Social Media Analysis and online Marketing

Radha Yelikar[1]

Department of Computer Engineering,
International Institute of
InformationTechnology
Pune, India

Minakshi Chilme[2]

Department of Computer Engineering,
International Institute of
InformationTechnology
Pune, India

Pearl Swaminathan[3]

Department of Computer Engineering,
International Institute of
InformationTechnology
Pune, India

Eshwari Bhujbal[4]

Department of Computer Engineering,
International Institute of
InformationTechnology
Pune, India

Prof. Sandeep Patil[5]

Department of Computer Engineering,
International Institute of
InformationTechnology
Pune, India

---***---

Abstract—Social Media is incredibly trending platform for online marketing nowadays. It also provides people with the opportunity to share the post and information of business. Social Media Plays quite essential role in the growth of business using online marketing. Among millions of users, famous celebrities choose these platforms to establish their reputations. Social platforms have solved many issues of customers by responding to their complaints as quickly as possible. The idea behind all this is to create the web application which will allow users to get all the help in kick-starting their business on Social Media. We aim to provide a complete guide from suggesting which social media platforms will be good for their particular business. We are also providing report of monthly growth of business. User doesn't need to go on another website. The project is based on an observation made in the most growing field of Digital Marketing.

Keywords—Social Media, Online marketing, Digital Marketing

1. INTRODUCTION

Sharing information, topics is always have been a part of social media. Nowadays people are interestingly responding to the online world. Social media is one of the most used platform for promoting websites, products, business as it is a platform where friends and family gathering and connecting together virtually. Recommending friends about product and sharing the information using digital advertising always have been a cost-effective process than seeing a commercial on television.

Based on web technology development number of techniques and concepts are developed to implement online marketing strategy. In which most popular technique is data analysis. Data is gathered from the user's data which involves likes, comments and followers of the user's post. We are also providing report of monthly growth of business. By analyzing the previous data, in future we can suggest user the best platform for marketing.

Where there are number of social media platforms which is used for online promoting and marketing business. Facebook, Instagram, Twitter are trending platforms from all. As varieties of people use different sites to create and edit post for marketing purpose our proposed system will help the user to create and edit his/her own post also will provide a complete guide from suggesting which social media platforms will be good for their particular business, there are a lot of platforms which can be used for marketing in the digital world to analyzing data and of your audience base to managing paid campaigns.

2. LITERATURE REVIEW

We have referred "Research Trends of Online Marketing in Social Media Research" [1], They have predicted which online strategy will be popular for online marketing. For which they have examined the previous research paper. Where research papers which was in pdf format converted into text files. First step includes the tokenizing and filtering of the text files. Then the text files have analyzed with data mining technique. Furthermore, irrelevant information from the document removed using word length and stop words filtering techniques. Then text mining is done using the different methods such as similarity, total occurrences, TF-IDF, Association rule and clustering.

A. Data preparation

The first step in building the dataset has done with collecting a list of Instagram influencers, that is users who, as a part of their occupation, make marketed Instagram posts.

Finding this kind of information is a big challenge on its own, as such lists are not easily available to the public domain. We have filtered 70+ pages of influencers to obtain their handles in a usable list. Filtering the Instagram profiles of these users involves reading the metadata from the profile (number of followers/following, number of posts, description of profile). Instagram's API has a limit of 60 requests/hour to their backend servers, which created indirect barrier for any real application or gather of data. The alternative to official APIs is searching each page programmatically. The scraper we have used have coded using Selenium, a framework that is aimed at building functional tests for web applications. The scraper initially linearly scans the latest posts of a user, then opens each post to retrieve more granular information related to each image. 16539 images from 972 Instagram influencers have collected to train and test our model. Some brief data metrics, we can see that the number of likes has a high standard deviation, 61 224.20. It appears there is a high disparity as the mean is 24 416.38 but 75% of the posts have less than 18 359 likes.

	numberPosts	numberFollowing	numberFollowers	numberLikes
count	16537.00	16537.00	16537.00	16537.00
mean	2315.91	2590.88	997926.14	24416.38
std	2655.50	59094.23	1934497.75	64224.20
min	15.00	0.00	124965.00	0.00
25%	787.00	191.00	198039.00	3232.00
50%	1481.00	396.00	393843.00	7351.00
75%	2932.00	645.00	935310.00	18359.00
max	27671.00	1838511.00	22130730.00	1115123.00

