Data Mining on Social Networks for Target Advertisement in Automobile Sector

Sudhanshu Chhabra[1], Renuka Mahajan[2], Sunil Kumar Khatri[3]

[1,2,3]*Amity Institute of Information Technology,*
Amity University Uttar Pradesh Noida India
[1]*sud.chabra@gmail.com,* [2]*rmahajan@amity.edu,* [3]*skkhatri@amity.edu, sunilkkhatri@gmail.com*

Abstract: There is huge potential to predict hidden social ties between different people sharing common area of interest or common friends. This paper presents data analysis on social network data for target advertising of the automobiles. The interrelated groups are derived from client's interaction data in social networks. Based on these cohesive subgroups, the likelihood of client's preference of a product category is inferred. Using this information, we construct a targeted advertising system. The experimental results on the social network dataset yields better quality of advertisement.

Keywords: Word Cloud, Data Mining, Predictive Analysis, Behavioral Aiming, Advertisement, Decision Matrix.

I. INTRODUCTION

The importance of advertising using client's feedback has long been acknowledged by businesses [9]. A lot of attention is attracted towards this research area for many reasons. There is a huge amount of product/service information available to clients. Hence, it is desirable for having different ways with which clients can wade through information that could help them find product/services. Secondly, it is important to understand what clients (both current and potential) needs, as a part of client relationship management (CRM). If one could identify the needs of clients more accurately and efficiently, they can advertise the products and services in a better way. It would increase the retention power of the company for its clients, the growth of the company and its profits.

Social data mining helps the organizations in targeting the consumer using behavioral and sentimental approach [25]. There are various popular social networking sites such as Facebook, Twitter, Instagram and Google+. People are also Connected to each other through various E-commerce sites [26]. Companies need to quantify and classify the feedback to make different business strategies by understanding emerging trends in a virtual world. [2].

This analysis is useful both for marketers as well as the clients. On one hand, it helps clients to understand popular preferences about a product or service. It is particularly useful for companies in evaluating the online feedback related to their products. These ideas and opinions propagating on a social network have been studied to improvise products and design new strategies of viral marketing [10], [11],[12][24].

II. RELATED WORK

Similar work was undertaken in paper [1], where development of an android application based on BLE device, was used to track the location of user. This app sends the data surfed by the user to admin and will advertise the user on device as per the interest and shops available near the location.

Previous work also shows mining and analyzing the data provided by the government. Authors in [4] have analyzed the data of one particular society to advertise the product as per the customer's need. By dividing groups into subgroups, the social data for targeting the correct group for the advertisement of the product to attract the consumers is done by authors of paper [3].

Earlier, the database of previous transactions was manually analyzed and the features that are associated with potential clients were extracted. This was done with some statistical tools and would identify clients that might respond to advertisements of products. With these new technologies that are coming up now, we could identify potential clients using automatic tools. Hence, many recommender systems have emerged over the past couple of years since the basic idea behind these systems is to advertise products according to the preferences of the users. These preferences could be by getting the ratings that are stated explicitly or inferred implicitly from their previous records, web logs or even cookies.

The structure of the paper is as follows. The proposed work is explained in section III. Section IV gives the conclusion and directions for future work.

III. PROPOSED WORK

To conduct this research the datasets for mining the social data was required. The data was fetched from twitter using R studio as a tool.

The code was developed for fetching the live clickstream data from twitter using APIs and Key Tokens provided by twitter. The data was thus fetched using hashtags of different automobile companies. All the dataset was then saved in .csv

978-1-5386-0514-1/17/$31.00 ©2017 IEEE

format in an excel file. After that data filtration was done on the basis of factors like "location of tweet, ratings of car, about which feature of car people are doing tweets." Further sentimental data analysis was done on the filtered dataset. Frequency of occurrence of words was recorded by writing code in R studio. The Word Cloud of words frequently searched in twitter is represented in the Fig 1.

Fig. 1. Word Cloud representation of cars frequently searched on twitter

Data analysis in Fig 2 shows the count of persons over the internet that are discussing about that particular automobile company. The graph shows the popularity of automobile companies Maruti, followed by BMW, Volkswagen, Toyota, Nissan, Lexus, Lamborghini, Hyundai, Honda, Ford, Ferrari, Chevrolet and Audi over the internet. With the help of this graph companies can easily get to know that how popular their brand is.

