A PRELIMENERY REPORT ON

IMMIGRANT SEGMENTATION FOR APPROPRIATE ACCOMMODATION USING EDA AND CLUSTERING ANALYSIS

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF
BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

GROUP 4

Devika Dhumal	Exam No :71914969L
Sunny Khade	Exam No :71915010J
Sanskriti Patole	Exam No :71915056G
Omkar Takle	Exam No :71915104L

DEPARTMENT OF COMPUTER ENGINEERING

TSSM's
PADAMBHOOSHAN VASANTDADA PATIL INSTITUTE OF TECHNOLOGY
BAVDHAN -21

SAVITRIBAI PHULE PUNE UNIVERSITY
2021 -2022

CERTIFICATE

This is to certify that the project report entitles
“IMMIGRANT SEGMENTATION FOR APPROPRIATE ACCOMMODATION USING EDA AND CLUSTERING ANALYSIS”

Submitted by

Devika Dhumal	Exam No :71914969L
Sunny Khade	Exam No :71915010J
Sanskriti Patole	Exam No :71915056G
Omkar Takle	Exam No :71915104L

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of Prof. **P. P. Dandawate** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

(Prof. P. P. Dandawate)
Guide
Department of Computer Engineering

(Prof. S. V. Bodake)
Head,
Department of Computer Engineering

(Dr. C.M. Sedani)
Principal,
Padmabhooshan Vasantdada Patil Institute of Technology, Bavdhan Pune – 41

Place: Pune
Date:

ACKNOWLEDGEMENT

We take this opportunity to express my deepest gratitude to those who have generously helped me in providing the valuable knowledge and expertise during my training.

It is great pleasure to represent this report on the project name "**Immigrant Segmentation for appropriate Accommodation using EDA and Clustering Analysis**" undertaken by me as part of my B.E (CS) curriculum. We are thankful to **Padmabhooshan Vasantdada Patil Institute of Technology** for offering me such a wonderful challenging opportunity.

We express my sincere gratitude to **Dr. C. M. Sedani** (Principal, PVPIT) for providing us an opportunity to undergo this major project as the part of the curriculum. We are thankful for our project guide **Prof. P. P. Dandawate** for her support, co-operation and motivation provided to me during training for constant inspiration, presence and blessings. We would also like to thank my H.O.D **Prof. S. V. Bodake** for her valuable suggestions which helped a lot in completion of this project.

NAME OF THE STUDENTS:

Devika Dhumal	Exam No :71914969L
Sunny Khade	Exam No :71915010J
Sanskriti Patole	Exam No :71915056G
Omkar Takle	Exam No :71915104L

ABSTRACT

A person can migrate due to various reasons such as job, studies, better lifestyle, urbanization etc. and the process of doing so is not very easy to begin with. There are many different aspects that need to be taken under consideration before one makes a shift to newer regions. Aspects such as budget, Amenities, Food preferences, Distance from a desired location, etc. need to be taken under consideration. From this observation we came up with the problem statement of "appropriate accommodations" and decided to run exploratory data analysis and clustering algorithms to see the trends and predictions that can be derived to make these big decisions a little easier.

While studying the accommodations we were able to derive the features that could be the reason for selecting a place or not. We future dived into understanding if any correlations existed. We were able to learn few things such as how amenities had correlations with budget. These findings were done with the help of visualization tools upon selected data set. With help of such understandings, we went ahead with our clustering algorithms. The clustering algorithm was able to cluster these correlations.

We got the most optimum value for k using k-mean algorithm. This value will help get suitable location using REST APIs to fetch the same.

In conclusion, this project will work on targeting key concepts like exploratory data analytics. This will help improve prediction probabilities. Like each project many literature surveys were done and all pros and cons are taken under consideration. This project has helped understand the limits and limitations for exploratory analysis and prediction probabilities while giving a great insight in the whole workflow on a data scientist.

Keywords: Exploratory Data Analysis (EDA), K-mean Clustering, Unsupervised Learning, Machine Learning, Data Visualization, Data Preprocessing, Geolocation Data.

INDEX

LIST OF ABBREVATIONS

ABBREVIATION	ILLUSTRATION
EDA	Exploratory Data Analysis
REST	Representational State Transfer
ML	Machine Leaning
AI	Artificial Intelligence
API	Application Programming Interface

LIST OF FIGURES

LIST OF TABLES

01. INTRODUCTION

1.1 Introduction to Exploratory Data Analysis for machine learning:

In Machine Learning when getting started with a project, one needs to keep in mind that data is everything. It is the main component that feds to any machine learning algorithm. When deriving the trend and insight from a set of data could prove to be a difficult and error-filled job. To have a good start for a machine learning project it helps to analyze data at the beginning, this practice describes data by paths of statistical and visualization techniques to bring important features of data into focus for future analysis. During this process, it is important that we get a good understanding of:

- The properties found in the set of data, like views, schemas and stat properties;
- The quality of the set of data used, like missing value and inconsistent data
- The prediction power of data used, such as the correlation of variables, features etc against a target

This process forms the basis for selecting the following features and engineering steps and provides a solid foundation for building good ML models. There are many ways to conduct exploratory data analysis (EDA) out there, this may bring difficulties in terms of what to analyze and what to perform. Based on the expected outcome of the analysis one can then determine feature selection and engineer appropriate recommendation.