The following histogram confirms the values shown on the table, the majority of the posts have less than 200k likes. In the application we are interested in, it makes sense to remove the people with a very high number of followers (more than 1 000 000 followers) and high average number of likes (above 200k) as they are no longer considered influencers, but celebrities/stars. It will also help to reduce the error of our model.
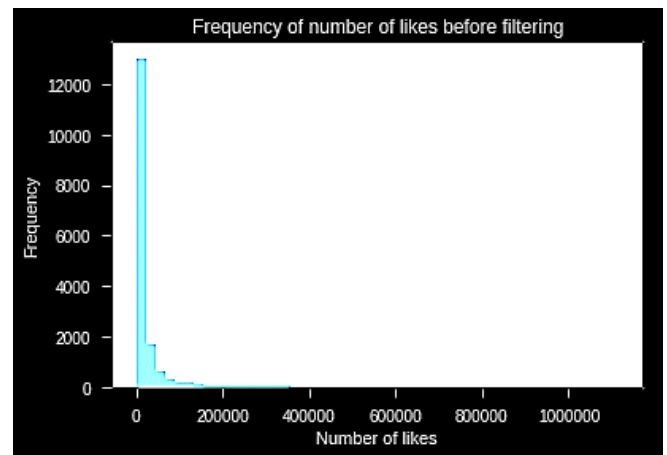

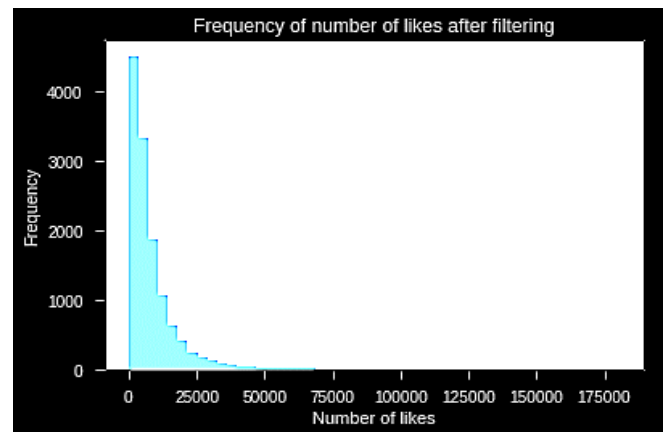Fig 1: Frequency of the number of likes before filtering

Fig 2: Frequency of the number of likes after filtering

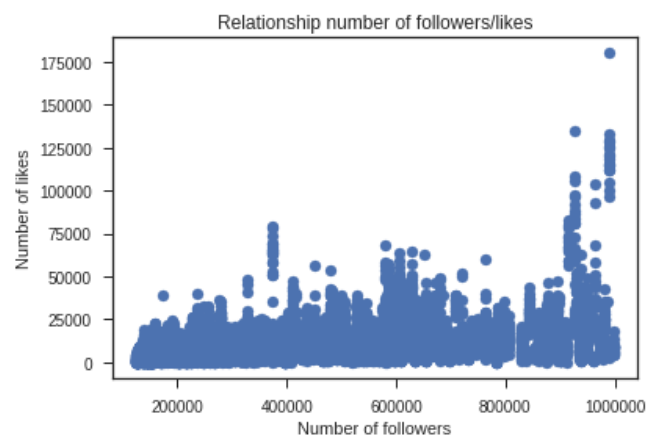We have plotted the number of likes versus the number of followers.

Fig 3: Relationship number of followers/likes.

III. System Description

Input: User will upload the image which they want to use as a post for further marketing. Later the image will be converted into the proper post.

Output: The website will provide a complete guide to user from suggesting which social media platforms will be good for marketing to their particular business.

