PRIVACY PRESERVING LOCATION DATA PUBLISHING - A MACHINE LEARNING APPROACH

S. Venkata Ramana[1], M. Kokila Reddy[2], N. Vaishnavi[2], M. Srija[2], M. Navya Reddy[2]

[1]Assistant Professor, [2]UG Scholar, [1,2]Department of Computer Science and Engineering

[1,2]Malla Reddy Engineering College for Women (A), Maisammaguda, Medchal, Telangana.

kokilareddymamidala2002@gmail.com, sri231900@gmail.com, srijamurthygari@gmail.com, nandalavaishnavi@gmail.com

ABSTRACT

Publishing datasets plays an essential role in open data research and promoting transparency of government agencies. However, such data publication might reveal users' private information. One of the most sensitive sources of data is spatiotemporal trajectory datasets. Unfortunately, merely removing unique identifiers cannot preserve the privacy of users. Adversaries may know parts of the trajectories or be able to link the published dataset to other sources for the purpose of user identification. Therefore, it is crucial to apply privacy preserving techniques before the publication of spatiotemporal trajectory datasets. In this paper, we propose a robust framework for the anonymization of spatiotemporal trajectory datasets termed as machine learning based anonymization (MLA). By introducing a new formulation of the problem, we are able to apply machine learning algorithms for clustering the trajectories and propose to use k-means algorithm for this purpose. A variation of k-means algorithm is also proposed to preserve the privacy in overly sensitive datasets. Moreover, we improve the alignment process by considering multiple sequence alignment as part of the MLA. The framework and all the proposed algorithms are applied to T-Drive, Geolife, and Gowalla location datasets. The experimental results indicate a significantly higher utility of datasets by anonymization based on MLA framework.

Keywords: Publishing of data, Machine learning, Privacy preserving.

1. INTRODUCTION

Publication of data by different organizations and institutes is crucial for open research and transparency of government agencies. Just in Australia, since 2013, over 7000 additional datasets have been published on 'data.gov.au,' a dedicated website for the publication of datasets by the Australian government. Moreover, the new Australian government data sharing legislation encourage government agencies to publish their data, and as early as 2019, many of them will have to do so [2]. Unfortunately, the process of data publication can be highly risky as it may disclose individuals' sensitive information. Hence, an essential step before publishing datasets is to remove any uniquely identifiable information from them. However, such an operation is not sufficient for preserving the privacy of users. Adversaries can re-identify individuals in datasets based on common attributes called quasi-identifiers or may have prior knowledge about the trajectories traveled by the users. Such side information enables them to reveal sensitive information that can cause physical, financial, and reputational harms to people.

One of the most sensitive sources of data is location trajectories or spatiotemporal trajectories. Despite numerous use cases that the publication of spatiotemporal data can provide to users and researchers, it poses a significant threat to users' privacy. As an example, consider a person who has been using GPS navigation to travel from home to work every morning of weekdays. If an adversary has some prior knowledge about a user, such as the home address, it is possible to identify the user. Such an inference attack can compromise user privacy, such as revealing the user's health condition and how often the user visits his/her medical specialist. Therefore, it is crucial to anonymize spatiotemporal datasets before publishing them to the public. The privacy issue gets even more severe if the adversary links identified users to other databases, such as the database of medical records. That is the very reason why nowadays most companies are reluctant to publish any spatiotemporal trajectory datasets without applying an effective privacy preserving technique.

A widely accepted privacy metric for the publication of spatiotemporal datasets is k-anonymity. This metric can be summarized as ensuring that every trajectory in the published dataset is indistinguishable from at least k − 1 other trajectories. The authors in [3], adopted the notion of k-anonymity for spatiotemporal datasets and proposed an anonymization algorithm based on generalization. Xu et al. [4] investigated the effects of factors such as spatiotemporal resolution and the number of users released on the anonymization process. Dong et al. [5] focused on improving the existing clustering approaches. They proposed an anonymization scheme based on achieving kanonymity by grouping similar trajectories and removing the highly dissimilar ones. More recently, the authors in [6] developed an algorithm called k-merge to anonymize the trajectory datasets while preserving the privacy of users from probabilistic attacks. Local suppression and splitting.

Lack of a well-defined method to cluster trajectories as there is not an easy way to measure the cost of clustering when considering the distances among trajectories rather than simply the locations. • The existing literature focuses on pairwise sequence alignment, which results in a high amount of information loss [3], [6], [8]–[10]. • There is no unified metric to evaluate and compare the existing anonymization methods.

In this paper, we address the mentioned problems by proposing an enhanced anonymization framework termed machine learning based anonymization (MLA) to preserve the privacy of users in the publication of spatiotemporal trajectory datasets. MLA consists of two interworking algorithms: clustering and alignment. We have summarized our main contributions in the following bullet points.

By formulating the anonymization process as an optimization problem and finding an alternative representation of the system, we are able to apply machine clustering algorithms for clustering trajectories. We propose to use k 0 -means 1 algorithm for this purpose, as part of the MLA framework.

- We propose a variation of k 0 -means algorithm to preserve the privacy of users in the publication of overly sensitive spatiotemporal trajectory datasets. • We enhance the performance of sequence alignment in clusters by considering multiple sequence alignment instead of pairwise sequence alignment.
- We propose a utility metric to evaluate and compare the anonymization frameworks.

MLA and all algorithms associated with it are applied on two real-life GPS datasets following different distributions in time and spatial domains. The experimental results indicate a significantly higher utility levels while maintaining k-anonymity of trajectories.

