University of Sindh Journal of Information and Communication Technology
(USJICT)
Volume 7, Issue 2

ISSN-E: 2523-1235, ISSN-P: 2521-5582
Website: http://sujo.usindh.edu.pk/index.php/USJICT/


# Profiling Reviewer Mobility Behavior Using Yelp Reviews

Muhammad Usman Zafar[1], Muhammad Bilal[2,3,*], Nadia Malik[4]

[1] Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Chiniot-Faisalabad Campus, 35400, Chiniot, Pakistan

[2] Department of Pharmaceutical Outcomes and Policy, Malachowsky Hall for Data Science and Information Technology, University of Florida, Gainesville 32610, Florida, USA

[3] Department of Computing and Information Systems, School of Engineering and Technology, Sunway University, Petaling Jaya 47500, Selangor, Malaysia

[4] Department of Software Engineering, University of Sargodha, Sargodha 40100, Pakistan
muhammadusmanzafar763@gmail.com; muhammad.bilal@ufl.edu; nadiamalik.csit@gmail.com

***Abstract:*** Social media data has become a popular source for uncovering hidden patterns in user mobility. However, the volume of online customer reviews has increased enormously, making it difficult for individuals and businesses to identify and analyze these patterns. Moreover, very few studies exist that explore the mobility patterns of reviewers. Therefore, this study aims to profile the mobility of reviewers to identify hidden patterns. The data of 1217 reviewers and 139,187 reviews from Yelp is used in this study. Firstly, distance-based mobility profiles which include the average distance and total distance of reviewers are created. The correlation analysis showed that distance-based measures strongly correlate with the number of reviews written by a reviewer, the number of cities, and the number of states visited by the reviewer compared to other features. Clustering is done using k-means and Density-based spatial clustering of applications with noise (DBSCAN) to analyze the different nature of reviewers which showed that reviewers can be grouped into two categories that are "less traveler" and "more traveler". A comparison of reviewers who have visited the same number of businesses reveals that their mobility patterns can vary significantly. Finally, the classification of reviewers into less and more travelers is done using various machine learning algorithms. Random forest (RF) achieved Area Under the Curve (AUC) of 0.865, which is comparatively better than other algorithms. The feature importance calculated using RF showed that review count, city, and state are more important features compared to others.

**Keywords:** User Profiling; Mobility Behavior; Online Reviews; Yelp; Clustering; Classification.

## I. INTRODUCTION

Location-based social media data is a type of geotagged data that consists of direct or indirect geology references obtained through social websites or location-based applications [1]. The latest origin of continually growing data has aided in the discovery of links between human mobility and resource consumption [2]. Big data analytics is being used to make crucial decisions, and the interaction of fields, social transformation, and policy measures opens up vast possibilities for city evolution [3, 4]. The smart city idea covers a wide range of urban living domains to which technology and policy interventions may be applied. It can be used to aid in developing smart cities in various fields [5]. Digital data from detected mobility, Geospatial data technology, and online networks may be combined for social interaction and collaborative decisions [6]. Massive quantities of trajectories are being created regularly by numerous sources due to the fast rise of the Internet of Things applications [7]. Mobility patterns are made up of a variety of components and their interactions [8]. A huge number of researchers have been committed to estimating urban movement graphs utilizing GPS data, multilevel and multi-source large geospatial data, smart cards data, data on mobile positioning, and data from other sources. In addition, some researchers have attempted to establish universal rules by analyzing urban/human movement patterns [9].

Various social media platforms, such as Facebook, Twitter, and Instagram, now have billions of users worldwide [10, 11]. The massive data from geotagged online networking websites provides enormous quantities of spatial detail that is important to people's lives [12]. The data created by reviewers provide an opportunity to get details about a variety of locations and events along with details on users' mobility behavior, and social profile attributes [13, 14]. Exploring geography and movement at the level of a social community can provide details about the groups and the members' behavioral graph, social relationships, and social needs [15]. Huge connections, friend relationships, and movement data from online platforms may be used to identify and study different groups. Social Networks provide a massive amount of user-generated data on an individual's activity and mobility. However, only about 1% of published tweets on Twitter are geotagged [16, 17]. Moreover storing, processing, and indexing of the data are real-time challenges, and they affect the mining of the latest mobility patterns. In addition, this massive amount of data is very difficult to analyze, using human cognitive ability. It is a challenging task to profile reviewer mobility behavior using unstructured social media data. The existing studies on mobility are mainly focused on user mobility, however, reviewer mobility based on published reviews still needs to be explored. Therefore, this research aims to propose a reviewer mobility profiling approach to determine reviewer mobility behavior. These mobility profiles are then correlated with other reviewer attributes such as friends count, and the number of cities and states visited by a reviewer. Additionally, the study grouped reviewers based on their mobility profiles using clustering techniques. Furthermore, it compared the performance of various machine learning algorithms in classifying reviewers as either less travelers or more travelers, thereby providing insights into the effectiveness of different algorithms in understanding reviewer mobility behaviors.