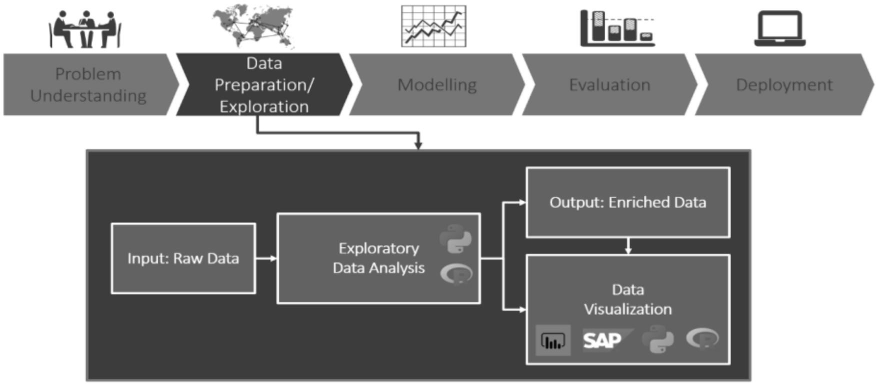

Fig1.1. Exploratory data analysis in Machine Learning.

In our project we are trying to target the same technology for the problem statement of "Appropriate immigrant accommodations". Through the help of EDA, we wish to a good insight into the features and aspects that would help any immigrant individual to make the best possible decision.

1.2 Motivation

As students coming from engineering background, we have had acquaintances and friends who were not locals. We had the opportunity to first hand see the struggles they went through while choosing the right locality while stills trying to manage an array of aspects. This gave us a great domain interest and insight. Hence our problem statement was made. We started asking around about all the aspects and thought process that takes place while choosing an accommodation. We took part in the quest and came up with our own findings. Majorly studying the student body, we were able to find reasons like, distance from a desired location, budget and food preferences being a few of main fields of consideration.

This close experience with finding the right accommodations for people we knew became the foundation and motivation for this project. We decided to make a project analyzing the aspects that would give the least to most favorable suggestions as expected outcome on a well visualized platform as for any non-technical person could understand with ease.

1.3 Problem Definition

Whenever the thought of relocation takes place the first thing, we focus on is our basic necessities like shelter and food. Taking this thought process in consideration we could easily derive that an appropriate accommodation makes in onto the priority list while migrating any place new. One could be relocating for a number of reasons like, job, higher studies etc. For the purpose of understanding and implementing our project we shall consider the migrating body to be of a student.

Being students ourselves we know how difficult it is settled in a new region, locality and area. Coming from a country as diverse as India we can observe a change in language, food, living standards over changing localities. And this may stand to be difficult for us younger generations. To overcome the awkwardness of adapting we shall try and target individual Points-of-Interests (POI).

To overcome the problems stated while relocating we shall use Exploratory Data Analytics (EDA) and try to find a good combination of aspects in a single locality which will help appease the individual and make the difficulty process of relocating a little easier.

02. LITERATURE SURVEY

2.1 Introduction

Data Collection and mining is the process of finding previously unknown patterns and insights in a database and using these findings we can build a good recommender system. This step dealing with data combines statistical analysis, machine Learning and technologies to extract hidden patterns and trends in relation to our targeted features. The World Migration Report 2020 enlightens the dramatic change in respect of migration. “The current United Nations global estimate is that there were around 281 million international migrants in the world in 2020, which equates to 3.6 per cent of the global population”. Migration is a complex concept. The Institute for Human Rights briefed recently regarding the overview of the human rights risks posed to migrant workers via the practices surrounding accommodation provision. This Literature survey will help struggle

understand all the techniques and technologies we’ll need to understand before we tackle our problem statement for appropriate accommodations.

2.2 Purpose of Literature Survey

This literature survey plays important role in our research process. It’s our source from where we research, brainstorm and understand all the prerequisites and approaches that will help us solve our problem statement. Depending on what we observed in the literature survey will determine the approach we adapt to.

Objectives:

- Understand the approach (i.e., concepts like EDA, clustering analysis, recommender systems)
- Broaden our understanding on the chosen domain.
- Understand the actual pros and cons behind every step

2.3 Literature Review

Numerous studies discuss the concept of Exploratory Data Analysis, Recommender system, clustering algorithm etc. They have applied different ways and algorithms to find the best and most accurate approach to create a good recommender system. Recommender systems recommend to users based on their interests and preferences. To overcome the problem statement of appropriate suggestions, recommender systems have come to exist to provide technological proxy (Chen, 2011), to determine if a user would like a specific item via making prediction, or recommending top items to the user based on her preferences and analyzing the user behavior. Recommender Systems (RS) can be widely applied in different areas.

Two types of recommender systems are widely known and adapted which are collaborative filtering and content-based filtering. In collaborative filtering, that is also referred as social filtering (Shardanand & Maes, 1995), items are selected based on the correlation between the current (active) user and other users of the system (Su & Khoshgoftaar, 2009). However, in content-based filtering items are recommended based on the correlation between items and the user preferences (Adomavicius et al., 2005). In addition to the mentioned techniques there are different hybrid ways which are developed by combination of these two in order to overcome certain limitations.

(J. Cao, Y. Dong, P. Yang, T. Zhou, and B. Liu, 2016) proposed a method for deciding user's visiting probability for certain POI, based on calculation of characteristics of meta-path. Here's where our idea of targeting individuals point-of-interests began. (M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, 2011) proposed an improved framework by integrating influence of social media friends and geographic location with original UB-CF. We can get a good understanding of where and which data and domain to study to get a good understanding of geographic targeted location recommender systems.