Functions: 1) Photo editor: we are providing a function to create and edit the post which will help to edit the post for marketing purpose. Also we are providing the access to the different social media platforms such as (Facebook, Instagram, Twitter). So that user can efficiently post from the same site.
2) Analyzer: It provides analysis of the posts based on different platform i.e. Facebook, Instagram, Twitter. And will suggest user Which platform will be best for their business.
3) Scheduler: This feature will help to schedule the post on any day and time. User need not to be present manually to post. Which will increase the efficiency of the website.
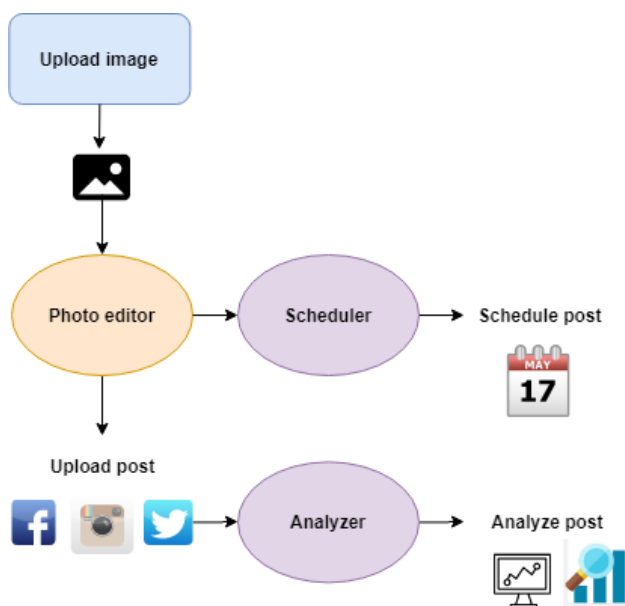

Fig.4.Architecture Diagram

We have first developed a base model containing some basic features obtained from the dataset. Then, added features obtained from Natural Language Processing (NLP) and finally, added features generated from a Convolutional Neural Net. To compare the different models, we have used two performance metrics: Root Mean Square Error (RMSE) and the R^2 value.

A. Base Model:

The base model consists of the following features:
Given Features:
- Number of followers
- Number of following
- Number of posts

Extracted features:
- Website: we classified the website provided in the user description into different categories: YouTube, Facebook, Twitter, Blog, Music and other. Then we one hot encoded the different categories.
- Day of the week: using the data of each data post we one hot encoded the day of the week

Generated feature:
- Average number of likes: we observed that the number of followers does not necessarily yield a high number of likes. Inactive and fake followers impact the results, a better metric appeared to be the average number of likes of each user. The rest of the features are supposed to vary the number of likes up or down depending on the average number of likes.

From the features generated we found that:
1) Website
11% of the user had a YouTube channel or video as their website, 4% a Facebook profile, 2% a blog and 1% a website related to music (Sound cloud or Spotify). 88% of the user had a website in their user profile.
2) Day of the week
The day where the highest number of pictures are uploaded is Sunday with 18%, then Saturday and Thursday with 15%, Friday and Wednesday with 14% and finally Monday with 11%. It appears that only two days have different statistics, Sunday has more posts and Monday less.
3) XGB
Applying an XGBoost model on these features with the following parameters: max_depth=4, learning_rate=0.01, n_estimators=59
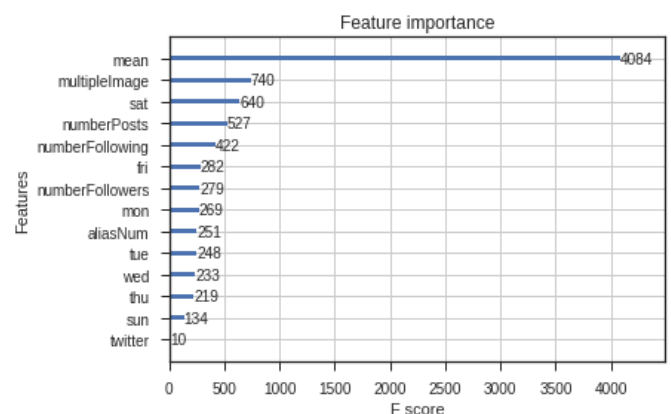
From the feature importance plot given by XGBoost we can see that the average number of likes significantly impact the outcome of the XGB model. The F score, 4084, is bigger than all the other features combined. In terms of basic features, number of posts, following and followers have a low F score. In terms of extracted features, Saturday appear to have a bigger impact than the other day. Perhaps suggesting that posting on Saturday yield a higher number of likes than the other. Earlier, it was said that the distribution of the posts during the week is relatively uniform with more posts published on Sundays and less posts on Mondays.