## II. Literature Review

This section provides a quick overview of the existing literature on user mobility pattern mining using social media data. Human movement modeling is one of the most essential challenges in understanding human dynamics, and it is also a key component in designing smart city applications. One important challenge of user mobility modeling is predicting a user's upcoming position based on previous mobility traces [18]. Previous studies attempted to analyze user mobility patterns which were using online networking data of location-based in order to track their social geographic behavior. A study attempted to analyze the location-based Weibo data to uncover users' mobility and activity. The Kernel Density Estimations (KDE) algorithm gives a very useful analysis of DBSCAN clustering that assists in the identification of groups depending on point density. By recognizing hot places visited by users, the clusters formed after using both methods enrich the discovery of mobility patterns [19]. Location-based data from social media platforms was used to examine metropolitan human mobility and activity patterns. Average mobility patterns were defined by identifying the proportions of various functional types over city geographic, and hence the intent activity delivery maps. The temporal distribution of visiting different areas based on activity category is then used to define individual activity patterns [20].

The benefit of leveraging user-generated content via social media for mobility demand prediction is investigated by a study. Individual mobility decisions are linked to their activity patterns as obtained from publicly available data in a disaggregated model for mobility demand prediction. A technique for estimating individual activity patterns is also presented, as well as an alternative way for grouping individuals into distinct groups. These functions combine to build a mobility prediction model, which was evaluated using social media data for two test cases [21]. Online data has been used to uncover trends that might help smart city planners make better judgments. Weibo social media data was utilized to follow citizen activity in Shanghai by analyzing social geographic human mobility. The primary purpose of this research was to determine whether geolocated Weibo information can be utilized to discover user mobility and behavior [22]. A study reported the findings of analysis aimed at determining the activity and movement patterns of Instagram users who visited EXPO 2015 organized in Milan, Italy. The authors collected and analyzed geotagged Instagram photographs from around 238,000 users who attended EXPO 2015, and over 570,000 posts were created during the trips, with 2.63 million unique visitors. The analysis revealed how visitor numbers changed over time, which buildings were most often visited, and nationality, as well as the main flows of visitors to Italian cities and regions in the following days of their EXPO trip [8].

A study looked at the mobility patterns of local Twitter users and visiting Twitter users using the flowing networks and constancy distributions of user behavior. The findings suggest that short-distance transportation is the most popular activity among locals and visitors alike. Furthermore, inter-county migration is the most common sort of movement among all Twitter users [23]. A study used mobility data mining and social network analysis approaches to aggregate comparable trajectories, highlighting hotspots of activity and people movements, as well as their fluctuations over time. The evaluation was performed using the trajectory mining algorithms on a big collection of routes derived from geolocated tweets collected for the 2012 Mobile World Congress held in Barcelona [15]. A study analyzed a dataset of 652,945 geotagged tweets created by 2,933 Swedish individuals over more than a year. Human movement predictability has been investigated from three perspectives using this dataset: 1) A chronology of movement spectrum that demonstrates how humans spread in space, 2) Predictability of entropy and mobility, and 3) Spatial boundaries and movement spectrum restrict prediction. The dependency of predictability on mobility range is extended when mobility is observed without regard to geographical boundaries, which reduces the predictability [3].

