Location Based Recommendation System with machine learning.

Ananya Chaudhary, Lakshmi Sharma, Dr. Anu Bala

School of Computing Science
and Engineering, Galgotias University, Greater Noida

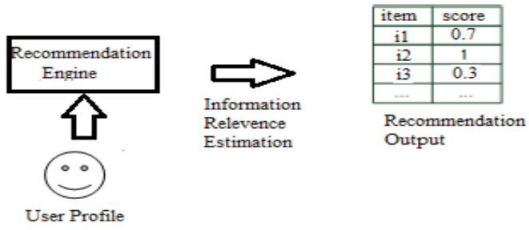
I. Abstract

Machine learning allows us to give vast volumes of data to computer algorithms and have them assess and make data-driven judgments and recommendations entirely based on the incoming data. This project will employ machine learning to analyze geolocation data and user preferences in order to provide the user with intelligent recommendations. In today's fast-paced and hectic world, it's common to be too exhausted to prepare a home-cooked supper. Even if you eat home-cooked meals every day, it's not uncommon to desire to treat yourself to a decent dinner now and then for social or recreational reasons. Consider the instance of someone who has recently relocated to a new location. They already have particular tastes and interests. . If a person lives near his favorite outlet, it will save him and his food providers a lot of time and effort. This project involves categorizing popular snack destinations for incoming students based on their mood and preferences for amenities, affordability, and proximity to the place using K-Means Clustering to find the easiest food location for students in Greater Noida (or any other city of your choice) by using K-Means Clustering to find the easiest food location for students in Greater Noida (or any other city of your choice) by using K-Means Clustering to find the easiest food location for students in Greater Noida.

Keywords --- Geolocation data, data driven judgement, machine learning, intelligent recommendation, k means clustering.

II. Introduction

People are frequently too exhausted to create home-cooked meals in the fast-paced and hectic environment in which they live. Of course, even if you eat homemade cuisine every day, it's not uncommon to want to go out to eat for social/recreational reasons once in a while. However, it is a widely held belief that, regardless of where you live, the food you consume is a significant component of your lifestyle. Consider a situation in which a student has recently relocated. They already have particular tastes and interests. If the student lives near his or her favorite outlet, this will save them a lot of time and money. . By forming clusters based on data similarity and then categorizing lodging and restaurant recommendations for incoming students based on what they are looking for, this analysis system uses K-Means Clustering to find the best locational recommendations for users in their own city or any other place of our choice!



Location based recommendation system basic architecture.[1]

III. Related work

With the rise of the human population, technology, and the burgeoning tourism business in the modern period, geolocation analysis, route suggestion systems, and destination recommendations acquired traction. For example, on-demand services are akin to a shared taxi that only operates when passengers want to travel between the origin and destination locations. It has a number of advantages over fixed-route buses, but riders dislike having to enter data such as origins, destinations, and deadlines. A forecasting site recommendation system would help riders during the reservation process as well as target potential riders while buses are idling. This study offers a flexible and

scalable foundation for such location recommendation systems. It is based on users' location histories and spatio-temporal correlations across locales, and it combines collaborative filtering algorithms, which are widely used in e-commerce, with link propagation, a popular data mining tool. Experiments on real-world data reveal that recommendations that include spatiotemporal data are more accurate than those that do not.

IV. Literature review

There are a number of algorithms that can be used to automatically calculate k using the information presented above. The majority of these methods are k-means or other clustering algorithms for fixed k that have been remixed. The vast majority of us have a proclivity for finding connections between objects or people. Food, behaviors, people, clouds, and so on, for example, all share commonalities. However, if an algorithm takes care of it, we won't have to worry about it. In this case, a machine learning algorithm can help. In machine learning, K-means clustering is a common unsupervised learning approach.[6] Clustering is a common data analysis technique for detecting patterns in data and identifying similar patterns in various settings. It can be described as categorizing data and then determining subcategories within those groupings where the data points are comparable. Because there are no labels to compare and assess the output, this method is categorized as unsupervised learning. Clustering is a technique for organizing data and analyzing its structure. One of the most extensively used clustering techniques is K-means. The K-means algorithm uses the clustering technique to divide the data set into k clusters, each with a centroid, and iteratively discovers the distance between the data points and centroids using distance metrics such as Euclidean distance, hamming distance, and so on. Finally, based on the properties of the data, linked data points are sorted into clusters.