B. NLP (Natural Language processing):

When trying to figure out what subset of the NLP features we would use for our final model, we tried to predict the number of likes of a post given a subset of NLP data. We chose the subset of data that reduced our RMSE (Root Mean Square Error) the most. After testing a couple subsets, we found that the best subset was this: the top 500 words/hashtags from individual post captions, and emoji's that show up over 175 times (37 emoji's). When we appended these features to the features we used for the base model, our RMSE actually worsened. What happened was, these extra NLP features were away importance from the mean feature in the base model features. Since the Base + NLP model relied less on the mean feature to predict likes, our RMSE actually worsened. Our RMSE score went up to 2895.90 and our R^2 score became 0.9163. Since there are over 500 features, the feature importance plot for this model is hard to read.

C. Transfer learning:

Another one of the approaches we took is related to image processing and computer vision. The aim is to identify features related to images that may be meaningful in determining the final number of likes. Transfer learning relates to the process of using a pre-trained deep ConvNet, and use that as a starting point to build a model that takes into account image features. There are two ways to use pretrained models, we have both tackled:

- **Fine-tuning the ConvNet with Inception V3:** that is taking the pre-trained weights of the ConvNet, remove the last fully connected layers, and expand the networks as needed. We then fine-tune the weights of the pretrained network by continuing the backpropagation. The model we used was Inception v3. The last two fully connected layers were removed, and we added three extra layers. We retrained it on our data points with a GPU on an EC2 machine.
- **Fixed feature extractor with VGG 19:** this is removing the last fully-connected layer, then treat the rest of the ConvNet as a fixed feature extractor

for the new dataset. VGG 19 is an image classification network pretrained on the 14+ million images on the ImageNet dataset. Removing the last fully-connected layer yields a feature vector of size 4096 for each image.
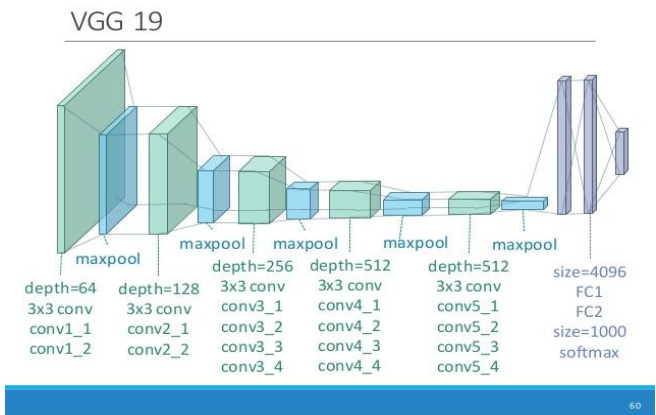

Fig. 5: VGG 19 (Used as features in base model)

IV. PROPOSED SYSTEM

Sharing information, topics is always have been a part of social media. Nowadays people are interestingly responding to the online world. Social media is one of the most used platform for promoting websites, products, business as it is a platform where friends and family gathering and connecting together virtually. Recommending friends about product and sharing the information using digital advertising always have been a cost-effective process than seeing a commercial on television.

To help a user for online marketing we have managed all social media to one platform. To avoid the user's ambiguities its necessary that our system is perfectly working as expected. To implement a system that can increase the efficiency and solve the user's problem we have included scheduler and analyzer feature. Where user need not to be present manually for marketing purpose. Also analyzer will complete guide the user from suggesting the best platform for marketing based on previous analysis.

A. Model building:

Firstly, the complete model of the system is made containing the GUI of the system which includes the dashboard of home page containing photo editor, Scheduler and Analyzer.
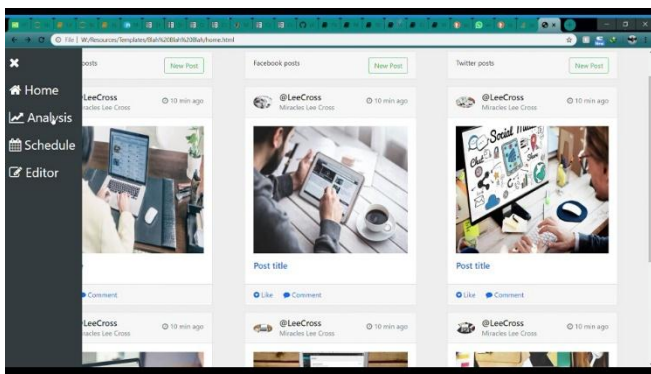
Fig. 6. Home Page