Local Business Service Systems (LBSSs) such as Yelp, TripAdvisor, and DianPing provide a review service as an important part of Location-Based Social Networks (LBSNs). Users of these systems can post reviews when they visit a Point of Interest (POI) and appreciate the service provided by the POI. This POI can be a business or shop that is listed in an LBSS [24]. In contrast to prior research, which focused on the possibility of identifiability of anonymized information on an individual basis, a study focuses on the potential for identifiability of anonymized records at the aggregate level. The authors used two data sets, each comprising the movement traces of thousands of individuals over one week, representing a reasonably representative group of a metropolitan region [25]. A study attempted to assess the Yelp cascades in respect of their length or briefness being predictable. While doing so, researchers tried to pinpoint the crucial characteristics of such cascades as well as the networks' characteristics along with the company associated with them cascades. However, the overall prediction accuracy of these algorithms varies [26].

Urbanization and check-in behavior in both time and place are crucial elements in the design and construction of urban sustainability, according to urban planners. Using location-based social network (LBSN) data to analyze check-in behavior by which public considerations appear to supplement traditional approaches. The spatial data introduces extra perspectives to the research of check-in habits and aids in the development of fresh methods and frameworks for LBSN data processing. A study examined the possibility of employing LBSN information to introduce a new perspective to study objective level check-in behavior. It was reported that women use it at high levels and that holiday and workday check-in patterns varied throughout all Shanghai neighborhoods. It may also be useful to track fluctuations in population concentration through time or can be used to predict the availability of services in the area [27].

To advance urban transportation through innovative solutions and research, understanding urban travel behavior is crucial. However, traditional methods of doing this are expensive and prone to inaccuracies. The explosion of urban big data offers an alternative way for data collection and mining tools. Human mobility influences many spatial and temporal events, including habitation, transportation, goods movement, and disease transmission. Over the past 50 years, the study of human mobility and its intersections with various fields has progressed significantly. Urban big data captures multiple aspects of cities, providing extensive information on human interactions with urban environments over time and space. This data allows researchers to study individuals, analyze driving habit variations, and understand travel patterns within and between communities [28].

## III. MATERIALS AND METHODS

The overall research methodology and flow of steps are shown in Fig. 1. A quick overview of each step is provided in this section.

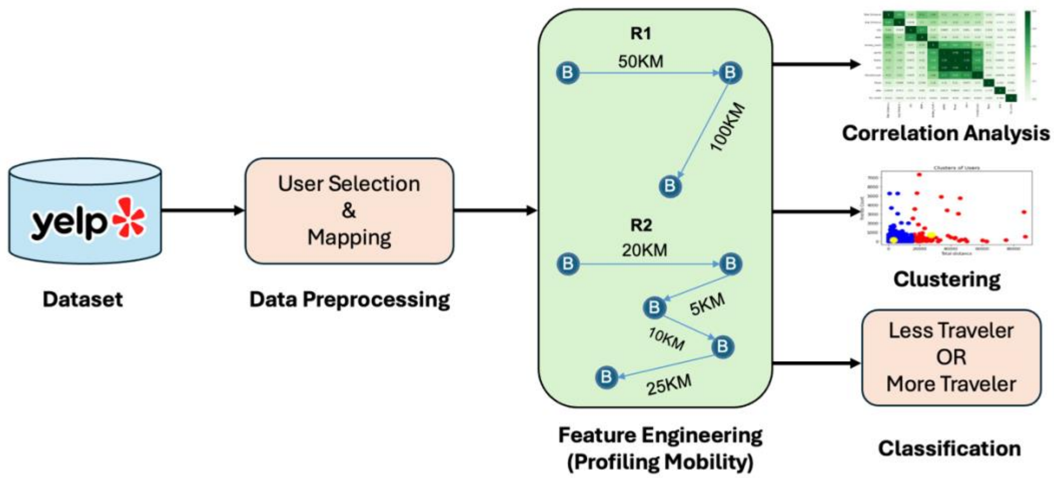


Figure 1. Proposed modified CNN architecture.

### A. Datasets Acquisition

The data from Yelp is used in this study. Since 2013, Yelp has provided access to its datasets for the Yelp Dataset Challenge. Our findings in this study are centered on the Yelp Open Dataset [29] that was obtained on March 5, 2022. Yelp dataset is a very rich dataset that provides information related to reviews, users, businesses, check-ins, tips, and photos. This dataset contains 6,990,280 reviews written by 1,987,897 users for 150,346 businesses that are located in 11 metropolitan areas. It also provides details on 200,100 pictures that are uploaded with the reviews. Furthermore, it contains data for 131,930 check-ins and 908,915 tips for different businesses.

