managers to forecast revenue, focus on expansion of salons, venture in new geographical potential areas.

In this study data was collected from a group of salons in Chennai city. It includes 141 salons.

Analysis was made on the following:

- Revenue generated from each salon – month wise, season wise
- Prediction of revenue from each Salon month wise.
- Identification of potential customers Eg: Top 10, Top 100…
- Revenue generated from customer in a particular salon

In this study data was collected from a group of salons in Chennai city. It includes 141 salons.

Analysis was made on the following:

- Revenue generated from each salon – month wise, season wise
- Prediction of revenue from each Salon month wise
- Identification of potential customers Eg: Top 10, Top 100…
- Revenue generated from customer in a particular salon
- Prediction of revenue from customer in a particular salon.
- Geographic location of Salons
- Clustering of Salons according to Revenue
- Clustering of Customers according to Revenue, Age

Data mining techniques allow enterprises to predict future trends by diving through large data sets, to identify patterns, unveil relationships in data through data analysis. [22].

II. LITERATURE REVIEW

A review of the relevant literature from different sources, focusing on the following objectives:

- To understand different perspectives to analyze the concepts of Data Mining in Service Sector.
- Potential benefits of application of Data mining in service sector.

Data mining in beauty industry aids in better decision making. The paper on Customer relationship management in hairdressing industry [13], which highlights use of K Means in effective clustering and RFM model. The paper on Relation between Data Mining and Business in the Four-Dimensional CRM Model [12], gives an insight of various DM techniques used in effective CRM. The Coffee shops analysis in Taiwan [20], use association rules and fuzzy c means to cluster customers. The RFM model used to identify the value of customers and possible churn of customers. Hot Spot analysis was done on crime spots [14] which gives a visual insight on geographical location. Though many such studies are conducted on different service sectors, in India the application of these strategies are still not much explored. The reasons may be the cost factor, expertise or the investors unfamiliarity of these data mining uses in the service industry.

Manuscript published on November 30, 2019.
* Correspondence Author
Padma R*, Department of Computer Science, School of Computing Sciences, Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India . Email:padma.scs@velsunivac.in
S Josephene Sheeba*, Department of Computer Science, School of Computing Sciences, Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India. Email: sheebasj@yahoo.co.in
Dr. A.Thirumurthi Raja, Department of Computer Science, School of Computing Sciences, Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India . Email: atraja.scs@velsunivac.in

Since this current analysis is on Beauty salon, the analysis explores the economic implementation of data mining techniques in a group of salons in Chennai city. The previous research papers on service industry growth seems to have used the following algorithms effectively: Self-organizing maps (SOM), K-means, ARIMA Ensemble, VARCLUS, DT, NN, Logistic regression and Bagging.

III. BUSINESS OBJECTIVE

This research analyses the various business intelligence needed for Service sector business improvement, expansion and even start-ups, through data mining, focusing on:

A. Client classification

Client classification approach is finding groups of customers with similar characteristics or transaction behaviours. The customers can be potential customers or most profitable customers.

B. Client churning & Identifying Potential Customers

Data Mining can be used to find the behavior of churned customers by analysing the customers previous transactions – recent visit, frequency, revenue. This will aid in holding the present active customers and also improving or altering services to retain customers.

C.Analysis of the hotspots

Hotspots are potential geographical areas of growth of present or new salon Identifying hotspots for the business, which enables entrepreneurs in their new venture or in expansion of the business.

D.Predictions

Predictive analysis are used to predict revenue from each Salon and revenue from client. These predictions help to make decisions that generate most sales.

E.Optimal utilization of resources

Idle time or free time of Staff is analyzed, for optimal utilization of resources. Idle time is inevitable, but this can be minimized by rescheduling or making some contingency plans.

IV. METHODOLOGY

The CRISP - Cross-Industry Standard Process for Data Mining methodology was adopted in this case study.