(KhanhQuan Truong,Fuyuki Ishikawa, Shinichi Honiden, 2015) helped understand Recommender System (RS) predicts user's preference , and then recommends highly-predicted items to user. It gave an insight on points like the number of attributes is often very large and so is the diversity amongst them, users who have similar preference in one category may have totally different judgement on attributes of another kind. (Mubaraka Sani Ibrahim and Charles Isah Saidu, recommend er Systems: Algorithms, Evaluation and Limitations) proposed the state of art machine learning based recommendation n models including Clustering models and Bayesian Classifiers. Further, we discuss the widespread ap plication of recommender systems. We got a good understanding and concluded It may provide inaccurate results if data entered incorrectly. Since the feature representation of the attributes are hand engineered to some extent, many techniques require a lot of domain knowledge. (Stamatelato s, G., Drosatos, G., Gyftopoulo s, S. et al., 2021) proposed a new idea on personalized interest-based recommender system which is based on lists of end user point of interests. The model can only make recommendat ions based on existing interests of the user. In other words, the model has limited ability to expand on the users' existing interests. (Dietz, L.W., Sen, A., Roy, R. et al, 2020) proposed the idea of mining trips from location-based social networks for clustering travelers and destinations. (Qing Li and Byeong Man Kim, 2003) Clustering techniques have been applied to the item-based collaborative filtering framework to solve the cold start problem. It also suggests a way to integrate the content information into the collaborative filtering. Extensive experiments have been conducted on MovieLens data to analyze the characteristics of our technique. The results show that our approach contributes to the improvement of prediction quality of the item-based collaborative filtering, especially for the cold start problem.

03. APPROACH

3.1 Introduction to the proposed Architecture

In our project we will be using Exploratory Data Analysis to understand the dataset and to derive the meaning behind all aspects and features. After initial understanding is gained using exploratory data analysis, we shall be implementing Clustering Analysis as our base algorithm. We can give our chosen features from the CSV dataset as an entry to the system. After taking the input the algorithms apply on that input that is K-Means Clustering. After the accessing data set the operation is performed and appropriate recommendations are produced.

The proposed system will add more parameters significant to appropriate accommodations with their individual Point-Of-Interests in mind like, Budget and Distance from desired location accordance to the priority levels. The accommodation recommendation system is designed to help the provide appropriate suggestions related to accommodations.

3.2 Proposed Method

In this section, we have presented a recommendation system with EDA and k mean clustering techniques. The k-mean clustering approach is a part of unsupervised machine learning methodology. The EDA is used in our research because it works effectively to present good results to other existing methods.

The diagram of the proposed recommendation system is shown in the figure below. The population from the codebook_food Kaggle dataset was taken. The project consists of the following stages:

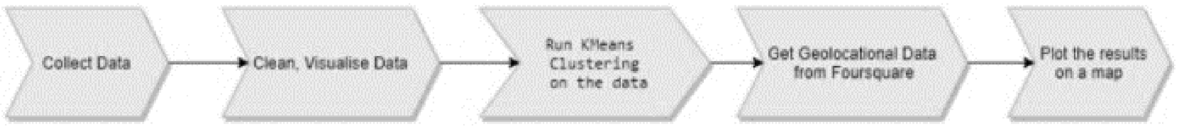
Collect Data
Clean, Visualise Data
Run KMeans Clustering on the data
Get Geolocational Data from Foursquare
Plot the results on a map

Fig 3.2: WorkFlow Diagram

We will follow the generic Machine Learning flow for understanding the exact approach our project will be taking.

3.3 Design Architecture and Implementation

We will be using EDA to get an accurate feature understanding. After the features are decided upon and data collection and cleaning are done, we shall push the data through our K-mean clustering algorithm which will result in the expected outcome of clustered locations. Now that we have the appropriate clusters and locational coordinates we can throw a request to our Foursquare REST API, and plot the same on a map which will allow the end-user to easily understand the analyzed locations instead of trying to understand the complex method working at the Backend.

Hence, we now understand how the process and architecture in place are working.