### B. Data Preprocessing

In this study, we gather user data from the User dataset available in the Yelp Dataset. After getting the user ID we fetch their review counts from the Review table. We then make buckets of 21-25, 46-50, 71-75, 85-90,95-100, 140-150, 190-200, and 251-350 based on review counts. After making buckets of review count, we map the user ID on the review dataset and fetch all the reviews from the dataset. Then we fetch the business ID from the business dataset based on the reviews we got from the review dataset. We also get other attributes counted from the business table which are latitude, longitude, cool, funny, state, city, and date. After that, we sorted the reviews for each user with respect to date and time. We get the review count of all users into user data, and we divide them into buckets of specific numbers. Then we get 150 random users from all buckets. After that, we extract the review IDs from the review table and the date and time of the review given to them. We also extract business IDs from the review table and add them to one location. After getting the business ID, we get the latitude and longitude of that business from the business table and city and state attributes.

### C. Feature Engineering

In this study, we select data attributes from three datasets that include User, Review, and Business. Studying user attitudes may be crucial for managing a company or website since it enables you to foresee customers' activities and put people into categories. Since analyzing user movement patterns is the main objective of the analysis, every item in the collection should include location data, such as latitude and longitude. As a result, the first step in the preparation process is to filter out or fill in any data that has missing values for these variables. For user profiling we calculate the total distance as well as the average distance of the specific user, we calculate the distance from the latitude and longitude of the user's business they visited by sorting them by date and time. We also count the city and state how many states and cities they visited and how much total distance they covered and by visiting every location how many friends increased. The pseudo-code of the reviewer mobility profiling algorithm proposed by this study is given in Fig. 2.

```
Algorithm 1:
1: Inputs:  User Data set
2: Inputs:  Review Data set
3: Inputs:  Business Data set
4:   Randomly sample user IDs
5:   Using the review ID
6:  Map the business ID from the business data set
7: Fetch:  latitude and longitude of the business
8: Calculate Average latitude and longitude
9: Add a dictionary containing the user ID
10: Repeat steps 2-7 for all the user Data set
11: Calculate Distance between the business location and
    the user's average location
12: Outputs:  Average and Total Distance
13: return
14: End
```

Figure 2. Proposed algorithm for calculating distance traveled.

### D. *Correlation, Clustering, and Classification*

The Correlation analysis of the features all the features is performed in this study. In our study, we used unsupervised learning because our data is unlabeled, and we have a clustering problem. We cannot go for classification directly because of unlabeled data. First, we labeled our data through clustering and then we went for classification. For this, we used clustering methods. There are many clustering methods, however, we used K mean and DBSCAN for this study. In k-means clustering, the ideal number of clusters is established using the elbow approach. By using various choices of k, the elbow approach plots the amount of the objective functions. The deviation here between generated clusters will be investigated using silhouette analysis. After analyzing the clusters, we use the cluster information to label the reviewer. After labeling the classification is performed using various machine learning algorithms and 10-fold cross-validation is for evaluation. The evaluation metrics used for classification include accuracy, precision, recall, F1-score, and AUC.

## IV. RESULTS AND DISCUSSION

We have collected 1217 random user IDs from the user's dataset based on their review count. A sample of the dataset used in this study is provided in Table 1. Labels of the attributes used for further discussion and used in correlation analysis are also provided in this table. of the given in the table are also provided.

The results of the correlation analysis are shown in Fig. 3. It can be seen from the results that distance-based measures strongly correlate with the number of reviewers written by a reviewer, the number of cities, and the number of states visited by the reviewer compared to other features. The review count is also seen as highly correlated with the review reactions such as funny, cool, useful, etc. Moreover, the number of friends and useful are also highly correlated, which reveals that having a higher number of friends leads to more useful votes.

The best number of clusters for the data is then determined to be the number of clusters that correspond to the elbow point. The optimal number of clusters based on the elbow method is 2. The density and dispersion of a cluster are measured by the silhouette coefficient. Its values vary from -1 to 1, with 1 denoting a tight and well-separated cluster, 0 denoting a cluster that is neither compact nor well-separated, and -1 denoting a cluster that is badly compact and separated. We can clearly see in Fig. 4 that based on the silhouette coefficient two clusters are suitable for this study.

