International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 71–75

Geographical Location Selection for Entrepreneurs

Palak Vora[1], Sneha Pawar[2], Jigisha Bhole[3], Sushma Nandiyawar[4], Pooja Vengurlekar[5]

Computer Department, Savitribai Phule Pune University

[1]palak.vora19@gmail.com

[2]snehapawar50@gmail.com

[3]bholejigisha97@gmail.com

[4]sushmanandiyawar5621@gmail.com

[5]pooja.v.85@gmail.com

Abstract

Selecting a suitable location to set up any business is a challenging task. Different types of factors must be taken into consideration for every type of business and not to forget the manual labour that goes into collecting the data for the parameters considered. This study presents an efficient system for knowing the probability of profitability of a location based on the type of business to be established. The methods explored include Analytical Hierarchy Process (AHP) and Machine Learning (ML).

Keywords— AHP, Machine Learning

I. INTRODUCTION

Entrepreneurs are critical to the forever changing economy as they hold the reigns to progress. In times of this fast-moving world, spending time in researching for a place may become expensive in terms of time and money. Considering the situation of India, this heritage rich country has been rooted to its culture which date back to ages. This has hindered the process of city planning which has resulted in a not so organised development. Taking that into consideration, the fast-paced advancements are difficult to read and decide on a location which may give good results. Deciding on a location for a business is the first and the most critical facet.

Main aspect is deciding on the parameters to be considered as well as the order of weightages which needs to be assigned. AHP as Multicriteria Decision Analysis (MCDM) can be regarded as the method which helps in deciding the proper weightages through formulas. When the AHP model is integrated with ML, it becomes a system which takes in geographical information and value preferences to give out facts on decision making.

Prioritizing the parameters can be defined as giving an order to the factors considered based on the importance for a particular business. The order is defined by the ratio of affect each of them has on the profitability of the business.

Through this, we address the problem of difficulty in data collection of Indian locations and the fact that there is no available system for location selection. Establishing criteria for decision making is a difficult and responsible task and in the past only a single criteria optimization has been debated over, that single criteria being economic. There have been various location recommendation systems which gives the best location but this may be inconvenient for the entrepreneur as the location may not be feasible for him/her.

ISSN: 2233-7857 IJFGCN

II. ANALYTIC HIERARCHY PROCESS

AHP involves an elicitation process based on the acquisition of pairwise preferences. FCM is a method for multi criteria decision making. Final decision about ranking of the factors under analysis can be obtained from the decision matrix. Further, Fuzzy AHP methodology can be applied to evaluate the factors more precisely. Solving a problem through AHP involves six basic steps.

First step involves identifying the problem and goal of the analysis. With correct recognition of problem, a clear vision to the pathway gets established. Next step is the hierarchy construction where all the attributes are arranged according to their level of goals. It reflects their mutual relationship. Highest level of the hierarchy will consist of the primary goals of the problem, the level just below that will have attributes of the secondary goals. All the lower levels combined, aim to achieve the goal of its consequent higher level. Third step is the Pair-wise comparison in which the attributes are assigned weights. The system of assigning weights is done by comparing them in pairs for every level of hierarchy. The results obtained are stored in comparison matrix. The method for comparing the attribute pairs involves a term called Saaty scale. Saaty scale ranges from 1-9 and was developed by Saaty in 1980, though its usage in AHP was done by Shapira and Simcha in 2009. If the value of Saaty scale is 1 then the attributes are of equal importance and as they go higher in value, the importance increases. The following step that is the fifth step involves synthesising matrix. It means that the values are normalised by dividing each element by the total value of the column. This reduces the errors and fixes the issue of large gaps in values. Fifth step is Prioritising factors in which the outcomes of the synthesized matrix is taken into consideration. This process is the continuation of the previous normalisation process. Priority vectors are obtained by summing the elements in each row and dividing them by the total of the row amount. The vector with the highest value after this process is given the utmost priority and from there it gets lower. Last step involves relative weight computation wherein the eigenvector is calculated.