We may use the remix clustering method to divide and mix the principles for centers as the process continues, allowing us to increase or reduce the value of k. After computing the BIC or Bayesian Information Criterion for each clustering model (BIC may be a mechanism for assessing and selecting a model), other scoring functions, in addition to BIC, are available. To find the simplest k, some scholars employ the MDL technique. The researchers also used the Minimum Description Length (MDL) framework, in which the outline length is a measurement value that indicates how well information fits the model. This algorithm starts with an oversized k value and removes the middle (reduces k) whenever the length of the outline is reduced. They employed the k-means approach to optimize the model's fit to the data during one of the k reduction steps.

V. Formulation of problem

- People in today's hectic environment are frequently too weary to prepare home-cooked meals and prefer to go out to dine once in a while for social/recreational reasons. However, it is a widely held belief that the food we consume is a vital component of our lifestyle, regardless of where we reside. Consider the following scenario: you've recently moved to a new place and don't want to go to the trouble of checking out every restaurant in the area for your chosen site.
- This is where the data model will save consumers a lot of time and effort by assisting them in making decisions rather than going through each and every place separately.
- Geospatial data can provide valuable information about the physical environment. It can be used by governments to save lives and improve public health. Businesses can use it to attract new clients by capturing the attention of the right people at the right time. Anonymized location data has a lot of value to offer without infringing on people's privacy.
- However, there are certain concerns concerning the accuracy and trustworthiness of location data. Many location data providers are focused with accumulating enormous volumes of data in order to entice and attract customers, and the quality is often subpar. As a result, firms only obtain poor results despite substantial study and resources. Even the most advanced algorithms can't tell the difference between good and poor data.
- The sources of location data you select to work with, as well as the suppliers you choose and their internal business practices,

all have an impact on data quality and, as a result, your business decisions and ROI.

- While processing data, many analysis/recommendation systems that are interested in location data have concerns. It can be challenging to clean and normalize location data for analytics or other corporate tasks.
- Understanding when data is useful for particular analytics is a related but distinct challenge.
- In order to access as much filtered data as possible and to comprehend the data purification process, we want an effective API. They must provide uniform documentation and solutions to help partners get the most out of their data.
- The value of location data is enormous. The main focus will be on fine-tuning a solid approach that addresses security concerns. The development and management of location data standardization will be required.

VI. Common use cases

- Those who want to employ IP-based geolocation to track down web visitors can do so in a variety of ways. Here are a few examples of common phrases:
- Using geolocation to highlight a different product to customers: geolocation can be used to showcase a different product to customers. Clients in one region may, for example, receive a service (such as food delivery), whereas customers in another region may receive restaurant ratings or information about popular attractions nearby.
- Display appropriate business hours for a user in a certain country or state: geolocation analysis can be used to display appropriate business hours for a user in a specific country or state.
- Creating a feeling of place: a corporation might use geolocation data to show the name of a relevant state or nation in a website title to create a sense of place and attract attention.
- Translating crucial messages into the user's language: a whole site translation could be time-consuming, costly, and difficult to maintain. Using the user's geolocation to translate crucial themes like headlines and calls to action can help reduce bounce rates and increase frequent user rates in many cases. For example, we can translate essential messages such as titles and calls to action.
- Geolocation can be used to establish location-based events and prizes. For example, Zomato frequently hosts location-based food festivals, and Amazon frequently hosts location-based seasonal discounts, such as Black Friday in the United States and Diwali in India.
- Redirect visitors to a site in their native language: You can employ location-based analysis to provide consumers with more personalized and relevant information.

VII. K-Means algorithm

K-Means algorithm based on dividing[4][5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Root-mean-squared error and error criteria are the basis of this algorithm with the aim of minimizing the cluster performance index. To find the best result, this algorithm tries to find a K pitch that meets certain criteria. First, select some points to represent the first cluster focus point (usually select the first K sample points of income to represent the first cluster focus point). Then collect the remaining sample points in focus according to the minimum distance criterion to get the first classification. If the classification is unreasonable, change it (recalculate all cluster foci) and repeat until you get a reasonable value. You will receive the classification. The split-based KMeans algorithm is a kind of clustering algorithm with the advantages of simplicity, efficiency, and speed.

VIII. Required tools

- Python
- For data - NumPy and pandas package.
- For plotting - matplotlib package & seaborn packages.
- For geospatial – geopy, folium package.
- For machine learning - sklearn (pre-processing and cluster), SciPy package
- For deep learning - minisom package.