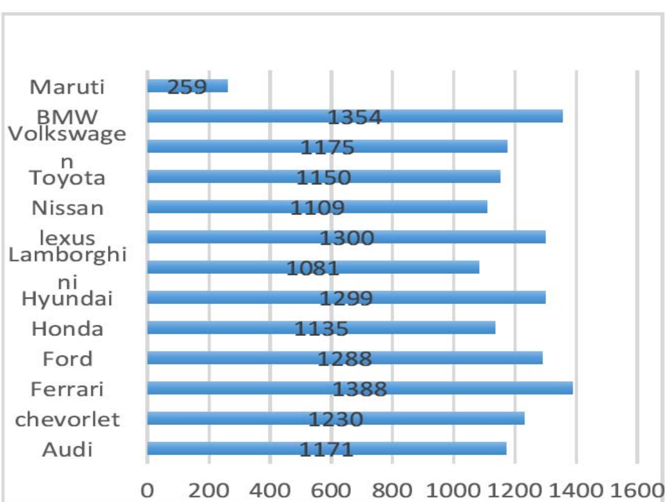

Fig..2. Graphical representation of companies most twitted

Benchmarks help the buyer and supplier both in deciding the type of car the particular group of user want to buy. To find the consumer reviews, different benchmarks were set by fetching the data. Fig 3 represents the benchmarks set while doing the research data analysis. This will help the consumer to know complete performance of a particular car and this will make decision easy for the buyer to buy that particular car.

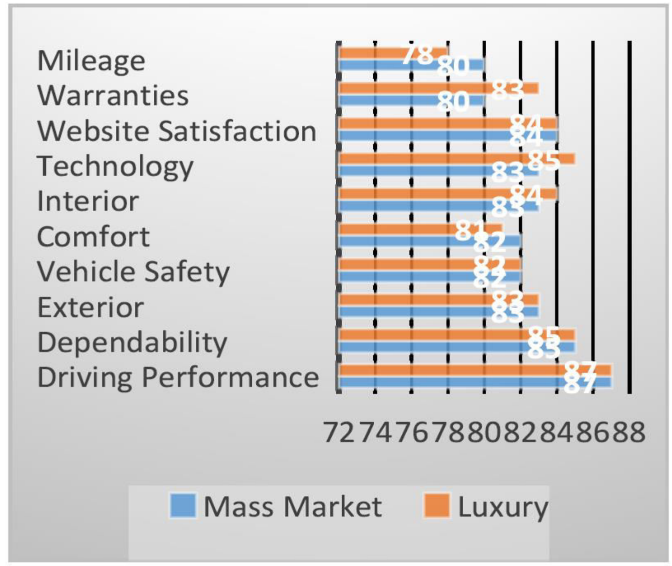

Fig. 3. Graphical representation of benchmarks

Fig 4 further shows popularity of brand between different age groups. This will help the company in targeting the right customer in right way of different age groups. The age group data was collected from 2000 existing users over the internet by creating a form.

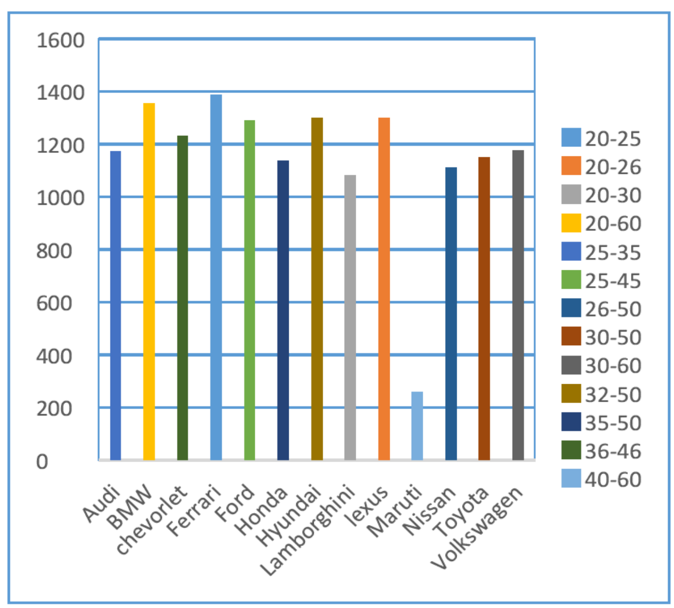

Fig. 4. Usage of car in different age groups

These results will help company and buyer with manufacturing decisions, improvisations and purchasing decisions about a particular car.