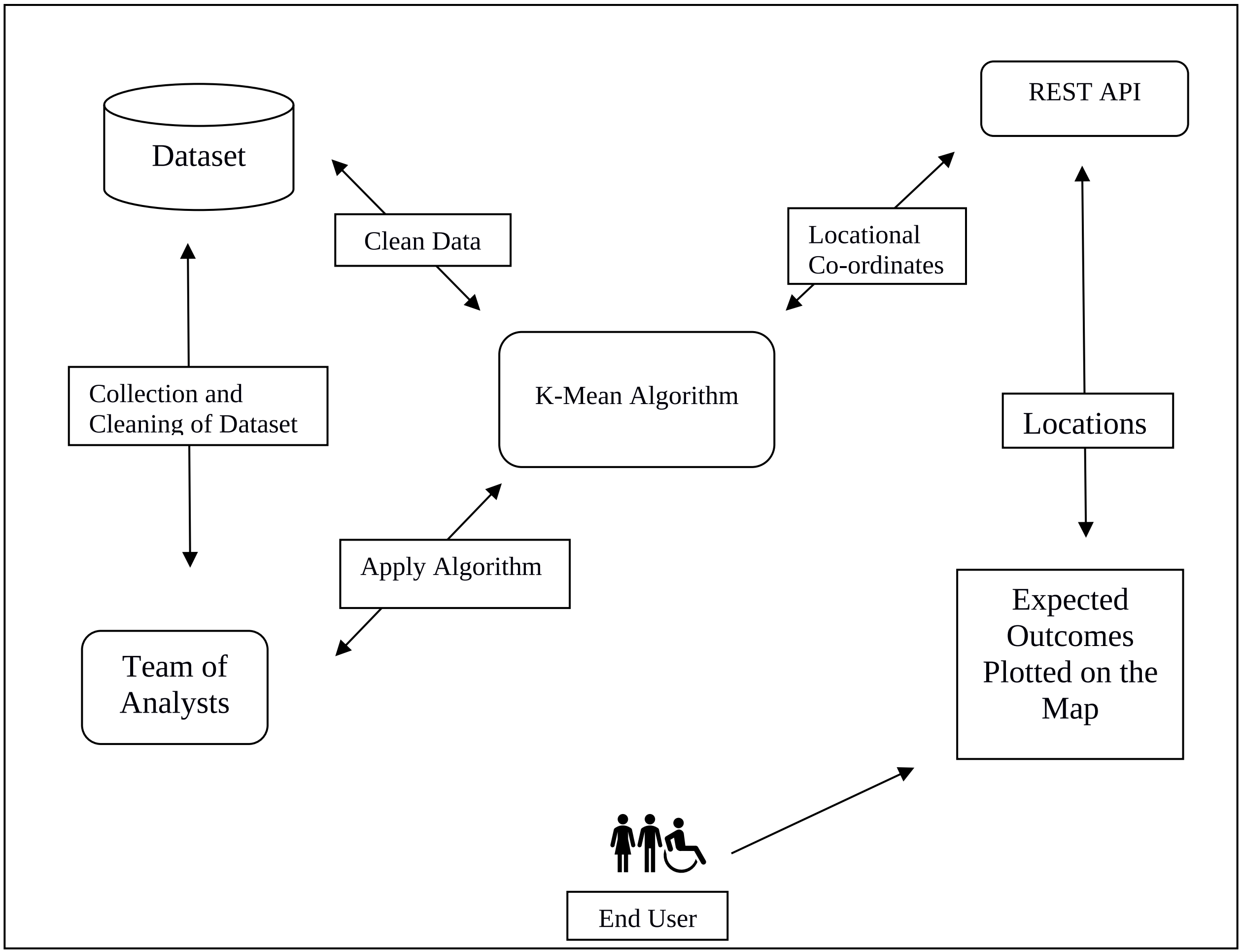

Fig 3.3: Architecture

k-mean Clustering is the task of forming groups or clusters of elements such that the observations of the same group are more similar to each other than those in other groups.

Affinity Propagation is a graph-based algorithm that assigns each observation to its nearest exemplar. All the observations 'vote' for which other observations they want to be associated with, which results in a partitioning of the whole dataset into a large number of uneven clusters.

Geolocational Analysis is the analysis that processes Satellite images, GPS coordinates and street addresses and applies to geographic models.

3.3.1 Parameters to be used

Our system followed following parameters to make the most accurate recommendations:

1) Distance: Distance will be taken under considersation when we are targeting a locality of preference. That can be a college for a student or an office for a working individual. Only locations falling under this targeted radius will be showed to the user. Also, for certain places and activities preference will be given to closer objects i.e for closer locations will contribute more to the calculation of affinity score.

   We have divided location into two categories:

   - Accommodation Location: where the user will be living.
   - Targeted Location: the location our user wishes to be in near locality for the ease pf travel.

2) Budget: A location with high number of stores nearby will be labelled as "Amenity Rich" i.e these localities will have better amenities like gyms, malls, societies, and better standard of living. While on the other hand locations with less amenities will be labeled as "Amenity Poor".

   Similar locations will be grouped or clustered together.

3) Food: Similarly, like the two stated features our user could target food to be a primary preference. Living in India we are already aware of how diversity can exist in cohabitation.

Hence, we can say that any preferred food will be available just need to be found and clustered in the appropriate locality for our user to benefit.

The various targeted features and preferences planned for the various clustering relations of the project perfectly corelate and successfully provide the appropriate foundation to create a good recommender system. The dataset found on Kaggle has all the chosen attributes and is well diverse to help gain a good insight without creating outliners that would disturb our machine. The K-mean algorithm helps show the most accurate and preferable accommodation and hence is chosen to be final.

3.4 Algorithm

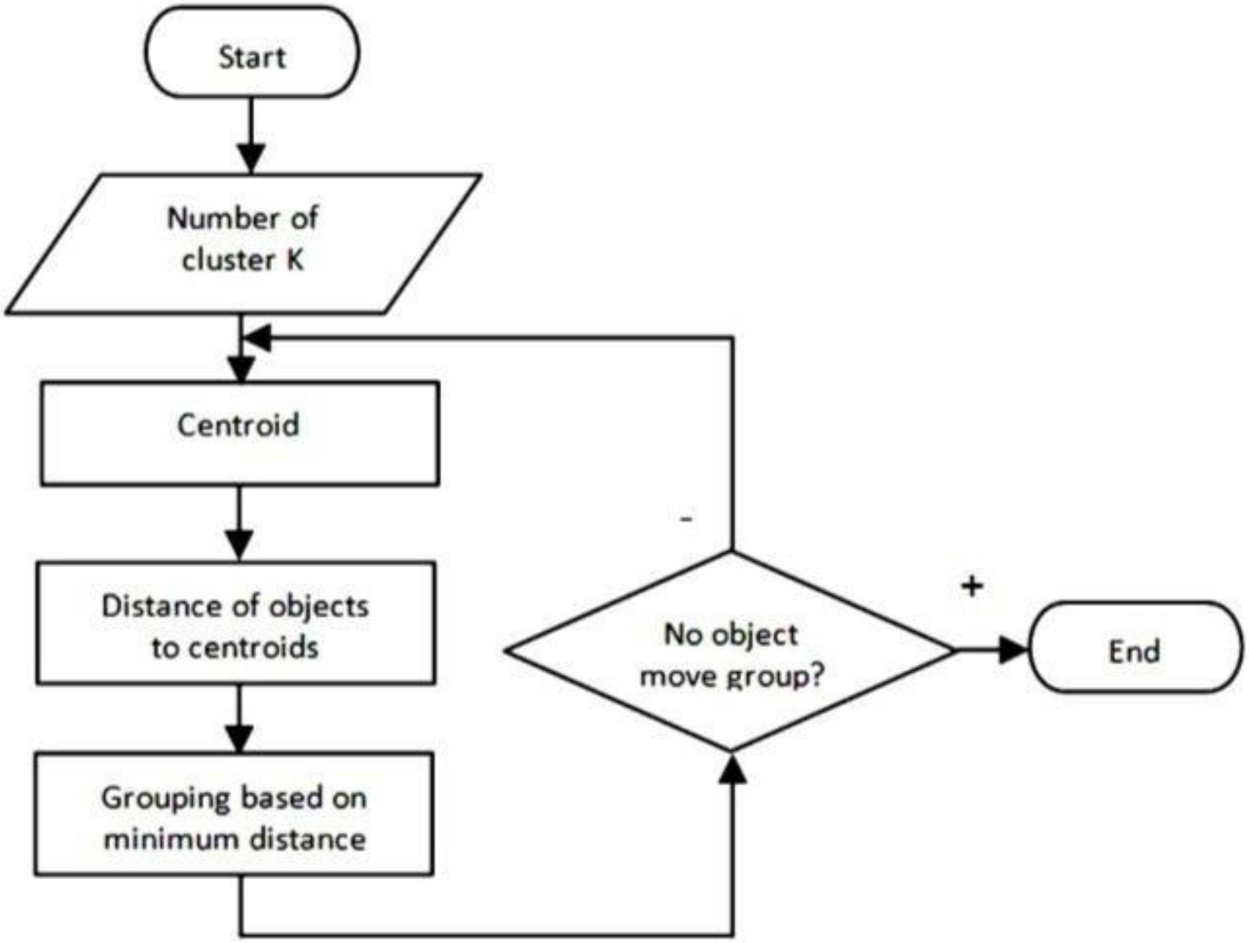
Start
Number of cluster K
Centroid
Distance of objects to centroids
Grouping based on minimum distance
No object move group?
-
+
End

