A NOVEL CLASSIFICATION APPROACH OF TRAVEL REVIEW DATASET BASED ON HEALTH

Dr.Ayyappan.G, Dr.A.Kumaravel
Associate Professor, Machine Learning Group, Department of Information Technology,
BIHER, Bharath Institute of Higher Education and Research, Chennai.
Professor, Machine Learning Group, Department of Information Technology,
BIHER, Bharath Institute of Higher Education and Research, Chennai.
ayyappangmca@gmail.com

Abstract - Recommender systems are progressing as a vital part of every industry with no exemption to travel and tourism segment. Considering the exponential proliferation in social media usage and huge volume of data being spawned through this channel, it can be considered as a vital source of input data for modern recommender systems. This in turn resulted in the need of efficient and effective mechanisms for contextualized information retrieval. Traditional recommender systems adopt collaborative filtering techniques to deal with social context. However, they turn out to be computationally intensive and thereby less scalable with internet and social media as input channel. A possible solution is to implement clustering techniques to limit the data to be considered for recommendation process. In tourism environment, based on social media interactions like reviews, forums, blogs, feedbacks, etc. travelers can be clustered to form different interest groups. This experimental study aims at comparing key clustering algorithms with the aim of finding an optimal option that can be adopted in tourism domain by applying social media datasets from travel and tourism context.

Keywords: Recommender Systems, Clustering Algorithms,Travel and Tourism, Cluster Evaluation

I. INTRODUCTION

In this section presents introduction of this research work.In end user perspective, travel and tourism is mostly explorative in nature and repetitive travels to same locations are minimal. So, travelers have to take decisions regarding their destinations and associated facilities to be consumed without adequate prior or personal knowledge. The best option available is to leverage social media and internet, but the amount of time required to extract relevant information is too high. Tourism recommenders are the best solutions in this scenario. Recommender systems helps in terms of automated filtering, processing, personalization and contextualization of the huge volume of data that is available and growing on a daily basis on the internet and the social media.
In this paper presents section 2 of this paper explains the detail on the related works. In section 3 presents the materials and methods adopted and section 4 presents the details of the experiments and discussions. Finally section 5 concludes the paper by sharing our inferences and future plans.

II. RELATED WORKS

In this section presents focuses the related works of this research work.A. Clustering in machine learning world is an unsupervised approach of grouping a set of entities together so that the entities in one group are more similar to each other than to the entities in another group. Unsupervised learning is applied while there is input data, but there is no corresponding outputvariables associated with it. Its goal is to understand and model the underlying distribution of data so as to learn more about it. Clustering has various applications like market segmentation for targeted advertisements and promotional offers, grouping of web contents in a search engines, text summarization, biological applications, astronomy, etc. Clustering reveals natural and meaningful groups among available data. Clustering algorithms aims to achieve highest intra-cluster similarity and least inter-cluster similarity. The concept of distance measure is used to calculate the similarity between objects. When the distance measure between two entities is very less, they are considered as similar. Based on the data under consideration appropriate distance measure can be chosen for clustering. A few of the most common distance measures include Euclidean, Manhattan, Cosine, Jaccardand Minkowski distances.

Clustering Algorithms can be generally categorized into three groups – partitioning [4], hierarchical and density based clustering. Partitioning clustering is used to categorizeobservations within a dataset based on their similarity. In this approach, the user has to identify the optimal count of clusters for the dataset in consideration and it need to be mentioned to the algorithm. The common partitioning clustering algorithms are k-means clustering [5][6], k-medoids clustering which is also known as Partitioning Around Medoids (PAM) [7][8], Clustering for Large Applications (CLARA) [8][9][10] .

III. MATERIALS AND METHODS

In this section presents the materials and methods of this research work.Reviews on destinations in 10 categories mentioned across East Asia. and average rating is used.This data set is populated by capturing user ratings from Google reviews. In this research work has implemented in Weka3.8.3. version.