A. Business Understanding

In this phase, the business perspective and important factors that could affect the output of the project was determined.
Goals:

- Predict the Revenue from Salons and Customers
- Retain present Customers
- Effective utilization of Employees
- Expansion of Salons geographically
- Better Customer Service

To achieve these goals:

- The clients need to be classified according to frequency, revenue, recent visit of the client, etc.
- Client retention is evaluated by analyzing and predicting client churning.
- The revenue across all salons is charted, which in turn identifies potential growth or identify hotspots, to minister growth.
- Service Efficiency can be achieved by predicting the peak times and the kind of service a client will demand and the frequency.
- Calculation of idle time of the resources and Utilizing the idle time, by training them to improve their expertise of the employees or re-scheduling.

B.Data understanding

After understanding the business requirements, the next phase was acquiring the data i.e., *collection of data* from reports and database or data load of 141 salons daily transactions for year 2018. The salons dataset consists of 94651 samples for Salons data in a metro city. The Customer dataset consists of customers transactions for each salon, one salon's customer data is analyzed.

Table- I: Summary of datasets

Description	Salons Dataset	Customers Dataset	Staff Dataset
Samples Count	94651	7980	12
Attributes	7	10	6

C.Data preparation

The refining of data, significant attributes and their relationships are identified using feature importance and correlation. Incomplete, missing or incorrect values are acquired. Wrapper Method – Backward elimination and RFE (recursive feature elimination) and Embedded Method – Lasso model, were used for feature selection.

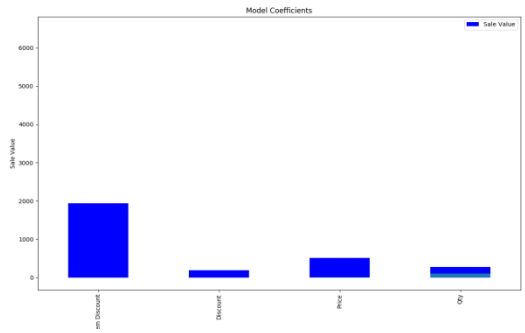

Fig. 1. Attribute selection. (*Model Coefficients*)

Familiarization with the data and identification of data quality and cleaning and transformation of data for modelling.

D. Modeling

The prepared data is analyzed to find the optimal model i.e., the predictive or descriptive model. Modeling data workflow are as follows:
In the prediction of revenue from all salons following algorithms are used on the salon dataset

- Logistic Regression
- LinearDiscriminantAnalysis
- DecisionTreeClassifier -CART
- GaussianNB
- KNeighborsClassifier –KNN
- SVM

E.Evaluation

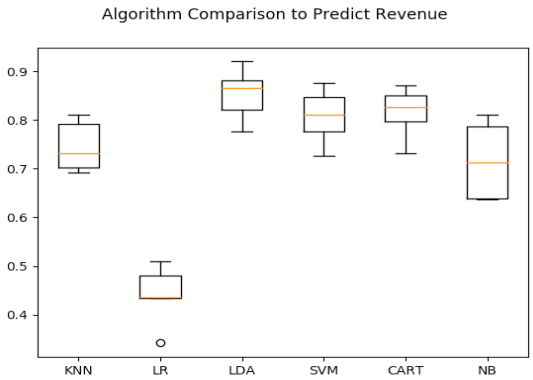

Fig. 2. Comparison of the Algorithms (Revenue Prediction)

from the above chart LDA was chosen as the predictor.

Table- II: RMSE of the Algorithms

Algorithm	RMSE
KNN	267.8
LR	484.5
LDA	104.5
SVM	645.4
CART	232.6
NB	411.5

Root Mean Square Error (RMSE) measures how much error there is between Predicted and Actual observed sales. The lowest RMSE is chosen.

Identification of potential customers the RFM concept was used. This was analyzed for each store.

R – Recency - Days since last customer visit

F – Frequency - Frequency of customer visit

M – Monetary / Revenue - Sales

Table- III: RFM for the dataset

Recency (R)	Frequency (F)	Monetary / Revenue (M)	Score
visited the salon over 12months ago	visited the salon once	spent in the salon < Rs.500	1
visited the salon over 9months ago	visited the salon 2-5 times	spent in the salon Rs.500 – Rs.2000	2
visited the salon over 6months ago	visited the salon 6 - 8 times	spent in the salon Rs.2000 – Rs.5000	3
visited the salon over 3months ago	visited the salon 8 times or more	spent in the salon > Rs.5000	4