2. PROPOSED SYSTEM

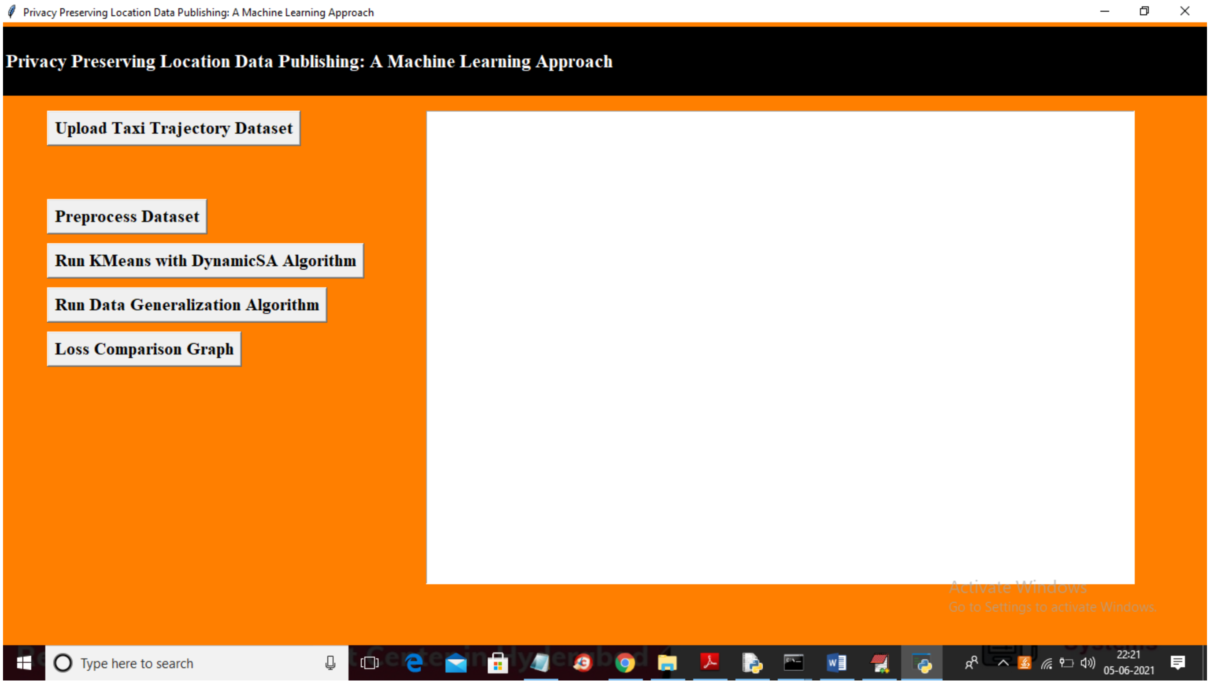
But above techniques are not reliable as malicious users can identify how to crack groups and noise data to know user location. To overcome from this problem author has introduce Machine Learning based data privacy preserving technique which consists of 3 models and this 3 models will provide more security and anonymize or generalized which cannot be easily understand or crack.

1. Clustering model: in this model user locations will be clusters by using KMEANS algorithm and then calculate loss value. Loss value indicates difference between correct value and predicted value and the lesser the loss the better is the algorithm. The loss value will be saved to compare

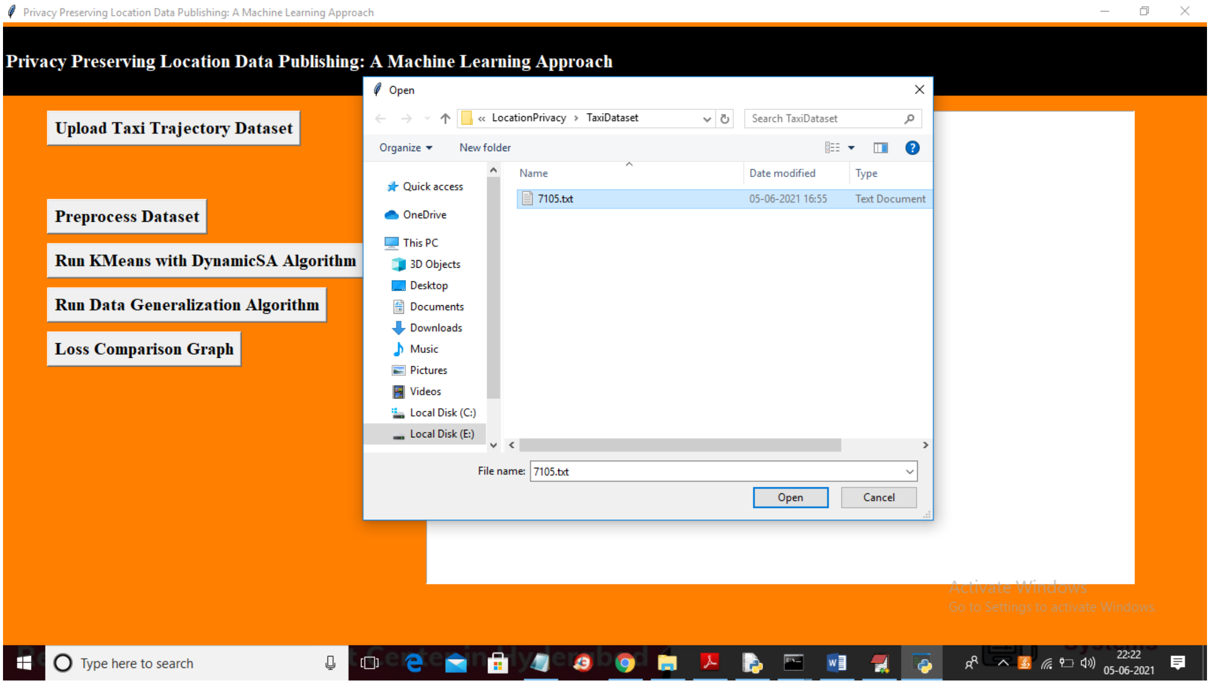
with Dynamic Sequence Alignment Loss and this Dynamic Sequence is called as Heuristic Clustering Algorithm.

2. Dynamic Sequence Alignment: In this module or algorithm we will take location form cluster member and then take random locations from original dataset and both this records will be aligned to get location which has minimal loss.
3. Data Generalization: in this module user location will be generalized or anonymised by summing up location with loss values.