Dataset Description

Reviews on attractions from 24 categories across Europe are considered. Google user rating ranges from 1 to 5 and average user rating per category is calculated. Each traveler rating is mapped as Excellent(4), Very Good(3), Average(2), Poor(1), and Terrible(0)

S.No	Attribute Name
1	Unique user id
2	Average ratings on churches
3	Average ratings on resorts
4	Average ratings on beaches
5	Average ratings on parks
6	Average ratings on theatres
7	Average ratings on museums
8	Average ratings on malls
9	Average ratings on zoo
10	Average ratings on restaurants
11	Average ratings on pubs/bars
12	Average ratings on local services
13	Average ratings on burger/pizza shops
14	Average ratings on hotels/other lodgings
15	Average ratings on juice bars
16	Average ratings on art galleries
17	Average ratings on dance clubs
18	Average ratings on swimming pools
19	Average ratings on gyms
20	Average ratings on bakeries
21	Average ratings on beauty & spas
22	Average ratings on cafes
23	Average ratings on view points
24	Average ratings on monuments
25	Average ratings on gardens

IV. EXPERIMENTS AND DISCUSSION

In this section discusses results and analysis of this research work. In Bayes classifier, BayesNet accuracy was 94.17% and NaiveBayes accuracy was 94.50%. In Lazy classifier, IBK(K Nearest Neighbor) accuracy was 93.95% and KStar accuracy was 90.36%. In Meta classifier, Bagging accuracy was 25.36 % and LogitBoost accuracy was 94.22%. In Rules classifier, Decision Table accuracy was 94.61% and JRip accuracy was 94.39% In Trees classifier, DecisionStump accuracy was 94.39% and J48 accuracy was 94.50%.

Table 1: Various Classifications with accuracy

S.No	Category of the Classifier	Name of the Classifier	Accuracy
1	Bayes	BayesNet	94.17 %
2		NaiveBayes	94.50 %
3	Lazy	IBK	93.95 %
4		kStar	90.36 %
5	Meta	Bagging	25.36 %
6		zLogitBoost	94.22 %
7	Rules	DecisionTable	94.61 %
8		Jrip	94.39 %
9	Trees	DecisionStump	94.39 %
10		J48	94.50 %