Fig 3.4: K-Mean clustering Algorithm

K-mean clustering is an unsupervised Machine Learning Algorithm that is used to solve segmentation and clustering problems in machine learning. K-mean clustering algorithm computes the centroids and iterates until we find the optimal centroid. The fundamental step for our algorithms is to determine the optimal number of clusters into which data may be clustered. The elbow method is one of the most popular methods to determine this optimal value of k. Elbow method requires drawing a line plot between SSE (Sum of Squared errors) vs number of clusters and finding the point representing the 'elbow point' (the point after which the SSE decreases). Elbow method is used to determine the most optimal value of k representing numbers of clusters in k-mean clustering algorithm.

Elbow Method: In K-Means Clustering algorithm it is necessary to define the value of 'k', that is the number of clusters, before the execution of the algorithm. To find the ideal value of 'k' we use the elbow method.

The Elbow method works by using the concept of Within Cluster Sum of Squares (WCSS).

For 'N' number of observations, if the number of clusters is 'N' then value of WCSS is zero or minimum. For 'N' number of observations, if the number of clusters is one, the value of WCSS is maximum. Hence, we have to find the value of 'k' or number of clusters such that the values of WCSS is low for 'N' number of observations.

In the Elbow method, a graph is plotted where x-axis has the Within Cluster Sum of Squares (WCSS) values and the y-axis has the number of clusters. After the curve is plotted, we have to look for a certain point after which the curve is pretty much flat. This point is known as the elbow point. The value of 'k' (number of clusters) at the elbow point is the ideal value of 'k' for the given data points.

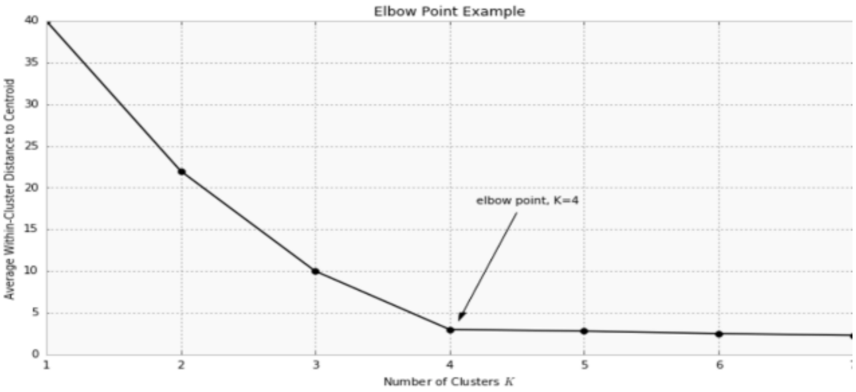

Fig 3.4: Elbow Method Representation

3.4.1 Project Flow Chart.

This will be the proposed flow chart that the system will look like.

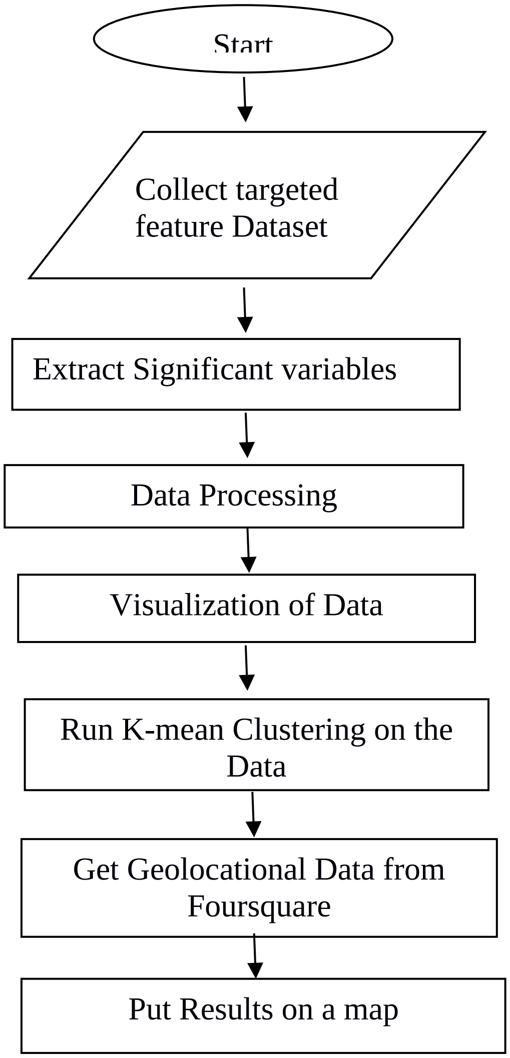
Start
Collect targeted feature Dataset
Extract Significant variables
Data Processing
Visualization of Data
Run K-mean Clustering on the Data
Get Geolocational Data from Foursquare
Put Results on a map

