Developing a Search Strategy

# **Research Topic:** Application of K-Means Clustering for Mining nearest Wi-fi Hotspots

# **Problem:**

These days wi-fi has become a necessity just like food and shelter. It is one of the most important technologies being used in the internet era because It allows the users to connect to the internet wirelessly. These days many restaurants, malls, cafes and libraries provide free public wi-fi. Naturally, it becomes a valid option to have an application that can assist you in searching nearest wi-fi accurately. People do not have any efficient solution to identify wi-fi hotspot near them so that they can just reach there and get connected. In the era of technology that we are living in, life without internet has become unimaginable.

With this project I intend to generate a java k-means application to gather wi-fi hotspots near a user and identify the nearest available hotspot. The main area of focus will be creating be a java K-means Clustering implementation with a special equal option, this means that even if there are no close enough Wi-Fi hotspots to the user, my k-means clustering implementation will still try to get 5 nearest wi-fi hotspots.

# **Objective:**

Contribution of this project is mainly to develop a k-means clustering model to deal with the large dataset of wi-fi hotspots, that offer free internet access to people in a city, to generate accurate navigation application that can provide intelligent solution to basic navigation queries of the people like

1. Where can the user find hotspot nearby?
2. How many wi-fi hotspots are present close to user?
3. How far does the user need to travel to get to nearest wi-fi?

This application can further be scaled to many other datasets like restaurants, grocery outlets, etc.

The main object of the project is to be able to achieve a variation of k-means that will ensure that each k-means cluster will have uniform cluster size and each cluster has a uniform number of points. With the project I intend to set the centroid dynamically for the k-means algorithm to provide the user the five nearest wi-fi hotspots.

The goal is to use machine learning algorithm to locate the nearest wi-fi hotspots by augmenting the current location of the user.

# **Research Questions:**

1. Is it feasible to use k-means Clustering?
2. Are there are any other machine learning algorithms available that can provide more efficient results?
3. How much accuracy will the k-means clustering algorithm provide?
4. What is my target audience and how can the research be applied in real world?
5. How many outlier data items are present in the dataset?
6. How should I clean the data before applying the k-means clustering algorithm?
7. Will the efficiency of clustering model decrease if I ignore the outlier data elements?
8. How should I visualize the data clusters to represent data?
9. What are the repercussions of switching ON the special equal option in k-means clustering?
10. Is the dataset large enough to get efficient clusters through k-means clustering?
11. Am I permitted to use public data set to create my application?

# **Key Terms used:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Group 1 | Group 2 | Group 3 |
| Term 1 | k-means Clustering | Mining | wi-fi hotspot |
| Term 2 | k-means Cluster | identifying | wireless network |
| Term 3 | k-means | Detecting | Wi-fi hotspots |
|  |  |  |  |

Search String:

(k-means Clustering OR k-means Cluster OR k-means) AND (Mining OR identifying OR Detecting) AND (wi-fi hotspot OR wireless network OR Wi-fi hotspots)

* IEEE: 395 Articles found
* Google Scholar: 250 Articles found
* ACM: 324 Articles found

**What should I do differently?**

I need to narrow my articles down so that I get a smaller number of relevant articles. I need to add more relevant keywords to the grid.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Group 1 | Group 2 | Group 3 | Group 4 |
| Term 1 | k-means Clustering | Mining | wi-fi hotspot | Large dataset |
| Term 2 | k-means Cluster | identifying | wireless network | Big data |
| Term 3 | k-means | Detecting | Wi-fi hotspots |  |
| Term 4 |  |  | Wi-fi |  |

**Search String:**

(“k-means Clustering” OR “k-means Cluster” OR “k-means”) AND (“Mining” OR “identifying” OR “Detecting”) AND (“wi-fi hotspot” OR “wireless network” OR “Wi-fi hotspots” OR “wi-fi”) AND (“Large dataset” OR “Big data”)

* Google scholar: 99 articles
* IEEE: 178 articles

**Top Sources for my research:**

* Featured Dataset. City of New York… New York City WiFi Hotspots, Every public WiFi hotspot in New York City. https://www.kaggle.com/new-york-city/nyc-public-wifi/version/1/data
* H. H. Bock (2007) Clustering Methods: A History of k-Means Algorithms. In: P. Brito, G. Cucumel, P. Bertrand, F. de Carvalho Selected Contributions in Data Analysis and Classification. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg
* Jong Hee Kang, William Welbourne, Benjamin Stewart, and Gaetano Borriello. 2004. Extracting places from traces of locations. In Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots (WMASH '04). ACM, New York, NY, USA, 110-118. DOI=http://dx.doi.org/10.1145/1024733.1024748
* Li Dazi, Qian Li, Jin Qibing, Tan Tianwei. The Improved Global K-means Algorithm And Its Application In Data Classification [J]. Infonnation and Control, 2011, 40(1): 100-104 (in Chinese)
* A. Abbsi, M. Younis, "A survey on clustering algorithms for wireless sensor networks", Computer Communications Elsevier, vol. 30, pp. 28262841, June 2007.
* Veervrat Singh Chandrawanshi, Rajiv K. Tripathi, Nafis Uddin Khan, "A comprehensive study on k-means algorithms initialization techniques for wireless sensor network", Signal Processing and Communication (ICSC) 2016 International Conference on, pp. 154-159, 2016.
* Cablefree (2017) The History of Wifi, WIRELESS EXCELLENCE LIMITED, Oxford Science Park, Oxford, OX4 4GA. UK
* K. Wagstaff, C. Cardie, S. Rogers, S. Schrödl - ICML, 2001, Constrained k-means clustering with background knowledge, Proceedings of the Eighteenth International Conference on Machine Learning, 2001, p. 577–584.
* A. Trevino (2016) Introduction to K-means Clustering. https://www.datascience.com/blog/k-means-clustering. June 12, 2016.