B. Input:

As input user should provide the image related to the business. Later uploaded image will be converted into proper post with the help of photo editor which user can use for marketing purpose.
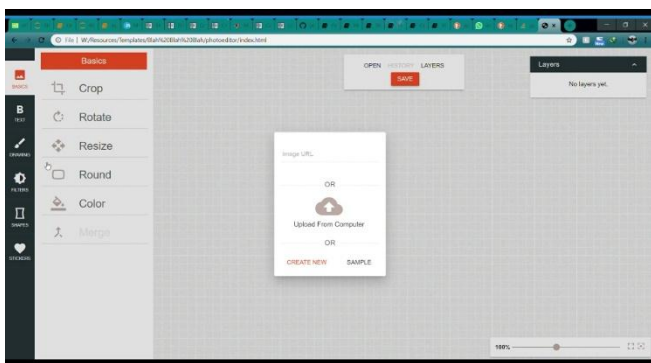

Fig. 7. Photo Editor

C. Analyzing phase

It provides analysis of the likes and comments of posts daily, weekly and monthly based on different platform i.e. Facebook, Instagram, Twitter. And will suggest user Which platform will be best for their business. It shows the graphical representation of the done analysis with histogram and pie chart where pie chart shows the social platforms engagement distribution in percentage % format.
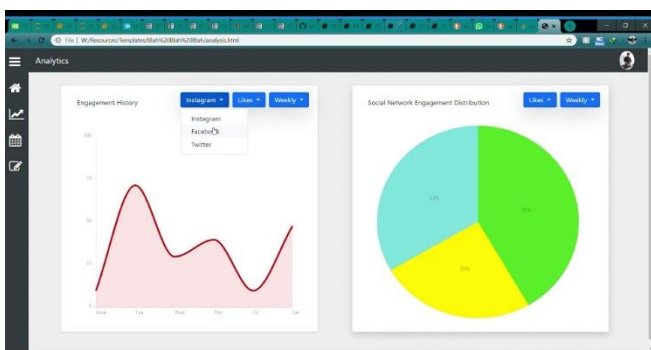

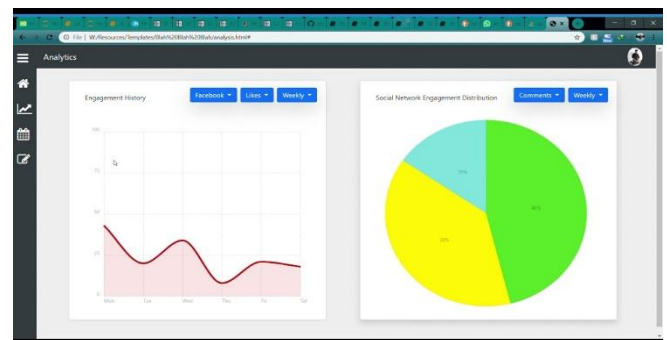Fig. 8. Weekly Instagram post analysis of likes.

Fig. 9. Weekly Facebook post analysis of comments.

D. Scheduling phase

In this process,editing th post user can schedule the perticular post on any date and time whereas user need not to be present manually at the time of uploading the post. Which also helps in increasing the efficiency of website.
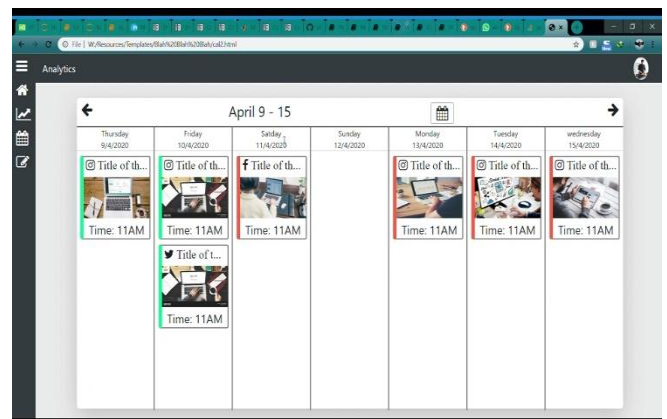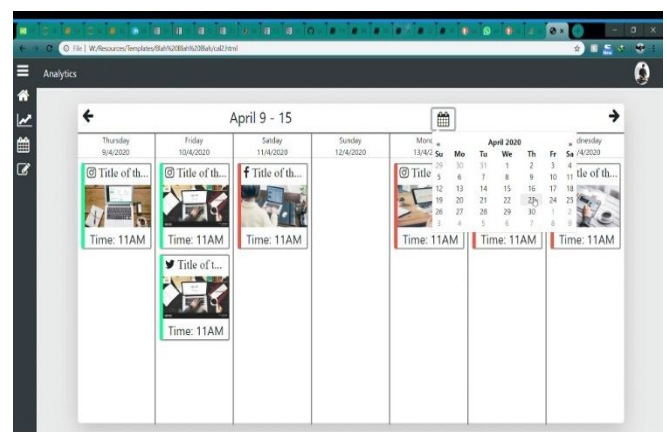