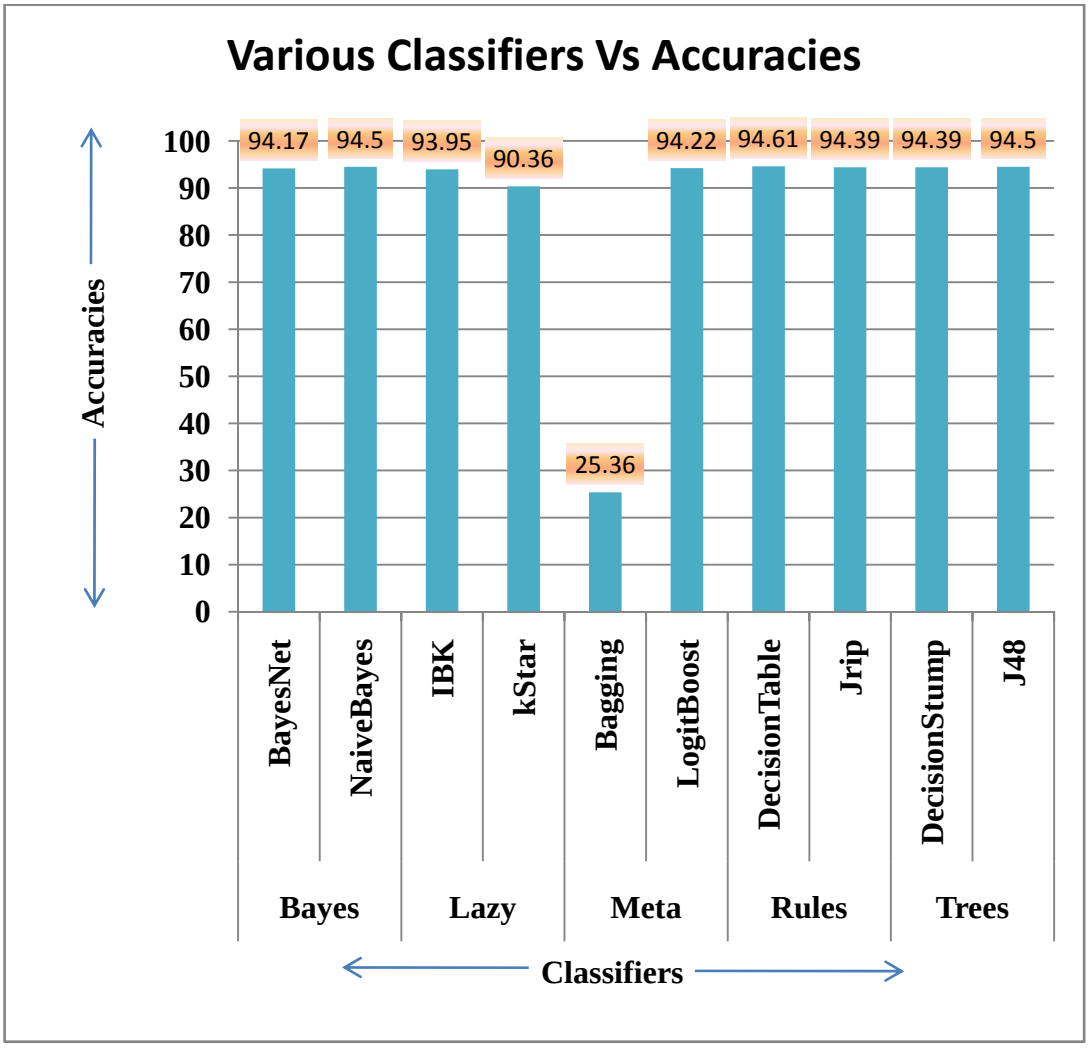

Figure 2 Graphical representations of various classifiers accuracy levels

The accuracies obtained from the selected classifiers are shown in Figure 2.This chart represents the comparison of all the categories of the classifiers. In Bayes classifier method, NaiveBayes has high accuracy when compared with BayesNet Classifier. In Lazy classifier, IBK(K Nearest Neighbor) has high accuracy when compare with KStarClassifier.In Meta classification, Bagging has very low accuracy when compared with LogitBoost Classifier. In Rules classification method, DecisionTable Classifier has high accuracy when compared with JRip Classifier. In Trees classification, J48 has high accuracy when compared with DecisionStump

Figure 2 represents high accuracy and similar tendency with regard to Decision Table classifier from Rules category, NaiveBayes classifier from Bayes category, J48 classifier from Trees category, JRip classifier from Rules category, DecisionStump classifier from Trees category, LogitBoost classifier from Meta category, BayesNet from Bayes category, IBK classifier from Lazy category, KStar classifier from Lazy category. Subsequently, low accuracy was observed under Meta category in Bagging Classifier.qw

V. CONCLUSION

Finally this work concludes that Clustering can help to propose most relevant solutions to customers based on their profiles. Any information that reflects customer traits can become an input to clustering process. In this work, we considered user reviews, feedbacks and rating information captured from forums and social media. Howevertravel managers have diverse opportunities to capture user traits and interests by tracking the types of queries coming to them, taking direct feedback via questionnaires or surveys, keeping track of the user transactions and monitoring the reviews on travel forums and portals. Depending on the data volume and data distribution pattern in consideration, they can adopt appropriate clustering algorithms to segment their customer base so that targeted marketing strategy and/or travel solutions can be offered.

REFERENCES

[1] Renjith, Shini, A. Sreekumar, and M. Jathavedan. 2018. â€œEvaluation of Partitioning Clustering Algorithms for Processing Social Media Data in Tourism Domainâ€ . In 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 127â€“31. IEEE.
[2] Renjith, Shini, and C. Anjali. "A personalized mobile travel recommender system using hybrid algorithm." In Computational Systems and Communications (ICCSC), 2014 First International Conference on, pp. 12-17. IEEE, 2014.
[3] Renjith, Shini, and C. Anjali. "A personalized travel recommender model based on content-based prediction and collaborative recommendation." International Journal of Computer Science and Mobile Computing, ICMIC13 (2013): 66-73.
[4] Estivill-Castro, Vladimir. "Why so many clustering algorithms: a position paper." ACM SIGKDD explorations newsletter 4, no. 1 (2002): 65-75.
[5] MacQueen, James. "Some methods for classification and analysis of multivariate observations." In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, no. 14, pp. 281-297. 1967.
[6] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A kmeans clustering algorithm." Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, no. 1 (1979): 100-108.
[7] Kaufman, Leonard, and Peter Rousseeuw. Clustering by means of medoids. North-Holland, 1987.
[8] Kaufman, Leonard, and Peter J. Rousseeuw. Finding groups in data: an introduction to cluster analysis. Vol. 344. John Wiley & Sons, 2009.
[9] Park, Hae-Sang, and Chi-Hyuck Jun. "A simple and fast algorithm for K-medoids clustering." Expert systems with applications 36, no. 2 (2009): 3336-3341.
[10] Wei, Chih-Ping, Yen-Hsien Lee, and Che-Ming Hsu. "Empirical comparison of fast clustering algorithms for large data sets." In System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on, pp. 10-pp. IEEE, 2000.