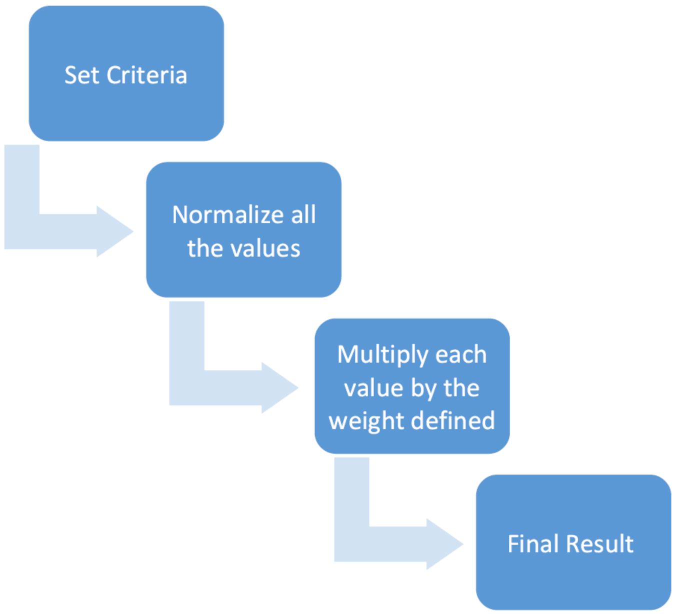

Fig. 5. Steps for making right decision

For smart and efficient decision which can be done by consumer to buy a car we propose a decision matrix on the basis of twit history about cars and its different features extracted which a consumer wants while buying a car. To make a smart decision, a flow chart is presented above in fig 5.

Flow Chart Steps

Step 1: Set the criteria for creating decision matrix table. For example: power, fuel economy, price, safety, reliability, cargo capacity, and the ability to hold your two teenagers comfortably. These are the key constraints which every consumer compare before buying a car.

Table 1, shows the value of each criteria of a car chosen by customers to show how the decision matrix works. It contains all the values of the cars along with the safety measures.

TABLE 1: Raw Data

Automobile	**Honda Jazz**	**Nissan Sunny**	**VW Polo**	**Audi Q7**	**Ford Endeavour**
HP	110	119	149	160	154
WT (kg)	1140	1250	1420	1350	1470
Power/weight	0.090	0.092	0.104	0.120	0.101
I/100k city	7.7	7.6	10.2	9.4	10.5
I/100k h-way	5.5	6	7	6.5	8.2
Avg I/100k	6.6	7	8.7	8.2	9.2
Rear legroom (mm)	866	976	886	920	920
Cargo up (L)	602	503	400	480	820
Cargo down (L)	1180	1420	1000	888	1867
Safety Stars					
Driver frontal	5	4	4	4	4
Pass frontal	5	4	4	4	4
Front side	5	5	5	3	5
Rear side	3	5	5	3	5
Rollover	4	4	4	4	3
Average	4.4	4.4	4.4	3.6	4.2
Reliability stars	5	3	1	4	3
Total price($)	22350	22460	26492	28265	32567

Step 2: Table 2 shows normalized value of the criteria which was chosen, Normalization is done by giving highest value of each category rank 1 and divide the rest value by that.

Step 3: Finally, weight the relative importance of each criterion.