The RFM score was calculated and the top 10, top 100 customers were identified.

```
          recency  frequency  monetary
Guest
MERIA BEJOY       3          9     10162
ANU               3          9      8060
STAFIE            1          8      5859
NASSIA UMMER      4          7      5558
BASIL             3          5      5541
                  recency  frequency  monetary  ...  f_quartile m_quartile RFM_Score
Guest                                           ...
MERIA BEJOY             3          9     10162  ...           1          1       111
ANU                     3          9      8060  ...           1          1       111
STAFIE                  1          8      5859  ...           1          1       111
NASSIA UMMER            4          7      5558  ...           1          1       211
BASIL                   3          5      5541  ...           1          1       111
GRACE PHILIP            8          3      5428  ...           2          1       221
SHEHNAZ SHAHAN          4          6      5204  ...           1          1       211
SARASWATI               4          8      4756  ...           1          1       211
NAVYA                   8          6      4591  ...           1          1       211
LESHA                  18          6      4531  ...           1          1       411
ANGALI                 19          4      4484  ...           2          1       421
KIRAN                   1          6      4460  ...           1          1       111
NISHA PAUL             21          2      4248  ...           3          1       431

[13 rows x 7 columns]
```

Fig. 3. Potential Customers using RFM (*First 13 potential customers*)

Geographic location of Salons

The Geographic location of Salons are shown in google map, marked as per the revenue generated by each salon.

This helps in identifying the hotspots were new salons can be open.

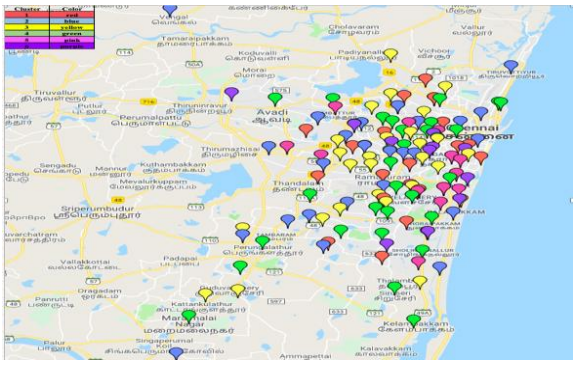

Fig. 4. Geographical clustering by K Means (Salons Cluster in Chennai)

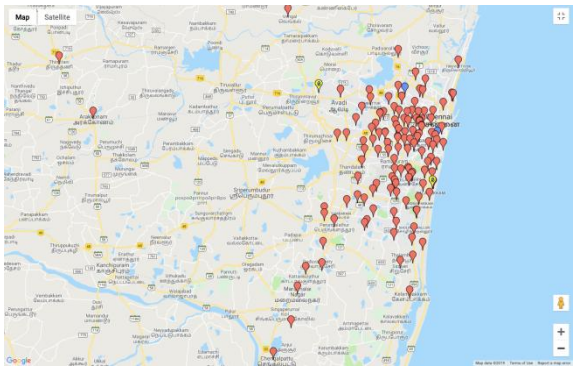

Fig. 5. Geographical location of Salons in Chennai (Salons in Chennai)

Clustering of Salons according to Revenue

The KMeans, GMM & DBScan was used to show clusters of Salons. This helps identifying the different class of low to high revenue salons.

Here the decision makers can concentrate on marketing for the low revenue generating salons.

Evaluation

To calculate the Silhouette Coefficient the mean intra-cluster distance and the mean nearest-cluster distance is used for each sample. The silhouette coefficient value is a measurement of similarities of an object is to its own cluster (cohesion) as compared to other clusters (separation). The silhouette ranges from −1 to +1. A high value indicates that the object is well matched to its own cluster and weakly matched to neighbouring clusters.