After the above six steps, another step which the AHP method has is the calculation of the weighted sum matrix. The weighted factor elements of the weighted sum matrix are obtained by multiplying each element of the pair wise comparison matrix by their respective priority vector. The weighted sum is the summation of the elements row wise. Eigen values are calculated by dividing weighted sum by priority vector. Another important term is Consistency Index (CI) which is obtained by the following=g formula where e_{max} is the eigen value and n the number of attributes:

$(e_{max}- n)/(n- 1)$

(1)

Consistency ratio (CR) is a measure of controlling the consistency of pair wise comparisons. It tells how consistent the judgement is if compared to random judgements done on large samples.

CR=CI/RC

(2)

In the above equation RC is the random consistency index.

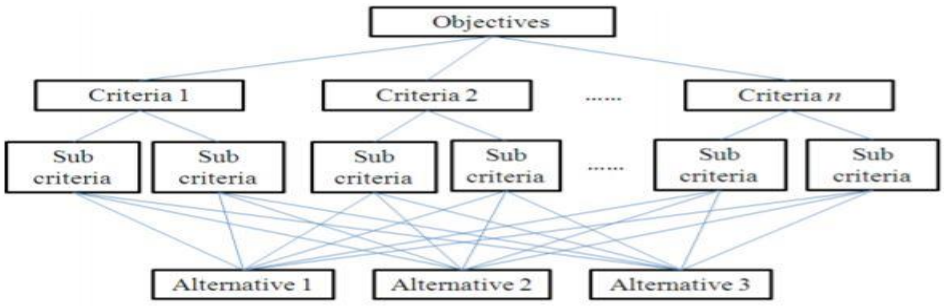

Fig 1. Typical AHP hierarchy structure

ISSN: 2233-7857 IJFGCN

III. STUDY AREA

Pune district of Maharashtra state was taken as study area. With 15.642 sq. km. area, it lies on the western margin of Deccan plateau. This gives rise to various types of terrains within the city which helps in covering many of the aspects of a scenario for business setup. It contains ten talukas and the rapid urbanisation of these very well reflect the situation of Indian cities. The twin cities of Pune and Pimpri-Chinchwad along with the three cantonment areas of Pune, Khadki and Dehu Road form the urban core of the PMR, which also includes seven municipal councils and 842 villages. Approximately 36% of the population lives in 486 slum areas. Pune has a population of more than 5,000,000 with a huge amount of migrating population which resulted in many villages to turn into cities. Of these, 45% slum households do not have in-house toilet facilities and 10% do not have electricity. One third of the slums are on mixed ownership land. The living conditions in slums varies considerably. Pune has a hot semi-arid climate (type BSh) bordering with tropical wet and dry (type Aw) with average temperatures ranging between 20 and 28 °C (68 and 82 °F).[96] Pune experiences three seasons: summer, monsoon, and winter. Typical summer months are from mid-March to mid-June, with maximum temperatures sometimes reaching 42 °C (108 °F). The warmest month in Pune is May. The city often has heavy dusty winds in May, with humidity remaining high. Even during the hottest months, the nights are usually cool due to Pune's high altitude. Pune has the fifth largest metropolitan economy and the sixth highest per capita income in the country.

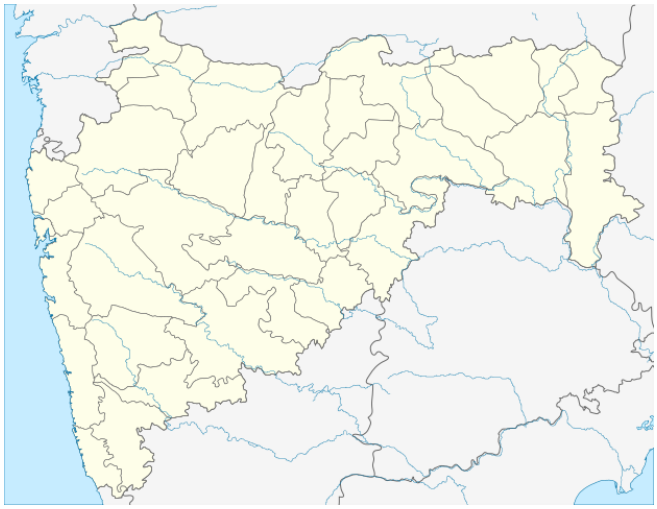

Fig 2. Location of Pune in India

IV. METHODOLOGY

The main objective of this system is to provide an easy to use system which gives the entrepreneurs an idea about how the business will profit if set up at a particular location. For any system, data collection is a crucial phase and the most important one. Collecting data of Indian streets is very difficult and not as easily available as other countries. Therefore, different resources were considered and the validity tested for a proper creation of the dataset. One important resource is Google Places API which gives information on the number of places of a certain kind that surrounds the given location. Some of these places' types include: airport, bakery, colleges, hotels, restaurants, fire station etc. It has features likes geocoding and geolocation which helps in converting location to coordinates and vice a versa as and when needed according to the user entry.