TABLE 1. SAMPLE OF DATASET.

| ID | Total Distance | Average Distance | City | State | Review Count | Useful | Funny | Cool | Friends Count | Days | Elite Count | Actual Review Count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Label | Total Distance | Avg Distance | city | state | Review_count | useful | funny | cool | freindsCount | days | elite | rev_count |
| 1 | 636.4137 | 9.091624 | 6 | 1 | 120 | 131 | 30 | 32 | 45 | 3622 | 5 | 20 |
| 2 | 169.4058 | 7.700264 | 13 | 1 | 28 | 41 | 17 | 13 | 0 | 4684 | 0 | 20 |
| 3 | 2163.301 | 10.01528 | 19 | 1 | 461 | 325 | 75 | 130 | 26 | 3448 | 2 | 20 |
| 4 | 2085.931 | 9.568491 | 2 | 1 | 577 | 1358 | 416 | 683 | 764 | 4493 | 10 | 20 |


Figure 3. Correlation of features used in this study.

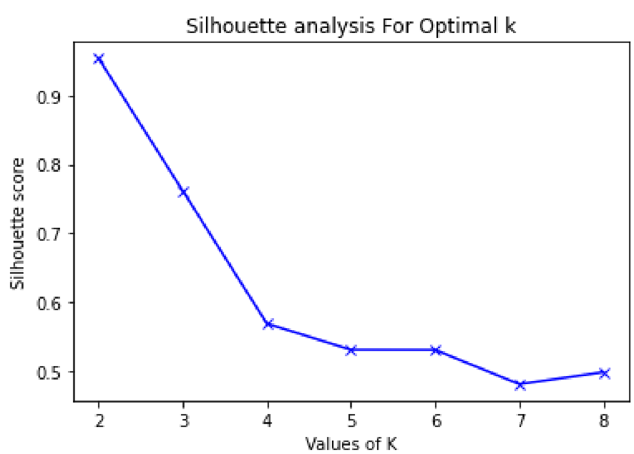


Figure 4. Optimal number of K.

We have performed clustering one by one for all features with average distance and total distance using both K-mean and DBSCAN. The clustering of average distance using K-mean with elite, cool, funny, friends count, and review count did not show that these features are not good for clustering based on average distance. In contrast, the city, state, and actual review count showed a clear separation of clusters. Two clear clusters can be seen for all features based on total distance using K-mean. A sample of clustering results using k-mean is shown in Fig. 5 and Fig. 6. Similarly, a sample of clustering results using DBSCAN is shown in Fig. 7 and Fig. 8.

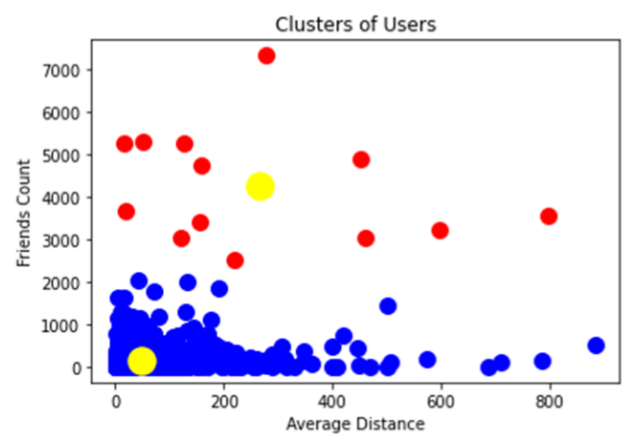


Figure. 5. K-mean cluster Friends Count with Average Distance.

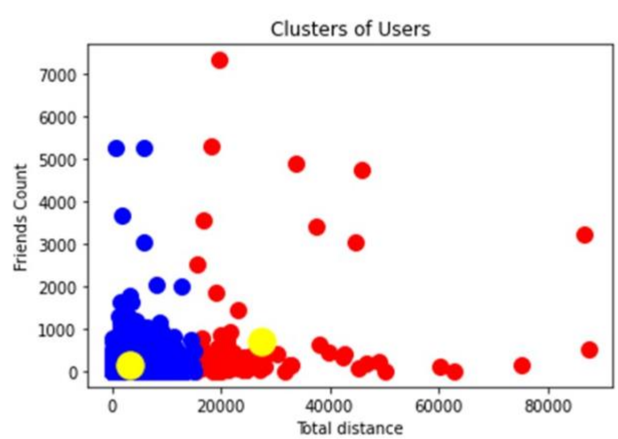

Figure. 6. K-mean cluster Friends Count with Total Distance.