Fig 3.4.1: Algorithm

3.4.2 Data Flow Diagram

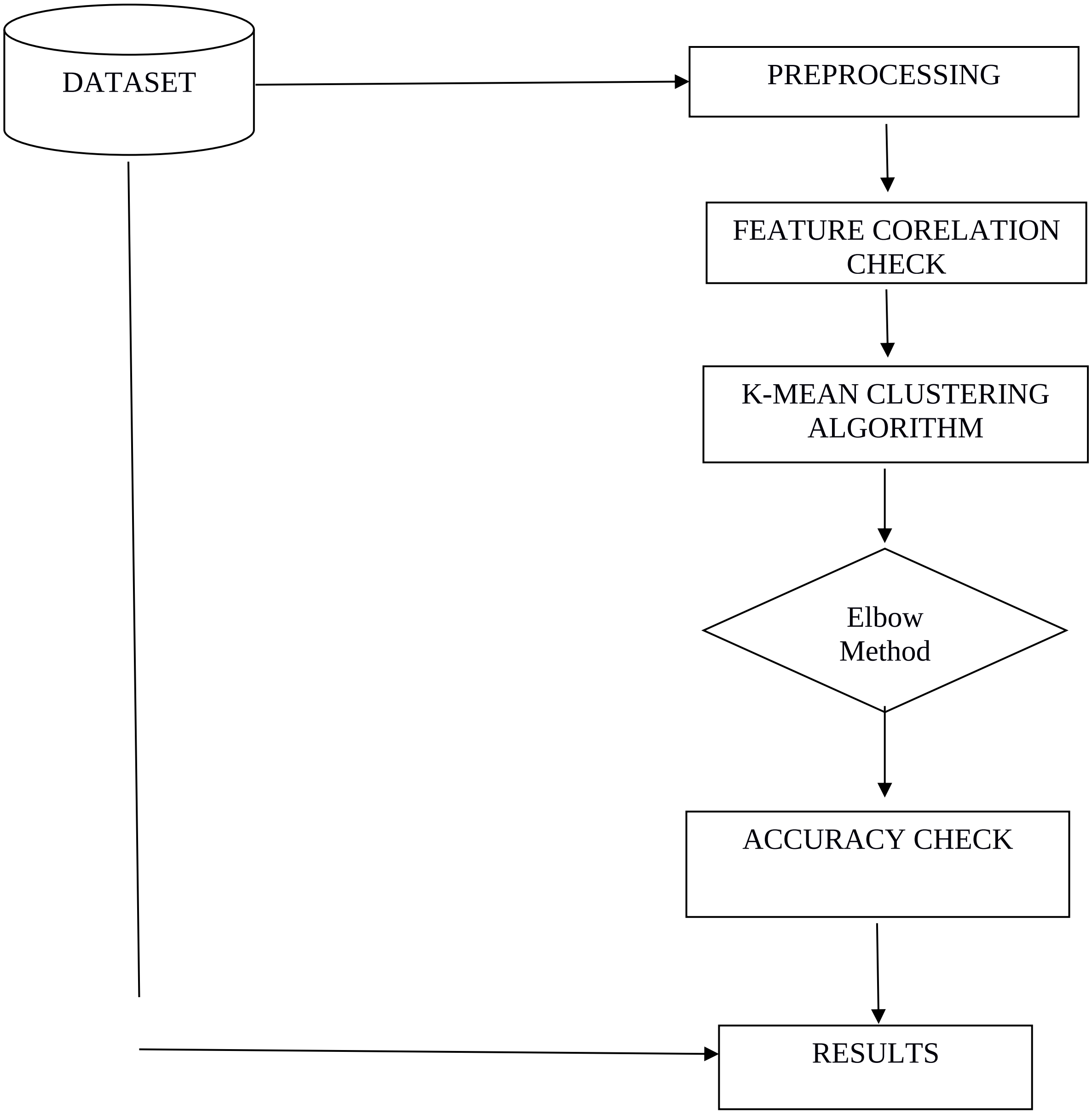

Fig 3.4.2: Dataflow Diagram

04. Research Methodology

4.1 Research Design

We will be using the analytical research design. It is an analysis-based methodology. Specifically using Exploratory Data Analysis. Basically, it is research conducted with analytical approach, where an in-depth case study is done over similar research to understand the domain knowledge. To get insight over which other problem statements have used similar methods and what kind of advantages or disadvantages they had while applying the same. This is more research-based recommendation system. It is an effective method as it is targeting many research case studies and has the end goal of targeting individuals point-of-interests.

4.2 System Development Methodology

The methodology of our system development is the method of managing the project/ ML model development. There are many models of the methodology are available as Waterfall model, Incremental model etc. However, we still need to take our domain under consideration. The same software development models wont accurately fit our machine learning project. So, we have chosen an analytical model for our project. The methodology is useful to manage the project efficiently and able to help developer from getting any problem during time of development. Also, it helps to achieve the objective and scope of the projects. In order to build the project, it needs to understand the domain requirements.