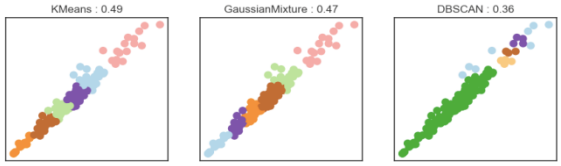

Fig. 6. Silhouette Score of Salons clustering (Silhouette score)

Silhouette score of KMeans is higher, so KMeans clustering is chosen.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

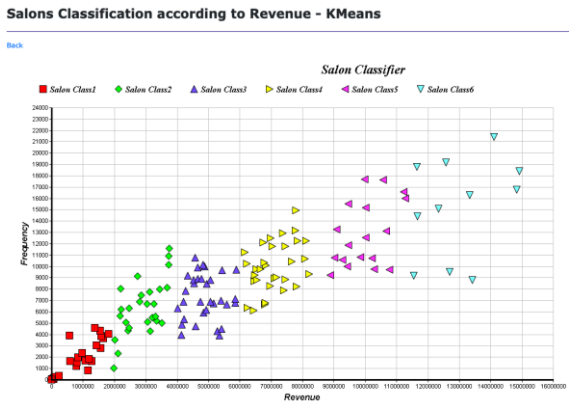

Fig. 7. Salons Classification with Salon Name List (KMeans clustering)

Clustering of Customers according to Revenue, Age

KMeans was used to show clusters of customers according to their age and revenue from them in a salon.

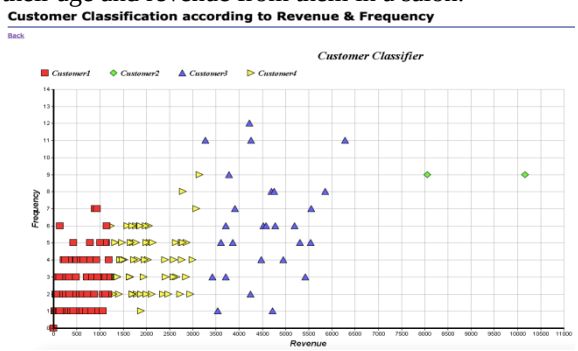

Fig. 8. Customers Classification - Revenue Vs Frequency (KMeans clustering)

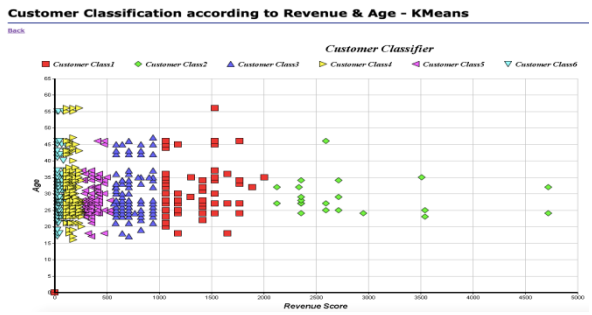

Fig. 9. Customer Classification – Revenue Vs Age (KMeans clustering)

Prediction of Sales – Month wise for all Salons

The attributes selected using their correlation to Sales. To predict the Revenue from customer, Linear regression is used

Table- IV: RFM for the dataset

Algorithm	RMSE
LinearRegression	12.07
Ridge	50.58
DecisionTreeRegressor	190.7
XGBRegressor	132.6

Evaluation

Evaluating the *RMSE*, we can get an idea of how far away each prediction was from the actual value in the test data, on average. We want this value to be as low as possible.

Optimal Utilization of Resources

The actual worked hours and the proposed hours, along with number of services, sales are calculated for each staff. The chart displays these details for a salon. This gives an idea of the idle time of staff.

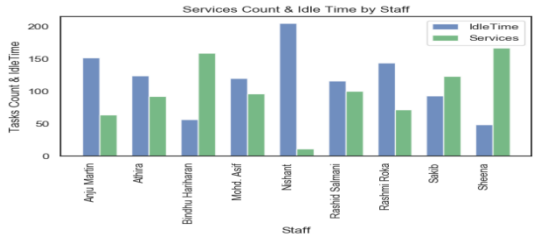

Fig. 10. Staff Services count & Idle time (Worked hours)

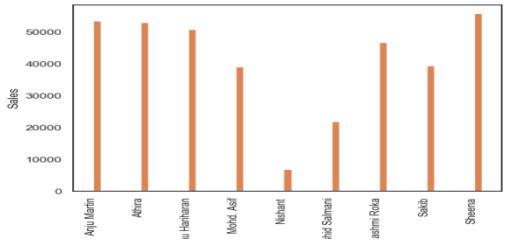

Fig. 11. Staff Sale Value (Sales per Staff)

Services

Peak service hours are analyzed. This helps in allocation of resources and plan tasks.