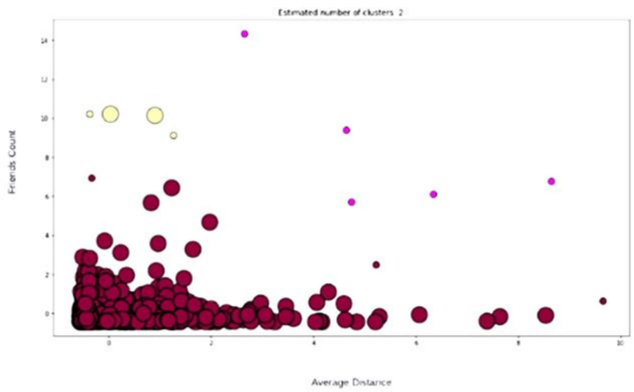


Figure. 7. DBSCAN cluster Friends Count with Average Distance

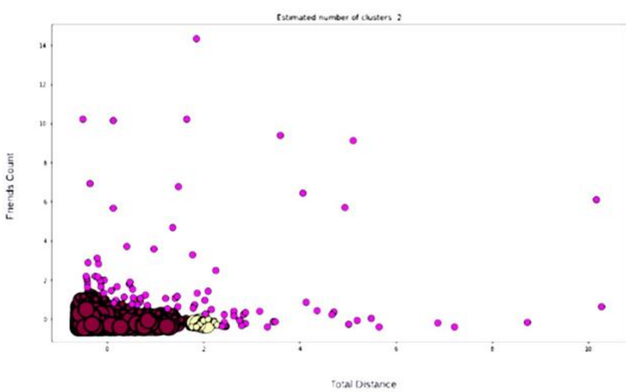


Figure. 8. DBSCAN cluster Friends Count with Total Distance

This study also provides findings by comparing two reviewers' profiles that have reviewed the same number of businesses. It is interesting to note that despite both reviewers visiting 48 businesses, the total distance traveled by them significantly varies. Fig. 9 and Fig. 10 provide a comparison of both reviewers.

The reviewers are labeled as "less travelers" and "more travelers" in this study. Those who have traveled less than 50 kilometers are labeled "less travelers" and those who have traveled more than 50 kilometers are labeled "more travelers". The machine learning algorithms used in this study include logistic regression (LGR), support vector machines (SVM), decision trees (DT), RF, multi-layer perceptron (MLP), and gradient boosting trees (GBT), and neural networks (NN). The evaluation results showed that NN achieved the highest accuracy at 79.63%, closely followed by MLP which achieved an accuracy of 78.89%. RF showed the best AUC score of 0.865, followed closely by NN with an AUC of 0.861. GBT demonstrated the highest precision at 87.35%, while DT achieved the highest recall of 99.00%. NN also achieved the highest F1 score of 86.97%, followed closely by LGR with F1 score of 86.38%. The classification results are given in Table 2.

Table 3 provides the feature weights calculated using RF that reflects the importance of features in predicting the reviewer as less or more traveler. "Review Count" emerges as the most important attribute followed by "City" and "State" which highlight the importance of geographical features, suggesting that regional factors significantly influence user mobility behavior. Metrics related to review interactions, such as "Useful", "Funny" and "Cool" also show moderate importance. While the remaining features appeared as less important for the classification of reviewers.

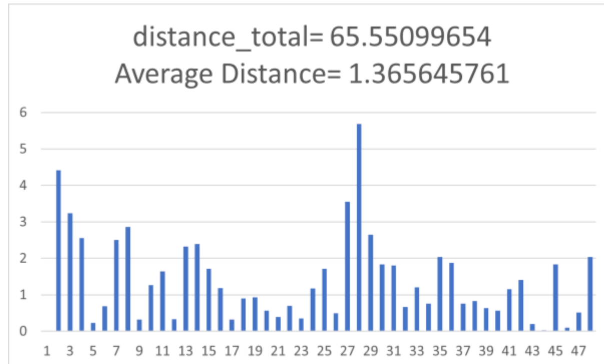


Figure. 9. Travel history of reviewer 1 with less distance covered.