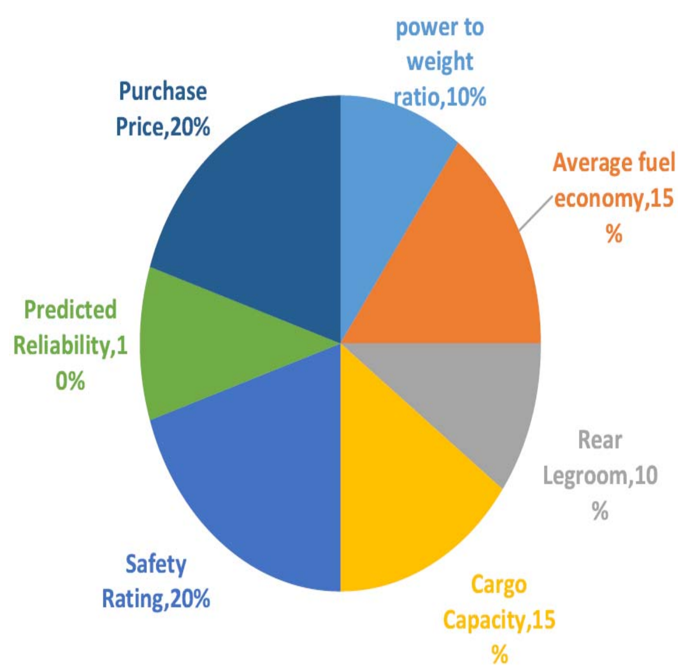

Fig. 6. Weight of different criteria.

Step 4: Table 4 shows final result of the research. On the basis of this decision matrix, consumers can easily buy the car of the choice which is suitable to him as per his requirement.

TABLE 2: Normalized Criteria Rating

	Jazz	Sunny	Polo	Q7	Endeavor
Power/wt.	0.68	0.74	0.93	1	0.96
Avg fuel economy	0.70	0.81	0.91	0.96	0.92
Legroom	0.73	0.72	0.97	0.89	1
Cargo	0.69	0.74	0.89	0.84	1
Safety rating	0.88	1	0.90	0.94	0.94
Reliability	0.73	0.61	0.48	0.58	1
Price	0.63	0.76	0.53	0.47	1

TABLE 3: Final Result

	Weight %	Jazz	Sunny	Polo	Q7	Endeavour
Power/wt.	10	6.8	7.4	9.3	10	9.6
Avg. fuel economy	15	10.5	12.15	13.65	14.4	13.8
Legroom	10	7.3	7.2	9.7	8.9	10
Cargo	15	10.35	11.1	13.35	12.6	15
Safety	20	17.6	20	18	18.8	18.8
Reliability	10	7.3	6.1	4.8	5.8	10
Price	20	12.6	12.2	9.6	11.6	20
Total score	100	72.45	76.15	78.4	82.1	97.2
Rank		5	4	3	2	1

This decision matrix also helps companies to advertise their product in correct manner and this will also help them in targeting the customer by advertisement.

IV. CONCLUSION

In this paper, the propensity of clients sharing common interest can be used to mine and predict client feedback on cars. Our methodology demonstrates that meaningful hashtag categories can be inferred. From the data analysis on the basis of popularity, it will be easy for company in hitting the correct groups. The trends detected will help in analyzing the popularity. The benchmarks thus discovered will help the consumer in purchasing the car wisely. They help automobile companies in predicting the demand of all the cars amongst the users. The graphical results show shared interest and feedback of the clients that can be used by both car manufacturers as well clients.

ACKNOWLEDGEMENT

Authors express their deep sense of gratitude to The Founder President of Amity Group, Dr. Ashok K. Chauhan, for his keen interest in promoting research in Amity University and have always been an inspiration for achieving great heights.

REFERENCES

[1] Archana Singh and Ajay Rana, "Mining of Customer data in an Automobile Industry using Clustering Techniques", International Journal of emerging Technologies in Computational and Applied Sciences, 2013, pp.251-258.

[2] Armstrong, Kotler "Marketing: an introduction", Prentice-Hall; 1999.

[3] David Kmpe, Jon Kleinberg, and Ev´ a Tardos, "Maximizing the spread of influence through a social network," Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2003, pp.137-146.

[4] M.M. Mostafa, "More than words: Social Networks' text mining for consumer brand sentiments", Expert system with application, 2013, vol. 40, Issue 10, pp.4241-4251.

[5] Mathew Richardson, and Pedro Domingos, "Mining knowledge-sharing sites for viral marketing," Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2002, pp.61-70.

[6] Ming-Hung Wang and Chin-Laung Lei, "Boosting Election Prediction Accuracy by Crowd Wisdom on Social Forums", 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2016, pp.348-353.

[7] Monu Kumar and Dr. Anju Bala, "Analyzing Twitter Sentiments through Big Data" International Conference on Computing for Sustainable Global Development (INDIACom), 2016, pp.2628-2631.

[8] Pedro Domingos, and Matt Richardson, "Mining the network value of customers," Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001, pp.57-66.