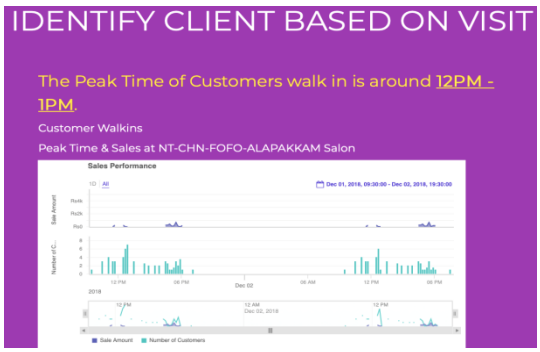

Fig. 12. Walk in Peak Time (Peak time of Sales)

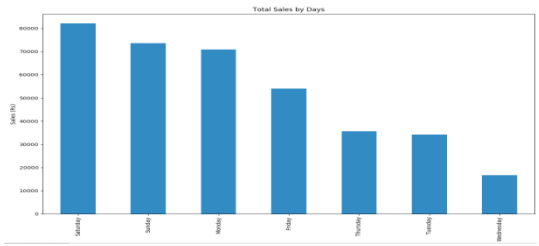

Fig. 13. Weekday Sale Value (Sales per day of week)

Retrieval Number: D6873118419/2019©BEIESP
DOI:10.35940/ijrte.D6873.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

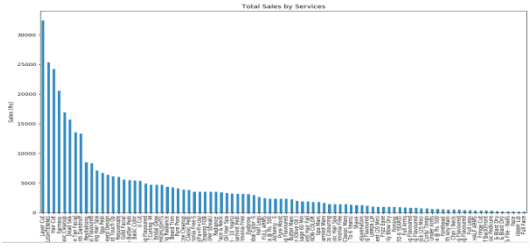

Fig. 14. Most Wanted Services (Frequent Services)

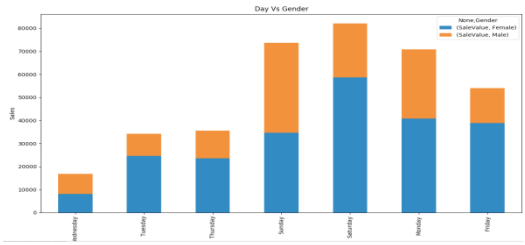

Fig. 15. Gender Vs Days of Week Sales (Gender based Services)

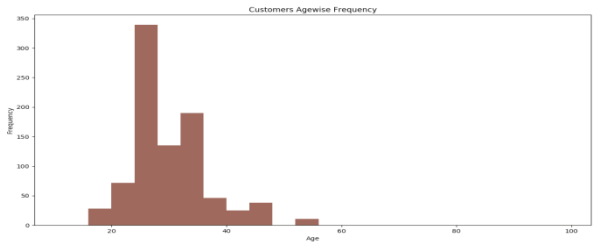

Fig. 16. Age range of frequent customers (Age wise walk-ins)

V. CONCLUSION

This analysis of salon data highlights the wide opportunity of data mining in beauty salons to improve their sales and invest in new ventures. In this study, the prediction of sales from salons and customers using the Linear Discriminant Analysis, Linear Regression algorithm. The grouping of salons and customers use the K Means clustering algorithm. This provides a suggestive grouping of salons and customers according to their revenue. The Geographical location of hotspots provide a visual view for expanding of new salons. The gender wise analysis provides an insight for opening all women or men salon instead of a unisex salon. Peak daytime and most wanted services analysis enables the staff scheduling accordingly. This analysis work can be expended further by adding include socio- economic factors that influence sales.

This result of this work assists in marketing decisions to promote sales and gain customers like rewarding customers for referrals, create a peak time salon services package to leverage more services, offer free services for purchase of gift card and so on. This can provide more business and job opportunities in the community.