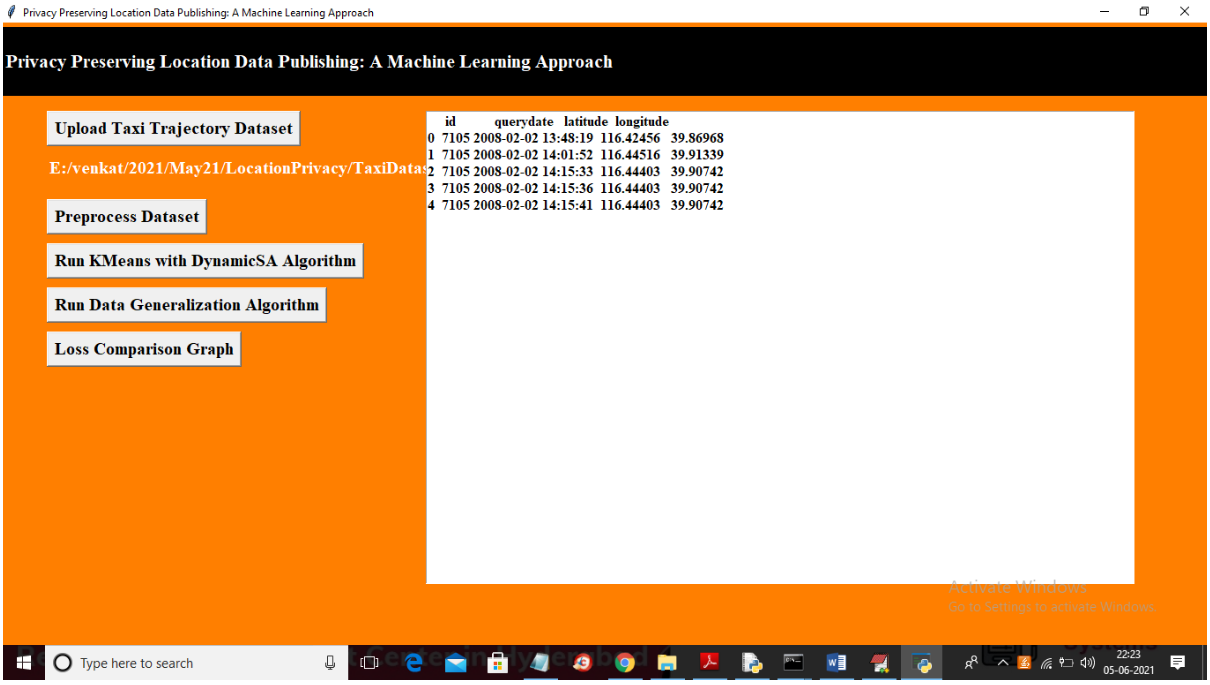
3. RESULTS


Privacy Preserving Location Data Publishing: A Machine Learning Approach
Privacy Preserving Location Data Publishing: A Machine Learning Approach
Upload Taxi Trajectory Dataset
Preprocess Dataset
Run KMeans with DynamicSA Algorithm
Run Data Generalization Algorithm
Loss Comparison Graph
Type here to search
22:21
05-06-2021

In above screen click on ‘Upload Taxi Trajectory Dataset’ button to upload dataset

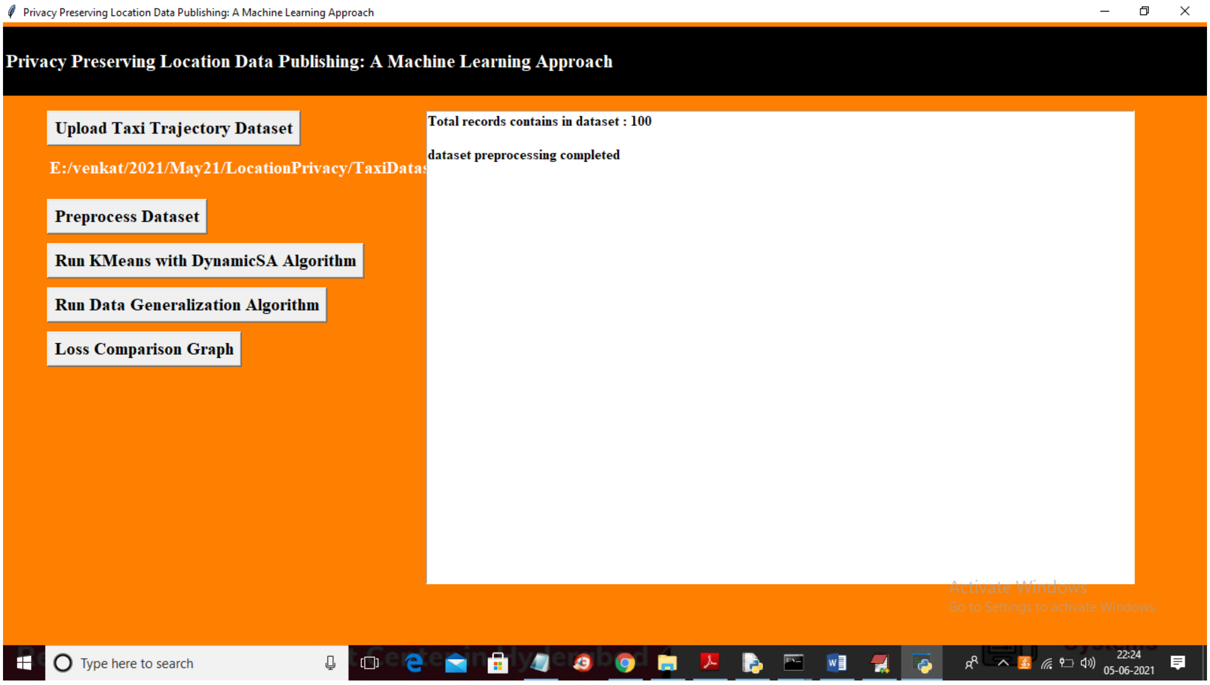


In above screen selecting and uploading taxi trajectory file and then click on ‘Open’ button to load dataset and to get below screen


Privacy Preserving Location Data Publishing: A Machine Learning Approach
Privacy Preserving Location Data Publishing: A Machine Learning Approach
Upload Taxi Trajectory Dataset
E:/venkat/2021/May21/LocationPrivacy/TaxiDatas
Preprocess Dataset
Run KMeans with DynamicSA Algorithm
Run Data Generalization Algorithm
Loss Comparison Graph
Type here to search
22:23
05-06-2021