Numpy –

NumPy is a popular Python data analysis library. NumPy allows you to speed up your workflow and interact with other Python ecosystem tools such as scikit-learn, which use NumPy under the hood. NumPy was created in the mid-2000s as an offshoot of an even earlier package called Numeric.[2] Because of its durability, practically every data

analysis or machine learning tool for Python makes use of NumPy in some way.

Matplotlib-

Matplotlib is a Python package that enables the creation of static, animated, and dynamic visualisations.. Matplotlib makes simple things simple and difficult things possible. It is a Python low-level graph plotting package that serves as a visualisation tool. John D. Hunter designed it. It is open source, and we can freely utilise it. The majority of it is written in Python, with a few pieces written in C, Objective-C, and Javascript for platform compatibility.
2D and 3D toolkits are available.
Among its applications are:

• Produce plots suitable for publication.
• Create interactive figures that can be zoomed, panned, and updated.
• Change the visual style and layout.
• Export to a variety of file formats.
• Integrate with JupyterLab and Graphical User Interfaces.

Minisom-

MiniSom is a minimalistic and Numpy based implementation of the Self Organizing Maps (SOM). SOM is a type of Artificial Neural Network able to convert complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display.

Sklearn:

Scikit-learn (Sklearn) is the most usable and robust Python machine learning package. It provides a set of efficient tools for machine learning and statistical modelling, such as classification, regression, clustering, and dimensionality reduction, via a Python interface. This Python-based library is based on NumPy, SciPy, and Matplotlib.

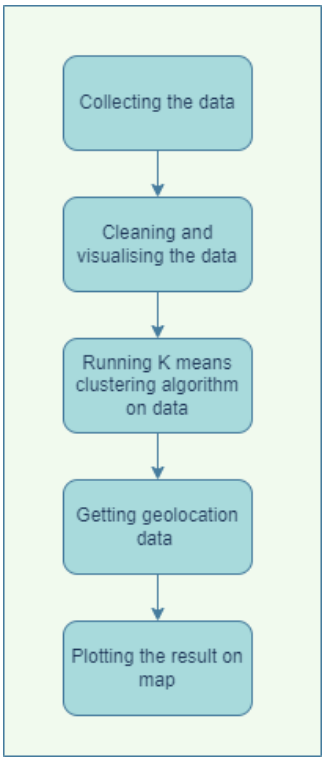
Geopy –

Geopy is a Python client for a number of well-known geocoding web services. It allows Python developers to easily find the coordinates of addresses, cities, countries, and landmarks all over the world by utilising third-party geocoders and other data sources.[8]It contains geocoder classes for OpenStreetMap Nominatim, Google Geocoding API (V3), and a variety of other geocoding services. The complete list can be found in the Geocoders documentation section. Geocoder classes are found in the geopy geocoders package. It has been tested with CPython (versions3.5,3.6,3.7,3.8, and3.9) and PyPy3. CPython 2.7, 3.4, and PyPy2 were also supported by the geopy 1.x line.

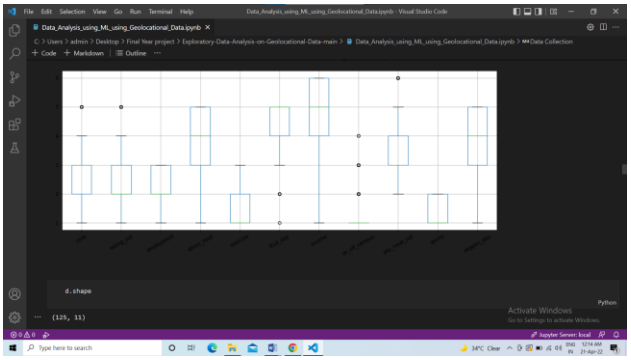
Seaborn -

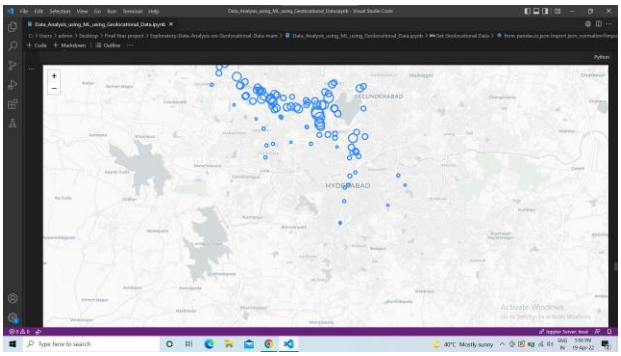
Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing
attractive and informative statistical graphics.[7]