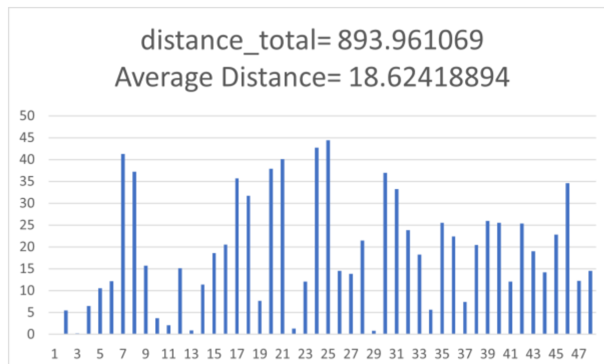


Figure. 10. Travel history of reviewer 2 with more distance covered.

TABLE 2. CLASSIFICATION RESULTS.

| Algo | Accuracy | AUC | Precision | Recall | F1 measure |
|---|---|---|---|---|---|
| LGR | 78.23% | 0.812 | 80.08% | 93.85% | 86.38% |
| DT | 74.53% | 0.777 | 74.63% | 99.00% | 85.10% |
| RF | 77.82% | **0.865** | 81.96% | 89.71% | 85.62% |
| GBT | 76.34% | 0.857 | 87.35% | 79.31% | 83.09% |
| SVM | 77.57% | 0.819 | 78.22% | 96.42% | 86.35% |
| NN | 79.63% | 0.861 | 82.15% | 92.50% | 86.97% |
| MLP | 78.89% | 0.819 | 81.53% | 92.38% | 86.53% |

TABLE 3. FEATURE WEIGHTS.

| Attribute | Weight |
|---|---|
| Review Count | 0.201 |
| City | 0.139 |
| State | 0.120 |
| Useful | 0.109 |
| Funny | 0.093 |
| Cool | 0.080 |
| Friends Count | 0.074 |
| Days | 0.072 |
| Actual Review Count | 0.070 |
| Elite | 0.040 |

## V. CONCLUSION

This study profiles the reviewer traveling behavior of 1217 reviewers using their Yelp reviews. For each selected reviewer, the reviews written by them are collected and stored. The latitude and longitude of each business that the reviewer reviewed is mapped against each review which is further used to calculate the total and average distance traveled by reviewers. The mobility profile for each reviewer is generated based on the distance traveled by them for reviewing different businesses. These reviewer mobility profiles help us to understand the mobility behavior of the reviewers in terms of their mobility and the number of businesses they have reviewed. Correlation analysis is performed to see how reviewer mobility features (average distance and total distance) are related to other features such as review count, number of cities, number of states, number of friends, etc. Correlation analysis shows that review count, city, and state are highly correlated with average distance and total distance compared to other features. K-Means and DBSCAN are used to perform unsupervised classification of reviewers. For the K-Means clustering elbow method and silhouette analysis are used to identify the optimal numbers of K, which reported that two clusters are optimal. DBSCAN results showed that the number of clusters varies from one to three. The comparison of different users is also performed based on the number of businesses they reviewed. It is seen that the reviewers who have visited the same number of businesses can have a significantly different mobility profile. The performance of various machine learning algorithms such as LGR, SVM, NN, DT, RF, GBT, and MLP for the classification of reviewers is also studied. The reviewers with an average distance less than the average distance of all the reviewers (50 km) are labeled as less travelers and the remaining are labeled as more travelers. The classification results show that RF performed better in comparison to the remaining algorithms based on AUC. The feature weights show that review count, city, and state are more important features than others.

There are a few limitations that are attached to the work done in this study. Firstly, the reviewer mobility profiles are generated using data from Yelp. The future work will try to use data from other review platforms and analyze similarities and differences in reviewer mobility patterns on different review platforms. Second, threshold-based labeling of reviewers as less travelers and more travelers is done. Future work will consider increasing the number of reviewers and using clustering labels for the classification task. Third, the work done in this study did not take into consideration business categories and textural features. These missing features such as business category and textual features will be considered in future work that can provide more insights into reviewer mobility behavior.