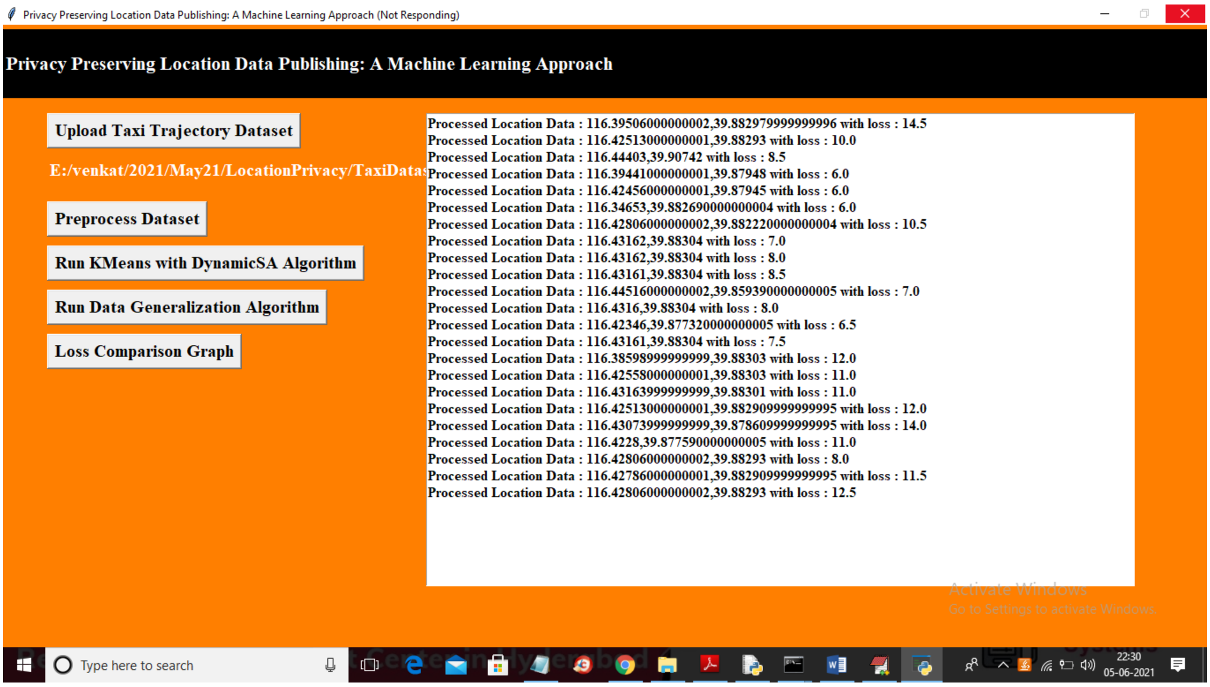
```
     id          querydate   latitude  longitude
0  7105 2008-02-02 13:48:19  116.42456   39.86968
1  7105 2008-02-02 14:01:52  116.44516   39.91339
2  7105 2008-02-02 14:15:33  116.44403   39.90742
3  7105 2008-02-02 14:15:36  116.44403   39.90742
4  7105 2008-02-02 14:15:41  116.44403   39.90742
```

In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to remove empty values and then extract latitude and longitude location from above dataset


Privacy Preserving Location Data Publishing: A Machine Learning Approach
Privacy Preserving Location Data Publishing: A Machine Learning Approach
Upload Taxi Trajectory Dataset
E:/venkat/2021/May21/LocationPrivacy/TaxiDatas
Preprocess Dataset
Run KMeans with DynamicSA Algorithm
Run Data Generalization Algorithm
Loss Comparison Graph
Type here to search
22:24
05-06-2021

```
Total records contains in dataset : 100

dataset preprocessing completed
```

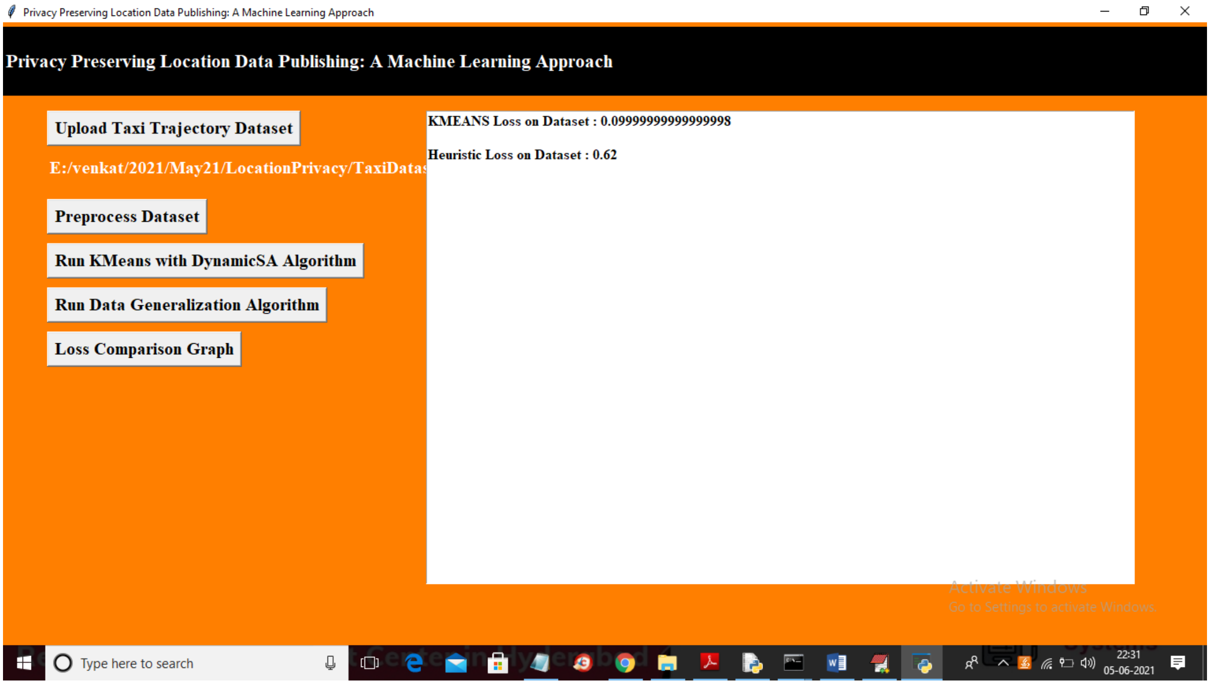
In above screen dataset preprocessing completed and now click on ‘Run KMeans with DynamicSA Algorithm’ button to run KMEANS on dataset with Dynamic SA. This algorithm will group all similar location into same cluster and then perform DYNAMIC SA.


Privacy Preserving Location Data Publishing: A Machine Learning Approach (Not Responding)
Privacy Preserving Location Data Publishing: A Machine Learning Approach
Upload Taxi Trajectory Dataset
Preprocess Dataset
Run KMeans with DynamicSA Algorithm
Run Data Generalization Algorithm
Loss Comparison Graph
Type here to search
22:30
05-06-2021