Methodology provides a framework for undertaking the proposed machine learning model. The methodology is a system comprising of steps the help transform a feature rich data set into individual point-of-interest targeting recommender system.

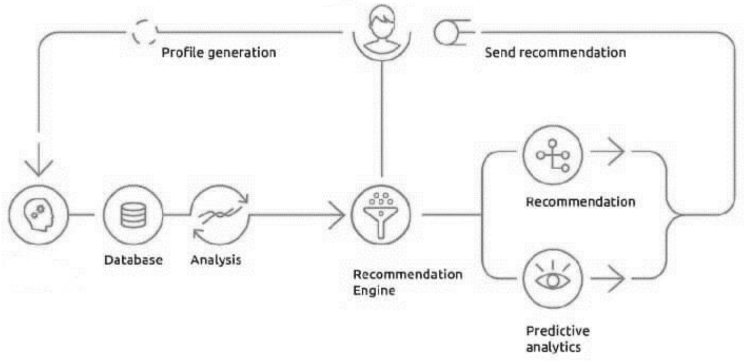

Fig 4.2: Methodology Diagram

There are the following phases in our model:

1)Profile generation: Profile of the user is generated using the data
available in the dataset i.e., preferences of the user from the immigrant preference dataset.

2) Domain knowledge: Domain knowledge consists of the data in the datasets being used for analysis i.e., the preferences of the immigrants (budget, food preferences, amenities, locality etc.) from the immigrant preference dataset and the geolocational data of the city that is present in the geolocational dataset.

3) Analysis: Analyzing the preference dataset and performing clustering analysis on it. Performing Exploratory data analysis (EDA) on the geolocational dataset.

4) Recommendation/Predictive analysis: After performing clustering and Exploratory data analysis (EDA) on the datasets, we recommend certain locations to groups of population which will be ideal place of residence for them.

5) Send recommendation to end user: The final outcome will be a map of the city and certain locations will be highlighted with different colors; each color will be assigned to a population group. This final outcome i.e., the map will be sent to the user.

05. SYSTEM REQUIREMENTS

5.1 Tools

For Model development, the following Software Requirements are:

Operating System: Windows 7/10 or any Linux Ubuntu

Language: Python3

Tools: Jupyter Notebook, Tableau Desktop, Microsoft Excel (optional)

Technologies Used: Exploratory Data Analysis, Python

5.2 Software Requirements

Operating System	Any OS with clients to access the internet
Network	Wi-Fi Internet or cellular Network
GitHub	Versioning Control
Software	Tableau Desktop
Platform	Jupyter Notebook/google colab
Google Chrome	Medium to find references to do system testing, Display and run model

5.3 Hardware Requirements

For Model development, the following Hardware Requirements are:

Processor:	Intel or AMD Ryzen or higher
RAM:	8 GB
Space on Disk:	512 SSD

Device	Any device with internet access

Minimum space to execute:

The effectiveness of the proposal is evaluated by conducting case studies and result accuracy evaluations, configured with an Intel CORE i7-------- processor (GHZ, Cores, RAM, OS)

06. SYSTEM IMPLEMENTATION PLAN

Sr. No.	WORK	MONTH
1	Problem Statement Discussion	August
2	Domain Knowledge gathering	September
3	Literature Survey	September
4	Database Selection	October
5	Data Cleaning and Processing	October
6	Visualization of data	November
7	Study on appropriate algorithm	December
8	Algorithm Implementation	January
9	Rest API connectivity check	January
10	Validation	February
11	Testing	March
12	Deployment	Final Submission

07. Other_Specifications

7.1 Advantages:

A hassle free hassle-free method for immigrants (eg., students who are new to the city) to find a place of residence. For immigrants (students) to manually do research about every area in the city,shortlist locations and then visit every location to find the perfect one is very time consuming andinconvenient. By the help of this project, Immigrants (students) won't have too manually do the overwhelming task of finding a perfect place of residence. No need to use multiple tools or paying agencies to find a place of residence according to your preferences. Currently, there are a very few numbers of tools present and even they are not very accurate or have a very limited scope. Hence immigrants have to use multiple tools or pay a lot of money or do the entire process manually. This project saves the time as well as money of the immigrants (students). People can choose a place of residence that fits in their budget. Hence money can be saved. People can choose a place of residence that is in the vicinity of the amenities and services that they need. Hence, the cost of transportation is saved. Also, time is saved.

People can receive personalized recommendations by selecting the preferences that are more important to them. Analysis is done and recommendations are provided directly to the user. The user can then visit the recommended area to make a final decision on the place of residence rather than having to visit each and every potential location. Hence time and money are saved.

The service providers (such as food and other amenities) will also benefit as the number of customers will also increase. The geolocational dataset we are using includes a lot of information about restaurants, hotels, gyms, schools/colleges and other amenities. As the number of people living in the vicinity of these amenities and services increases, the number of people using these amenities and services will also increase as they would find it convenient. Hence it would be beneficial for both, the customers and the service providers.

Analysis is done and recommendations are provided directly to the user. The user can then visit the recommended area to make a final decision on the place of residence rather than having to visit each and every potential location. Hence time and money are saved.

7.2 Limitations:

Our project works on key skills and concepts like Exploratory Data Analysis, Clustering Algorithms and Recommender systems. In this section we shall discuss the limitations one could face in each of the mentioned phases.

Our chosen methodology of Exploratory Research wont ever be able to replace conclusive, quantitative research. This will always work as a limitation. This methodology gives a broader perspective which helps get an insight instead of a conclusion. The methods use samples, that might not be the most accurate representative since they have not been taken under consideration using probability. For instance, case studies can be selected simply because they represent extreme good or bad situations as opposed to average ones.