## REFERENCES


1. Bao, J., et al., Recommendations in location-based social networks: a survey. GeoInformatica, 2015. 19: p. 525-565.
2. Yuan, Y. and M. Raubal, Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. International Journal of Geographical Information Science, 2016. 30(8): p. 1594-1621.
3. Hilbert, M., Big data for development: A review of promises and challenges. Development Policy Review, 2016. 34(1): p. 135-174.
4. Bilal, M., et al., Exploring industrial demand trend's in Pakistan software industry using online job portal data. University of Sindh Journal of Information and Communication Technology, 2017. 1(1): p. 17-24.
5. Bilal, M., et al., Smart cities data: Framework, applications, and challenges. Handbook of smart cities, 2020: p. 1-29.
6. Miller, H.J., Beyond sharing: cultivating cooperative transportation systems through geographic information science. Journal of Transport Geography, 2013. 31: p. 296-308.
7. Jiang, R., et al. Deepurbanmomentum: An online deep-learning system for short-term urban mobility prediction. in Proceedings of the AAAI conference on artificial intelligence. 2018.
8. Cascetta, E., F. Pagliara, and A. Papola, Governance of urban mobility: complex systems and integrated policies. Advances in complex systems, 2007. 10(supp02): p. 339-354.
9. Gallotti, R., et al., A stochastic model of randomly accelerated walkers for human mobility. Nature communications, 2016. 7(1): p. 12600.
10. Bilal, M., et al. Predicting helpfulness of crowd-sourced reviews: A survey. in 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS). 2019. IEEE.
11. Bilal, M., et al., Profiling and predicting the cumulative helpfulness (Quality) of crowd-sourced reviews. Information, 2019. 10(10): p. 295.
12. Wang, Y., et al., Using social media for emergency response and urban sustainability: A case study of the 2012 Beijing rainstorm. Sustainability, 2015. 8(1): p. 25.
13. Liu, Z., et al., Urban traffic prediction from mobility data using deep learning. Ieee network, 2018. 32(4): p. 40-46.
14. Bilal, M., et al., Social profiling: A review, taxonomy, and challenges. Cyberpsychology, Behavior, and Social Networking, 2019. 22(7): p. 433-450.
15. Gao, S., Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. Spatial Cognition & Computation, 2015. 15(2): p. 86-114.
16. Manca, M., et al., Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. Online social networks and media, 2017. 1: p. 56-69.
17. Tiwana, M.S., et al. Comparative analysis of context based classification of twitter. in 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA). 2018. IEEE.
18. Yang, D., et al. Location prediction over sparse user mobility traces using rnns. in Proceedings of the twenty-ninth international joint conference on artificial intelligence. 2020.
19. McKercher, B. and A.A. Lew, Tourist flows and the spatial distribution of tourists. A companion to tourism, 2004: p. 36-48.
20. Lew, A.A. and B. McKercher, Trip destinations, gateways and itineraries: The example of Hong Kong. Tourism Management, 2002. 23(6): p. 609-621.
21. Adriansen, H.K. and T.T. Nielsen, The geography of pastoral mobility: A spatio-temporal analysis of GPS data from Sahelian Senegal. GeoJournal, 2005. 64: p. 177-188.
22. Liu, L., et al. Understanding individual and collective mobility patterns from smart card records: A case study in Shenzhen. in 2009 12th International IEEE Conference on Intelligent Transportation Systems. 2009. IEEE.
23. Liu, J., et al., Rethinking big data: A review on the data quality and usage issues. ISPRS journal of photogrammetry and remote sensing, 2016. 115: p. 134-142.
24. Wang, Y., et al., Exchanging registered users' submitting reviews towards trajectory privacy preservation for review services in Location-Based Social Networks. Plos one, 2021. 16(9): p. e0256892.
25. Kondor, D., et al., Towards matching user mobility traces in large-scale datasets. IEEE Transactions on Big Data, 2018. 6(4): p. 714-726.

26. Khan, M.R., Cascading behavior in yelp reviews. arXiv preprint arXiv:1712.00903, 2017.
27. Rizwan, M., et al., Using location-based social media data to observe check-in behavior and gender difference: Bringing weibo data into play. ISPRS International Journal of Geo-Information, 2018. 7(5): p. 196.
28. Wang, C. and D.B. Hess, Role of urban big data in travel behavior research. Transportation research record, 2021. 2675(4): p. 222-233.
29. Yelp. Yelp Open Dataset. March 05, 2022]; Available from: https://www.yelp.com/dataset.