E:/venkat/2021/May21/LocationPrivacy/TaxiDatas

```
Processed Location Data : 116.39506000000002,39.882979999999996 with loss : 14.5
Processed Location Data : 116.42513000000001,39.88293 with loss : 10.0
Processed Location Data : 116.44403,39.90742 with loss : 8.5
Processed Location Data : 116.39441000000001,39.87948 with loss : 6.0
Processed Location Data : 116.42456000000001,39.87945 with loss : 6.0
Processed Location Data : 116.34653,39.882690000000004 with loss : 6.0
Processed Location Data : 116.42806000000002,39.882220000000004 with loss : 10.5
Processed Location Data : 116.43162,39.88304 with loss : 7.0
Processed Location Data : 116.43162,39.88304 with loss : 8.0
Processed Location Data : 116.43161,39.88304 with loss : 8.5
Processed Location Data : 116.44516000000002,39.859390000000005 with loss : 7.0
Processed Location Data : 116.4316,39.88304 with loss : 8.0
Processed Location Data : 116.42346,39.877320000000005 with loss : 6.5
Processed Location Data : 116.43161,39.88304 with loss : 7.5
Processed Location Data : 116.38598999999999,39.88303 with loss : 12.0
Processed Location Data : 116.42558000000001,39.88303 with loss : 11.0
Processed Location Data : 116.43163999999999,39.88301 with loss : 11.0
Processed Location Data : 116.42513000000001,39.882909999999995 with loss : 12.0
Processed Location Data : 116.43073999999999,39.878609999999995 with loss : 14.0
Processed Location Data : 116.4228,39.877590000000005 with loss : 11.0
Processed Location Data : 116.42806000000002,39.88293 with loss : 8.0
Processed Location Data : 116.42786000000001,39.882909999999995 with loss : 11.5
Processed Location Data : 116.42806000000002,39.88293 with loss : 12.5
```

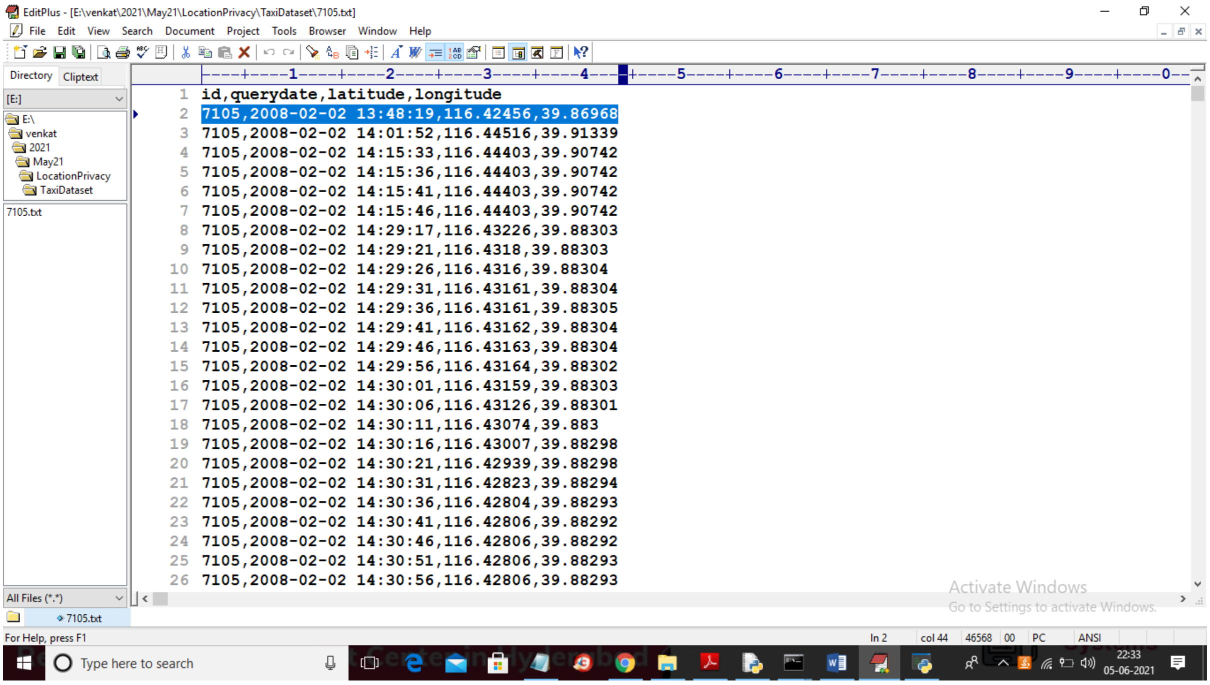
In above screen each location is processed and then calculating loss value with dynamic sequence alignment which align two locations by choosing minimal loss location.


Privacy Preserving Location Data Publishing: A Machine Learning Approach
Privacy Preserving Location Data Publishing: A Machine Learning Approach
Upload Taxi Trajectory Dataset
Preprocess Dataset
Run KMeans with DynamicSA Algorithm
Run Data Generalization Algorithm
Loss Comparison Graph
Type here to search
22:31
05-06-2021

E:/venkat/2021/May21/LocationPrivacy/TaxiDatas

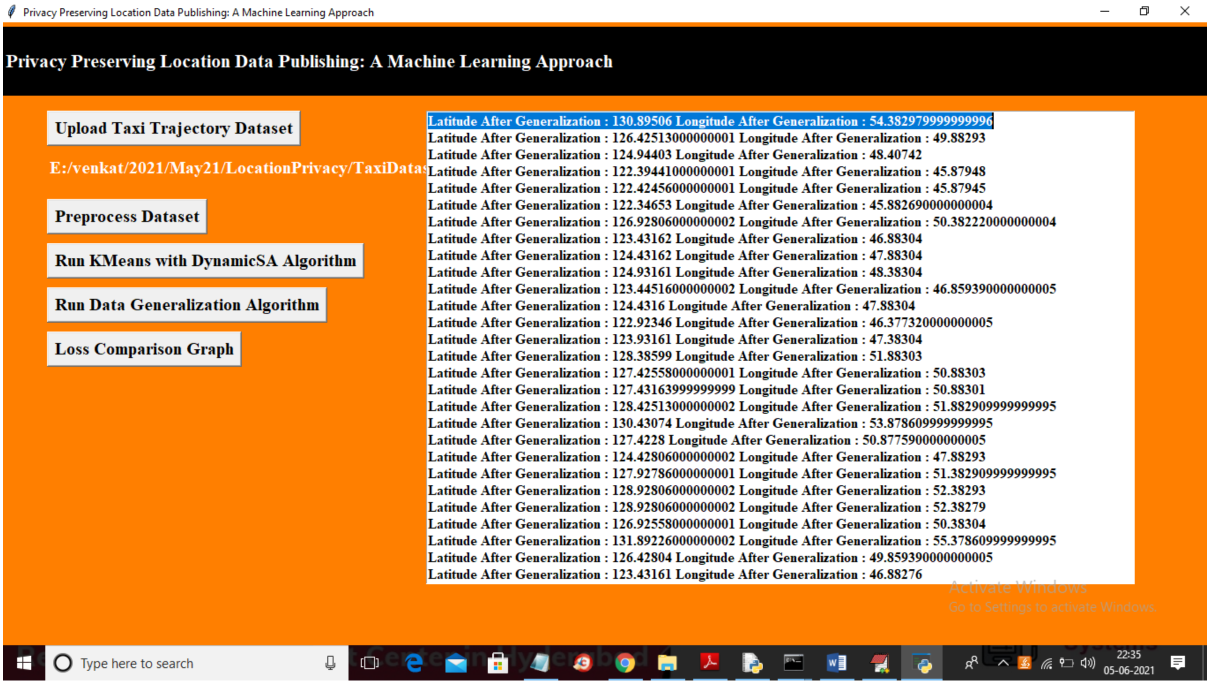
```
KMEANS Loss on Dataset : 0.09999999999999998

Heuristic Loss on Dataset : 0.62
```