REFERENCES

1. Esra Kahya Ozyirmidokuza, Kumru Uyarb, Mustafa Hakan Ozyirmidokuzc - A Data Mining Based Approach to a Firm's Marketing Channel
2. Mr.A.Sai Manideep, Dr.M.Siva Koti Reddy, Dr.P.Srinivasa Reddy - Competitiveness of Indian Wellness Industry A Conceptual Analysis
3. A.Rajasekar1, J.Samyuktha2, S. Meharaj - Applicability of Big Data Techniques to Smart Home
4. Ammar A. Q. Ahmed, Maheswari D - Churn prediction on huge telecom data using hybrid firefly based classification .
5. Danilo Corral-De-Witt 1,2,*, Enrique V. Carrera 1 ID , Sergio Muñoz-Romero 2,3 and José Luis Rojo-Álvarez - Statistical, Spatial and Temporal Mapping of 911 Emergencies in Ecuador
6. Elma Kolçe (Çela), Prof.Dr.Neki Frasheri - A Literature Review of Data Mining Techniques Used in Healthcare Databases
7. Fatih CIL1, Tahsin CETINYOKUS*2, Hadi GOKCEN2 - Knowledge Discovery on Investment Fund Transaction Histories and Socio-Demographic Characteristics for Customer Churn
8. Femina Bahari Ta, Sudheep Elayidom - An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour
9. Hemlata Sahu, Shalini Shrma, Seema Gondhalakar - A Brief Overview on Data Mining Survey
10. Hiziroglu, A & Seymen, Omer. (2014). Modelling customer churn using segmentation and data mining. Frontiers in Artificial Intelligence and Applications. 270. 259-271. 10.3233/978-1-61499-458-9-259.
11. Imas Sukaesih Sitanggang, Razali Yaakob, Norwati Mustapha - Application Of Classification Algorithms In Data Mining For Hotspots Occurrence Prediction In Riau Province Indonesia
12. Iva Šalov, Aleksandra Krajnović, Ante Panjkota. - Relation between Data Mining and Business Fields in the Four Dimensional CRM Model
13. Jo-Ting Wei a, Ming-Chun Lee b, Hsuan-Kai Chen c, Hsin-Hung Wu d - Customer relationship management in the hairdressing industry: An application of data mining techniques
14. K. Jaishankar, S. Shanmugapriya, V. Balamurugan (2004) - Crime Mapping in India: A GIS Implementation in Chennai City Policing, Geographic Information Sciences
15. Nikhil Chaudhari, Karthik Vardhrajan , Shashank Shekhar, Prabhjit Thind And Swarnalatha P - A Data Mining Approach To Language Success Prediction Of A Feature Film
16. Parneet Kaura,Manpreet Singhb,Gurpreet Singh Josanc - Classification and prediction based data mining algorithms to predict slow learners in education sector.
17. Sathyadevan, Shiju & S, Devan & S Gangadharan, Surya. (2014) - Crime analysis and prevention is a systematic approach for identifying and analysing patterns and trends in crime.
18. V. Maria Antoniate Martin, Dr. K. David, A.Vignesh - Big Data and Its Challenges
19. Wen-Yu Chiang - Identifying high-value airlines customers for strategies of online marketing systems An empirical case in Taiwan
20. Wen-Yu Chiang, (2018) - Applying data mining for online CRM marketing strategy: an empirical case of coffee shop industry in Taiwan
21. https://www.franchiseindia.com/wellness/future-and-growth-of-beauty-industry-of-india.11124
22. https://searchsqlserver.techtarget.com/definition/data-mining

AUTHORS PROFILE

Ms. Padma R is working as Assistant Professor in Department of Computer Science, Vels Institute of Science, Technology and Advanced Studies, Chennai.She is pursuing PhD in Computer Science from Bharathiar University, Coimbatore. She is presently guiding 1 Mphil Scholar . She has published many papers in various International journals including journals indexed in Scopus. She has presented many papers in International Conferences and attended many seminars and workshops conducted by various educational Institutions. Her research interest lies in the area of Big Data Analytics, Machine Learning.

Josephene Sheeba S, , Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies, Chennai.

Dr. A.Thirumurthi Raja , Assistant Professor, Department of Computer Science, Vels Institute of Science Technology & Advanced Studies, Chennai.