[9] Prabhakar, E., Parkavi, R., Sandhiya, N. and Ambika, M. "Public Opinion Mining for Government Scheme Advertisement", International Journal of Information Research and Review", 2016, vol. 3, Issue 4, pp. 2112-2114.

[10] Sanjeev Pippal, Lakshay Batra, Akhila Krishna, Hina Gupta, Kunal Arora, "Data Mining in social networking sites : A social media mining approach to generate effective business strategies", International Journal of Innovations & Advancement in Computer Sciences (IJIACS), 2014, vol. 3, Issue 2, pp.22-27.

[11] Svitlana Volkova, Yoram Bachrach, Benjamin Van Durme "Mining User Interests to Predict Perceived Psycho-Demographic Traits on Twitter" IEEE Second International Conference on Big Data Computing Service and Applications, 2016, pp.36-43.

[12] Vrinda Bhatia and Varun Hasija "Targeted advertising using behavioral data and Social Data Mining", Eighth International Conference on Ubiquitous and Future Networks, 2016, pp.937-942

[13] Wan-Shiou Yang, Jia-Ben Dia, Hung-Chi Cheng, Hsing-Tzu Lin "Mining Social Networks for Targeted Advertising ", Proceedings of the 39th Hawaii International Conference on System Sciences, 2006, vol. 6, pp.137a-137a.

[14] Rakesh Kumar Donthi, Vishwesh Nagamalla, Srujan Kumar "A Framework for Semantic Advertisement over Social Media" IJCSN International Journal of Computer Science and Network, Vol 2, Issue 1, 2013.

[15] Sharad Goel, Daniel C. Goldstein "Predicting Individual Behavior with Social Networks" Vol. 33, No. I, January—February 2014, pp. 82—93.

[16] Fostor Provost, Brian Dalessandro, Rod Hook, Xiaohan Zhang, Alan Murray "Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting" *KDD'09,* June 28–July 1, 2009, Paris, France. Copyright 2009 ACM 978-1-60558-495-9/09/06.

[17] Killian Thiel, Tobias Kotter, Dr. Michael Berthold, Dr. Rosaria Silipo, Phil Winters, Creating Usable Customer Intelligence from Social Media Data: Network Analytics meets Text Mining"

[18] Aston Zhang, Lluis Garcia-Pueyo, James B. Wendt, Marc Najork, Andrei Broder "Email Category Prediction" 2017 International World Wide Web Conference Committee (IW3C2), published under Creative Commons CC-BY-NC-ND 2.0 License. WWW 2017 Companion, April 3–7, 2017, Perth, Austraila. ACM 978-1-4503-4914-7/17/04.

[19] Lynne Grewe, Sushmita Pandey "Quantization of Social Data for Friend Advertisement Recommendation System" PDCTA 2011, CCIS 203, pp. 596-614

[20] Dejan Markovikl, Sonja Gievska, Michal Kosinski, David Stillwell "Mining Facebook Data for Predictive Personality Modeling" AAAI Technical Report WS-13-01 Computational Personality Recognition.

[21] Chris Sumner, Alison Byers, Rachel Boochever and Gregory. J. Park "Machine prediction of personality from Facebook profiles" DOI: 10.1109/IRI.2012.6302998

[22] Bogdan Batrinca, Philip C. Treleaven "Social media analytics: a survey of techniques, tools and platforms" AI & Soc (2015) 30:89–116 DOI 10.1007/s00146-014-0549-4.

[23] David Stuck, Haraldur Thomas, Greg Ver Steeg, Alessandro Epasto, Luca Foschini "The Spread of Physical Activity Through Social Networks" 2017 International World Wide Web Conference Committee (IW3C2), published under Creative Commons CC BY 4.0 License. ACM 978-1-4503-4913-0/17/04. http://dx.doi.org/10.1145/3038912.3052688.

[24] Data Fetched from twitter www.twitter.com

[25] Xu G, Zhang Y, Li L. Web Mining and Social Networking, Techniques and applications, 1st edition. Springer; 2011.

[26] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," Journal of Computer-Mediated Communication 13(1):210-230, 2008.

[27] Sanjeev Dhawan, Kulvinder Singh and Tanya Arora ,"Examining Polarization of Emotions in Online Shopping Websites," International Journal of Scientific & Engineering, Research, Volume 7, Issue 12, December-2016, pp: 379-382.

[28] eam.com/meta-tags