In above screen KMEANS loss is 0.09 and Heuristic Clustering (also known as Dynamic SA) loss is 0.62. Now click on ‘Run Data Generalization Algorithm’ button to generalized data with loss value. In below screen in first record you can see real location values from dataset and in next screen same location was generalized or anonymised with above algorithms


EditPlus - [E:\venkat\2021\May21\LocationPrivacy\TaxiDataset\7105.txt]
File Edit View Search Document Project Tools Browser Window Help
Directory Cliptext
7105.txt
All Files (*.*)
For Help, press F1
Type here to search
22:33
05-06-2021

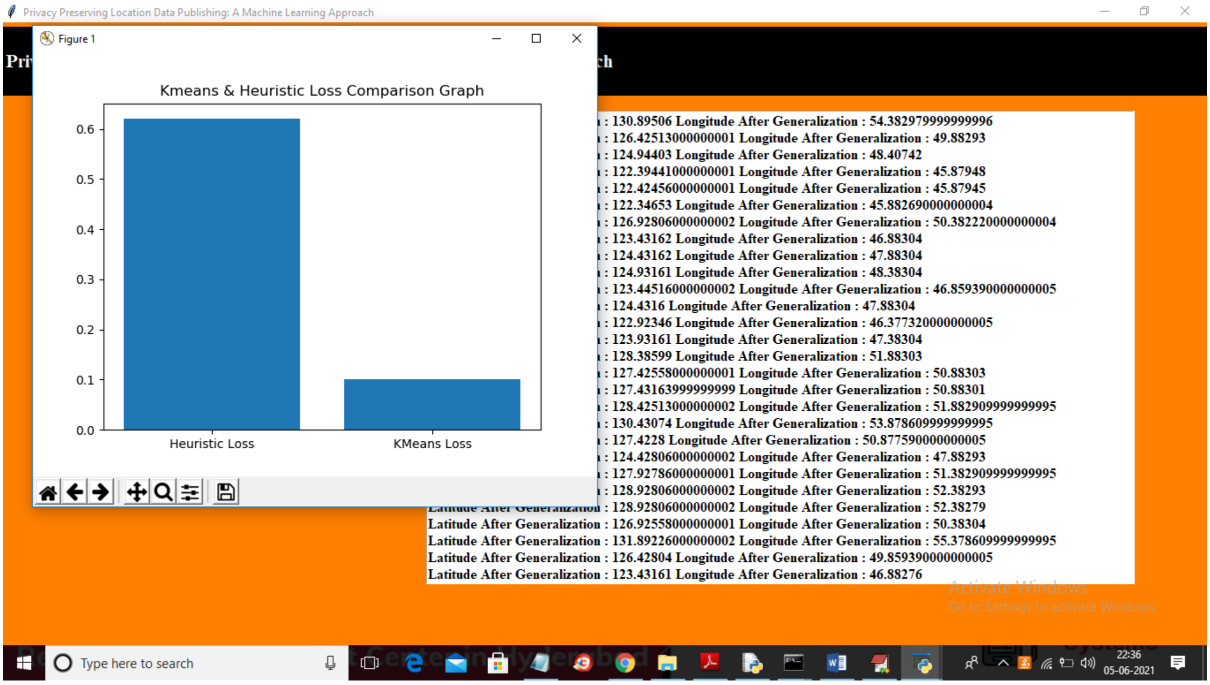
```
id,querydate,latitude,longitude
7105,2008-02-02 13:48:19,116.42456,39.86968
7105,2008-02-02 14:01:52,116.44516,39.91339
7105,2008-02-02 14:15:33,116.44403,39.90742
7105,2008-02-02 14:15:36,116.44403,39.90742
7105,2008-02-02 14:15:41,116.44403,39.90742
7105,2008-02-02 14:15:46,116.44403,39.90742
7105,2008-02-02 14:29:17,116.43226,39.88303
7105,2008-02-02 14:29:21,116.4318,39.88303
7105,2008-02-02 14:29:26,116.4316,39.88304
7105,2008-02-02 14:29:31,116.43161,39.88304
7105,2008-02-02 14:29:36,116.43161,39.88305
7105,2008-02-02 14:29:41,116.43162,39.88304
7105,2008-02-02 14:29:46,116.43163,39.88304
7105,2008-02-02 14:29:56,116.43164,39.88302
7105,2008-02-02 14:30:01,116.43159,39.88303
7105,2008-02-02 14:30:06,116.43126,39.88301
7105,2008-02-02 14:30:11,116.43074,39.883
7105,2008-02-02 14:30:16,116.43007,39.88298
7105,2008-02-02 14:30:21,116.42939,39.88298
7105,2008-02-02 14:30:31,116.42823,39.88294
7105,2008-02-02 14:30:36,116.42804,39.88293
7105,2008-02-02 14:30:41,116.42806,39.88292
7105,2008-02-02 14:30:46,116.42806,39.88292
7105,2008-02-02 14:30:51,116.42806,39.88293
7105,2008-02-02 14:30:56,116.42806,39.88293
```