Clustering analysis has a few limitations to keep in mind before using it in any system or project. We have used K-mean clustering algorithm in our project and one of the limitations or disadvantage we faced was that it gave varying results on different runs of an algorithm. These random choices of cluster pattern gave varying results. While doing the literature survey we came across limitations like k-Mean clusters tendency to assume the admin deals with spherical clusters having equal observation. This gave us insight in the method of how the algorithm would fail or cause a disadvantage while working on unusual size of clusters.

In conclusion of limitations on the recommender system we learned were a lack of data when we take a cold problem under consideration, the constant change in the domain of research, and when targeting a niche of interest, the constant varying preferences caused a great disadvantage when working with the system.

7.3 Applications:

Will studying a applying Exploratory Data Analysis we came across its effective nature in data manipulation, which helps the user get a good insight in the domain of work, to find the answers and interpretations one would need by discovering data patterns, spotting anomalies, checking assumption and testing out hypothesis. It can help detect errors that one might miss, identify outliers in the database, and understand corelations between targeted factors and attributes. It is majorly used in the field of Data Analytics by industries like Clinical and pharmaceuticals as well as corporate and retail.

Clustering analysis is majorly used in by big industries and companies for their market research, in pattern recognition, predictive and data analysis and image processing as well. It helps the marketers by discovering the right groups or clusters in their customer base hence helping target and improvise sales. In field of biology clustering analysis has helped gain insight into structures of inherent populations. Clustering Analysis if used by Government in the cold problem of land usage could help in allocation and housing predictions giving a probable conclusion. It can further be used in document classifications, in identifying outliers like fraud detection in credit cards.

Finally discussing applications for Recommender System, we understand how it aims to provide the best possible personalized services. It majorly works on targeting the point of interests and give a conclusive answer accordingly. One can observe it applications in 8 major categories like: government services, library systems, tourism industry, shopping and e-commerce, group activities and resource services.

08. CONCLUSION AND FUTURE SCOPE

Through this conclusion we were able to understand the whole end-to-end Machine Learning project, which was made upon a foundation of some key domain skills like Exploratory Data Analysis (EDA), Analytical methodology etc. We study the powerful upcoming technology of Recommender system which values extracting useful attributes from a good targeted dataset and problem statement. These systems can in future help solve many cold problem statements.

In this paper, we studied today problem of immigrant accommodation. We targeted a niche and started with doing some literature survey and Exploratory data analysis on the same. We were able to understand many attributes that can affect the decision while choosing a good suitable accommodation. After understanding these attributes, we were able to cluster then using the K-mean Clustering algorithm. These clusters helped map the appropriate accommodations on the map for the end users better understanding.

Our project helps show how far Exploratory Data Analytics can go to help understand a domain and problem statement. We have presented and evaluated Clustering Analysis as an approach in a recommender system. Our approach holds the promise of allowing clustering analysis to scale bigger datasets and much more complex problem statements at the time of producing an even higher quality recommender system

The project can work as a base for government related approach to relocate population in natural disaster, Immigrant and refugee situations etc. It can be used to target a locality and start improvement and development projects based on the clustering analysis done. We can use it to improve e-commerce business and to start predicting business modules long before the implementations come in scope. It has the potential to have great community impact as the days go by and the accommodation allocation starts becoming a bigger issue.

REFERENCES

1. J. Cao, Y. Dong, P. Yang, T. Zhou, and B. Liu, ‘‘POI recommendation based on meta-path in LBSN,’’ Chin. J. Comput., vol. 39, no. 4, pp. 675–684, Apr. 2016.

2. M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, ‘‘Exploiting geographical influence for collaborative point-of-interest recommendation,’’ in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. (SIGIR), 2011, pp. 325–334.

3. Dietz, L.W., Sen, A., Roy, R. et al. Mining trips from location-based social networks for clustering travelers and destinations. Inf Technol Tourism 22, 131–166 (2020).

4. Stamatelatos, G., Drosatos, G., Gyftopoulos, S. et al. Point-of-interest lists and their potential in recommendation systems. Inf Technol Tourism 23, 209–239 (2021). https://doi.org/10.1007/ s40558-021-00195-5

5. K. Gao et al., "Exploiting Location-Based Context for POI Recommendation When Traveling to a New Region, " in IEEE Access, vol. 8, pp. 52404-52412, 2020, doi: 10.1109/ACCESS.2020.2980982.

6. Singh, Ashwani & Soundarabai, Paulsingh. (2017). Collaborative filtering in movie Recommendation System Based on Rating and Genre. IJARCCE. 6. 465-467. 10.17148/IJARCCE.2017.63107.

7. Xu, Chonghuan & Ju, Chunhua & Qiang, Xiaodan. (2015). Efficient Collaborative Filtering Using Particle Swarm Optimization and K-Harmonic Means Algorithm. Journal of Computational and Theoretical Nanoscience. 12. 6334-6342. 10.1166/jctn.2015.4675.

8. Qing Li and Byeong Man Kim, "Clustering approach for hybrid recommender system," *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, 2003, pp. 33-38, doi: 10.1109/WI.2003.1241167.

9. Sahu, Satya & Nautiyal, Anand & Prasad, Mahendra. (2017). Machine Learning Algorithms for Recommender System - a comparative analysis. International Journal of Computer Applications Technology and Research. 6. 97-100. 10.7753/IJCATR0602.1005.

10. Basu, C., Hirsh, H., and Cohen, W. (1998). Recommendation as Classification: Using Social and Content-based Information in Recommendation. In *Recommender System Workshop '98*. pp. 11-15.

11. Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, Guangquan Zhang, Recommender system application developments: A survey, Decision Support Systems, Volume 74,2015, Pages 12-32, ISSN 0167-9236