IX. Complete work plan layout

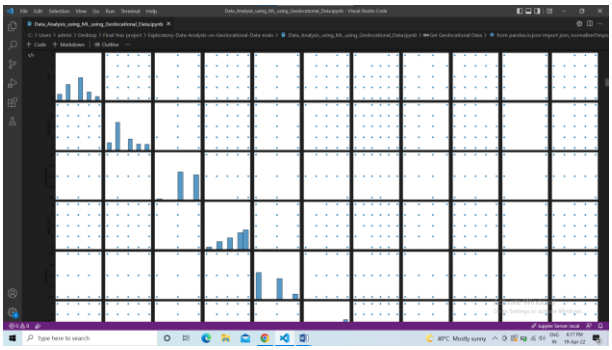


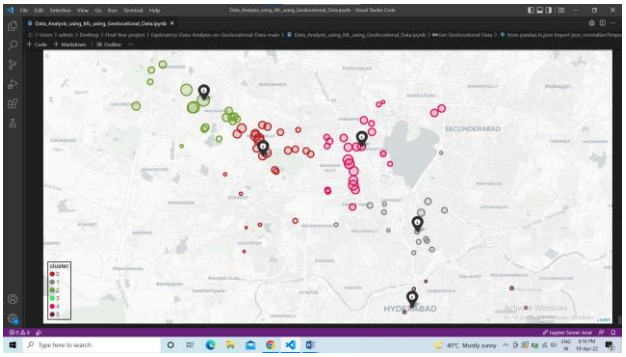
Working of software.





X. Brief look and output





XI. Applications

1. Diversification: The project's diversity can aid in the application of the software to various challenges involving a number of geographic suggesting systems in relation to geospatial data.

2. Small-scale application: The system can be applied to simple geolocational analytic problems at the district and city level.

3. Large-scale application: The method can be applied to larger-scale geolocational analytic problems at the national and international levels, albeit this will necessitate a larger user data collection.

4. Commercial use: The programme can be used to generate revenue by providing smart locational recommendations to users, increasing their odds of completing a purchase. This can be a profitable source of revenue.

XII. Areas for Improvement

Machine learning development is simple for ML developers, but validating its performance is crucial for achieving accurate and consistent results. There are, however, a number of strategies to boost the performance of your machine learning model. Machine learning models, which are typically written in Python, must be built while considering the various factors that affect their performance.[3]However, we've given a list of the most important factors to think about while you build your ML model. I.

XIII. Conclusion

Working on this project provided us with a lot of insight into the technologies that were utilized, as well as real-world experience as a software developer. This system can run efficiently without any fancy equipment and with just a computer system and a camera, making it suitable for all scale competitions. This project will be open-sourced and scalable, as there is always opportunity for development and changes, allowing it to offer more exact recommendations over time. It gave us a lot of insight into the technologies that were utilized, as well as real-world experience as a software developer, which is always valuable.

XIV. References

[1]Tiwari, S.; Tiwari, S.; Jagwani, P.; Kaushik, S. (2012). [Institution of Engineering and Technology IET International Conference on Wireless Communications and Applications (ICWCA 2012) - Kuala Lumpur, Malaysia (8-10 Oct. 2012)] IET International Conference on Wireless Communications and Applications (ICWCA 2012) - Location Based Recommender Systems: Architecture, Trends and Research Areas. , (), 071–071. doi:10.1049/cp.2012.2096

[2] Wes McKinney. 2018. Python for Data Analysis: Data Wrangling with pandas, NumPy, and IPython. O'Reilly

[3] Zhang, Qian; Lu, Jie; Jin, Yaochu (2020). *Artificial intelligence in recommender systems. Complex & Intelligent Systems, (), –.* doi:10.1007/s40747-020-00212-w

[4] Barnes, T. J. (2003) 'The place of locational analysis: a selective and interpretive history', Progress in Human Geography, 27(1), pp. 69–95. doi:

[5]ZhangYufang etc. A kind of improved K-means algorithm [J]. Computer Application,p31-33, 2003, (8).

[6] Mu-Hsing Kuo a,*, Liang-Chu Chen b , Chien-Wen Liang c 0957-4174/$ - see front matter 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.eswa.2008.02.014

[7] Hug, N., (2020). Surprise: A Python library for recommender systems. Journal of Open Source Software, 5(52), 2174. 21105/joss.02174

[8] Aggarwal, C. C., & others. (2016). Recommender systems (Vol. 1). Springer. doi:10.1007/ 978-3-319-29659-3