In below screen you can see the same location is generalized with other values


Privacy Preserving Location Data Publishing: A Machine Learning Approach
Privacy Preserving Location Data Publishing: A Machine Learning Approach
Upload Taxi Trajectory Dataset
E:/venkat/2021/May21/LocationPrivacy/TaxiDatas
Preprocess Dataset
Run KMeans with DynamicSA Algorithm
Run Data Generalization Algorithm
Loss Comparison Graph
Type here to search
22:35
05-06-2021

```
Latitude After Generalization : 130.89506 Longitude After Generalization : 54.382979999999996
Latitude After Generalization : 126.42513000000001 Longitude After Generalization : 49.88293
Latitude After Generalization : 124.94403 Longitude After Generalization : 48.40742
Latitude After Generalization : 122.39441000000001 Longitude After Generalization : 45.87948
Latitude After Generalization : 122.42456000000001 Longitude After Generalization : 45.87945
Latitude After Generalization : 122.34653 Longitude After Generalization : 45.882690000000004
Latitude After Generalization : 126.92806000000002 Longitude After Generalization : 50.382220000000004
Latitude After Generalization : 123.43162 Longitude After Generalization : 46.88304
Latitude After Generalization : 124.43162 Longitude After Generalization : 47.88304
Latitude After Generalization : 124.93161 Longitude After Generalization : 48.38304
Latitude After Generalization : 123.44516000000002 Longitude After Generalization : 46.859390000000005
Latitude After Generalization : 124.4316 Longitude After Generalization : 47.88304
Latitude After Generalization : 122.92346 Longitude After Generalization : 46.377320000000005
Latitude After Generalization : 123.93161 Longitude After Generalization : 47.38304
Latitude After Generalization : 128.38599 Longitude After Generalization : 51.88303
Latitude After Generalization : 127.42558000000001 Longitude After Generalization : 50.88303
Latitude After Generalization : 127.43163999999999 Longitude After Generalization : 50.88301
Latitude After Generalization : 128.42513000000002 Longitude After Generalization : 51.882909999999995
Latitude After Generalization : 130.43074 Longitude After Generalization : 53.878609999999995
Latitude After Generalization : 127.4228 Longitude After Generalization : 50.877590000000005
Latitude After Generalization : 124.42806000000002 Longitude After Generalization : 47.88293
Latitude After Generalization : 127.92786000000001 Longitude After Generalization : 51.382909999999995
Latitude After Generalization : 128.92806000000002 Longitude After Generalization : 52.38293
Latitude After Generalization : 128.92806000000002 Longitude After Generalization : 52.38279
Latitude After Generalization : 126.92558000000001 Longitude After Generalization : 50.38304
Latitude After Generalization : 131.89226000000002 Longitude After Generalization : 55.378609999999995
Latitude After Generalization : 126.42804 Longitude After Generalization : 49.859390000000005
Latitude After Generalization : 123.43161 Longitude After Generalization : 46.88276
```

In above screen you can all location values are generalized so no malicious users can understand correct location. Now click on ‘Loss Comparison Graph’ button to get below graph


Privacy Preserving Location Data Publishing: A Machine Learning Approach
Figure 1
Kmeans & Heuristic Loss Comparison Graph
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Heuristic Loss
KMeans Loss
Activate Windows
Go to Settings to activate Windows.
Type here to search
22:36
05-06-2021

```
: 130.89506 Longitude After Generalization : 54.382979999999996
: 126.42513000000001 Longitude After Generalization : 49.88293
: 124.94403 Longitude After Generalization : 48.40742
: 122.39441000000001 Longitude After Generalization : 45.87948
: 122.42456000000001 Longitude After Generalization : 45.87945
: 122.34653 Longitude After Generalization : 45.882690000000004
: 126.92806000000002 Longitude After Generalization : 50.382220000000004
: 123.43162 Longitude After Generalization : 46.88304
: 124.43162 Longitude After Generalization : 47.88304
: 124.93161 Longitude After Generalization : 48.38304
: 123.44516000000002 Longitude After Generalization : 46.859390000000005
: 124.4316 Longitude After Generalization : 47.88304
: 122.92346 Longitude After Generalization : 46.377320000000005
: 123.93161 Longitude After Generalization : 47.38304
: 128.38599 Longitude After Generalization : 51.88303
: 127.42558000000001 Longitude After Generalization : 50.88303
: 127.43163999999999 Longitude After Generalization : 50.88301
: 128.42513000000002 Longitude After Generalization : 51.882909999999995
: 130.43074 Longitude After Generalization : 53.878609999999995
: 127.4228 Longitude After Generalization : 50.877590000000005
: 124.42806000000002 Longitude After Generalization : 47.88293
: 127.92786000000001 Longitude After Generalization : 51.382909999999995
: 128.92806000000002 Longitude After Generalization : 52.38293
Latitude After Generalization : 128.92806000000002 Longitude After Generalization : 52.38279
Latitude After Generalization : 126.92558000000001 Longitude After Generalization : 50.38304
Latitude After Generalization : 131.89226000000002 Longitude After Generalization : 55.378609999999995
Latitude After Generalization : 126.42804 Longitude After Generalization : 49.859390000000005
Latitude After Generalization : 123.43161 Longitude After Generalization : 46.88276
```

In above graph x-axis represents algorithm name and y-axis represents loss values generated for that algorithm and in above graph KMEANS got less loss so KMEANS is better in anonymization.

4. CONCLUSION

In this paper, we have proposed a framework to preserve the privacy of users while publishing the spatiotemporal trajectories. The proposed approach is based on an efficient alignment technique termed as progressive sequence alignment in addition to a machine learning clustering approach that aims at minimizing the incurred loss in the anonymization process. We also devised a variation of k 0 -means algorithm for guaranteeing the k-anonymity in overly sensitive datasets. The experimental results on real-life GPS datasets indicate the superior spatial utility performance of our proposed framework compared with the previous works

REFERENCES

[1] S. Shaham, M. Ding, B. Liu, Z. Lin, and J. Li, “Machine learning aided anonymization of spatiotemporal trajectory datasets,” arXiv preprint arXiv:1902.08934, 2019.

[2] A. Government, “New australian government data sharing and release legislation,” 2018.

[3] A. Tamersoy, G. Loukides, M. E. Nergiz, Y. Saygin, and B. Malin, “Anonymization of longitudinal electronic medical records,” IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 3, pp. 413–423, 2012.

[4] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, and D. Jin, “Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data,” in Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017, pp. 1241–1250.

[5] Y. Dong and D. Pi, “Novel privacy-preserving algorithm based on frequent path for trajectory data publishing,” Knowledge-Based Systems, vol. 148, pp. 55–65, 2018.

[6] M. Gramaglia, M. Fiore, A. Tarable, and A. Banchs, “Towards privacy-preserving publishing of spatiotemporal trajectory data,” arXiv preprint arXiv:1701.02243, 2017.

[7] M. Terrovitis, G. Poulis, N. Mamoulis, and S. Skiadopoulos, “Local suppression and splitting techniques for privacy preserving publication of trajectories,” IEEE Trans. Knowl. Data Eng, vol. 29, no. 7, pp. 1466–1479, 2017.