Fig. 10. Scheduling phase

Fig. 11 .Scheduling post on perticular day and time

V. RESULT AND DISCUSSION

Our main outcome is overall providing guidance and suggestion on which platform is best for marketing

particular business based on previous analysis. We have successfully implemented website which also provides scheduling featured. We created a one stop solution to all social media marketing needs for the upcoming a growing business who can't afford the retainer ship of Agencies. It is real time application which uses data analysis and digital behavior of the vast audience by using NLP (Natural Language Processing and transfer learning.
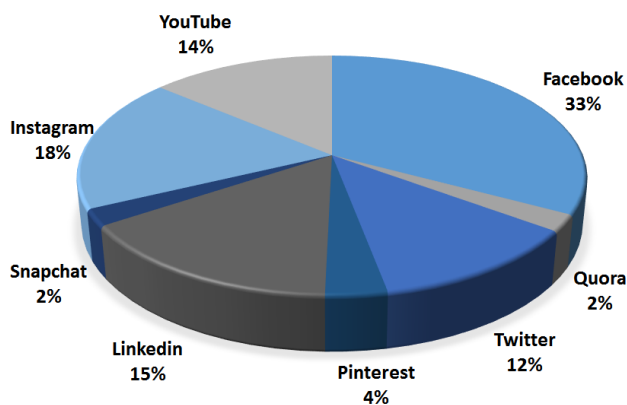
Success graph of social media marketing:

Fig. 12 . Social Media marketing success graph

VI. CONCLUSION AND FUTURE WORK

After deep analysis and various surveys in the digital marketing community and businessmen we came to the conclusion that there are few or no platforms that provide one stop solution to know, learn and executive everything in terms of social media. The aim of this project is to create a one stop solution to all social media marketing needs for the upcoming a growing business who can't afford the retainer ship of Agencies. To develop a website where the user will be able to schedule, analyze and monitor their social media posts. To develop a website where the user will be able to create and edit their creative/templates on our website. In this system, we have tried to create real time application which uses data analysis and digital behavior of the vast audience by using NLP (Natural Language Processing and transfer learning. In Future we are trying to provide security to our system which can secure the users personal information. In future we will try to develop mobile application which will manage all social media sites to one platform. Also, there is a scope and need of online marketing in the field of business.

REFERENCES

[1] Marina V. Terskikh "Strategy of Sports Brands Promotion in Social Networks" IEEE, 2018.
[2] Harsh Namdev Bhor, Rajat Malviya "Digital Media Marketing using Trend Analysis on Social Media", IEEE Xplore Compliant, 2018.
[3] Prajaks Jitngernmadan, Prawit Boonmee "Research Trends of Online Marketing in Social Media Research." Digital Media and Interaction Laboratory, 2018.
[4] Shilpa h. Chheda "Impact of social media marketing on performance of micro and small businesses" July-2018.
[5] Lu Liu, Haoting Liang "Influence analysis for celebrities", China-Communications August 2016.
[6] Helena Alves, Cristina Fernandes, Maria Reposo "Social Media Marketing: Implications of social media marketing", Psychology and marketing, Dec 2016.
[7] Manoj Kumar Srivastav, Asoke Nath " Study on Mathematical Modeling of Social Networks" ,International journal of Emerging technology and advanced engineering,2018.
[8] Ying Hsun Hung, Seng-Cho T. Chou "Exploring the Management of Government Social Media", IEEE /ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),2018.
[9] Dave Evans,Jake McKee "Social Media Marketing: The Next Generation of Business Engagement".
[10] Dr. Lalit Khullar, Dr. Sanjay Kaveshwar, Dr.Sangeeta Deshpande "Business Remodeling",2018.