[8] M. E. Nergiz, M. Atzori, and Y. Saygin, "Towards trajectory anonymization: a generalization-based approach," in Proc. of the SIGSPATIAL ACM GIS. ACM, 2008, pp. 52–61.

[9] S. Gurung, D. Lin, W. Jiang, A. Hurson, and R. Zhang, "Traffic information publication with privacy preservation," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 5, no. 3, p. 44, 2014.

[10] R. Yarovoy, F. Bonchi, L. V. Lakshmanan, and W. H. Wang, "Anonymizing moving objects: How to hide a mob in a crowd?" in Proc. of the 12th International Conference on Extending Database Technology: Advances in Database Technology. ACM, 2009, pp. 72–83.

[11] B. Liu, W. Zhou, T. Zhu, L. Gao, and Y. Xiang, "Location privacy and its applications: A systematic study," IEEE Access, vol. 6, pp. 17 606–17 624, 2018.

[12] G. Poulis, G. Loukides, S. Skiadopoulos, and A. GkoulalasDivanis, "Anonymizing datasets with demographics and diagnosis codes in the presence of utility constraints," Journal of biomedical informatics, vol. 65, pp. 76–96, 2017.

[13] T. Takahashi and S. Miyakawa, "Cmoa: Continuous moving object anonymization," in Proceedings of the 16th International Database Engineering & Applications Sysmposium. ACM, 2012, pp. 81–90.

[14] X. Zhou and M. Qiu, "A k-anonymous full domain generalization algorithm based on heap sort," in International Conference on Smart Computing and Communication. Springer, 2018, pp. 446–459.

[15] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Incognito: Efficient full-domain k-anonymity," in Proceedings of the 2005 ACM SIGMOD international conference on Management of data. ACM, 2005, pp. 49–60.

[16] S. Yaseen, S. M. A. Abbas, A. Anjum, T. Saba, A. Khan, S. U. R. Malik, N. Ahmad, B. Shahzad, and A. K. Bashir, "Improved generalization for secure data publishing," IEEE Access, vol. 6, pp. 27 156–27 165, 2018.

[17] G. Poulis, A. Gkoulalas-Divanis, G. Loukides, S. Skiadopoulos, and C. Tryfonopoulos, "Secreta: A tool for anonymizing relational, transaction and rt-datasets," in Medical Data Privacy Handbook. Springer, 2015, pp. 83–109.

[18] K. Sreedhar, M. Faruk, and B. Venkateswarlu, "A genetic tds and bug with pseudo-identifier for privacy preservation over incremental data sets," Journal of Intelligent & Fuzzy Systems, vol. 32, no. 4, pp. 2863–2873, 2017.

[19] M. E. Rana, M. Jayabalan, and M. A. Aasif, "Privacy preserving anonymization techniques for patient data: An overview," in Third International Congress on Technology, Communication and Knowledge (ICTCK 2016), 2016.

[20] M. Jayabalan and M. E. Rana, "Anonymizing healthcare records: A study of privacy preserving data publishing techniques," Advanced Science Letters, vol. 24, no. 3, pp. 1694–1697, 2018.

[21] D. Narula, P. Kumar, and S. Upadhyaya, "Privacy preservation using various anonymity models," in Cyber Security: Proceedings of CSI 2015. Springer, 2018, pp. 119–130.

[22] J. Ding, "Trajectory mining, representation and privacy protection," in Proceedings of the 2nd ACM SIGSPATIAL PhD Workshop. ACM, 2015, p. 2.

[23] A. E. Cicek, M. E. Nergiz, and Y. Saygin, "Ensuring location diversity in privacy-preserving spatio-temporal data publishing," The VLDB JournalThe International Journal on Very Large Data Bases, vol. 23, no. 4, pp. 609–625, 2014.

[24] C. Romero-Tris and D. Meg´ıas, “Protecting privacy in trajectories with a user-centric approach,” ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 12, no. 6, p. 67, 2018.

[25] F. T. Brito, A. C. A. Neto, C. F. Costa, A. L. Mendonc¸a, and J. C. Machado, “A distributed approach for privacy preservation in the publication of trajectory data,” in Proceedings of the 2nd Workshop on Privacy in Geographic Information Collection and Analysis. ACM, 2015, p. 5.

[26] E. Naghizade, L. Kulik, and E. Tanin, “Protection of sensitive trajectory datasets through spatial and temporal exchange,” in Proc. of the 26th International Conference on Scientific and Statistical Database Management. ACM, 2014, p. 40.

[27] K. Jiang, D. Shao, S. Bressan, T. Kister, and K.-L. Tan, “Publishing trajectories with differential privacy guarantees,” in Proc. of the 25th International Conference on Scientific and Statistical Database Management. ACM, 2013, p. 12.

[28] L. Sweeney, “k-anonymity: A model for protecting privacy,” International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 05, pp. 557–570, 2002.

[29] B. Chowdhury and G. Garai, “A review on multiple sequence alignment from the perspective of genetic algorithm,” Genomics, 2017.

[30] X. Chen, C. Wang, S. Tang, C. Yu, and Q. Zou, “Cmsa: a heterogeneous cpu/gpu computing system for multiple similar rna/dna sequence alignment,” BMC bioinformatics, vol. 18, no. 1, p. 315, 2017.

[31] Q. Le, F. Sievers, and D. G. Higgins, “Protein multiple sequence alignment benchmarking through secondary structure prediction,” Bioinformatics, vol. 33, no. 9, pp. 1331–1337, 2017.

[32] J. MacQueen et al., “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. Oakland, CA, USA, 1967, pp. 281–297.

[33] S. K. Pal and P. P. Wang, Genetic algorithms for pattern recognition. CRC press, 2017.

[34] . Fischer and D. Picard, “Convergence rates for smooth k-means change-point detection,” arXiv preprint arXiv:1802.07617, 2018.

[35] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, “Mining interesting locations and travel sequences from gps trajectories,” in Proceedings of the 18th international conference on World wide web. ACM, 2009, pp. 791–800.

[36] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, “Understanding mobility based on gps data,” in Proceedings of the 10th international conference on Ubiquitous computing. ACM, 2008, pp. 312–321.

[37] Y. Zheng, X. Xie, and W.-Y. Ma, “Geolife: A collaborative social networking service among user, location and trajectory.” IEEE Data Eng. Bull., vol. 33, no. 2, pp. 32–39, 2010.

[38] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, “T-drive: driving directions based on taxi trajectories,” in Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems. ACM, 2010, pp. 99–108