Other data was collected by writing a code to scrape data from different websites. The following step is cleaning the data and organizing the data based on the type of business used. After contemplating all

ISSN: 2233-7857 IJFGCN

the factors that needs to be used and having data for every factor for Pune city, it needs to be validated, and missing values needs to be handled. Any discrepancy in the dataset is looked for and fixed before moving on the next step that is AHP.

MCDM is used for the AHP process where all the factors are given as an input. The problem identified here is to know the importance of factors and how much effect will it have on the profitability of the business. Factors go through the hierarchy construction to determine their levels and eventually we get output as a list of parameters in order of their importance. The initial importance was decided by conducting a survey around the study. Different businessmen were approached and their reviews were taken into consideration for determine which factors must be prioritised over others.

There is one issue with AHP when it comes to large scale implementation and this scalability issue is fixed by AHP but which in turn gives rise to another problem of biases. To reduce the bias of a dominance hierarchy, ML is integrated with AHP.

When ML comes into picture there is a big question of which algorithm to choose and which fits best with the dataset. To solve this problem, different ML models were observed over the dataset and accuracy was calculated. After a thorough study over the behaviour of different ML algorithms with the dataset, the most accurate came out to be Multilinear regression.

TABLE I
COMPARISON OF DIFFERENT ML ALGORITHMS

ML Algorithms	Accuracy (%)
Random Forest	75
MultiLinear Regression	97
Decision Tree	82
SVM	90

Multilinear regression is good for datasets which have a greater number of independent variables to be considered. The dataset goes through this algorithm and many things happen simultaneously. The target variable for the training phase of the algorithm is determined by an equation generated through AHP. As AHP is not dynamic and adaptive to the current environment, ML is combined with it to give a better result. After the model is trained, the dataset reserved for testing is used to acquire the accuracy as well as the rightness of the model. Large number of entries needs to be used for training and the testing data should also have the same structure. The testing phase will allow the model to recalculate the weightages which were initially assigned by the AHP model. The new weightages are updated for the next phase of the model. Every time the model learns and revises its weightages which results in a very efficient, dynamic and adaptable system.

The user side of this will only include entering of a location along with the type of business and internally the factors assigned to the business will be fetched and assigned weightages and the trained model will be loaded for prediction. The output will be in the form of a rating on the scale of 1-5.

V. RESULT AND DISCUSSION

The model was successfully tested for Pune ventures with the owners already established satisfied with the result. ML algorithm training phase with a large dataset took a lot of time to complete but as soon as it got loaded on the server, the process became a little smooth. This system doesn't concentrate on giving fast results but it focuses on giving accurate results without time constraint. The data collection also took majority of the time as Indian census dates back to 2011 which isn't feasible for an

ISSN: 2233-7857 IJFGCN

advanced system. To wait for the next census which happens every 10 years is not efficient and not to forget the ever-changing scenarios. The dataset generation needs to be revaluated time and again and revised to the new standards of the environment which is possible by running the system for data gathering every now and then.

VI. CONCLUSION

This paper revised the already established method AHP and added ML with it to make a more effective system. The factors considered varied from industry to industry and the weightages also varied accordingly. The profitability of the venture was put into a form of rating which was calculated using the factors considered and situations fed into the ML model. Optimization through ML gives an upper hand from all the systems which use monotonous AHP methods.

REFERENCES

[1] M. Koilakuntla, V. S.. Patyal, S. Modgil, P. Ekkuluri, *A research study on estimation of TQM 'Factors Ratings' through Analytical Hierarchy process*, Procedia Economics and Finance, 2012, pp. 55-61, Vol 3.

[2] P. Chaudhary, S. K. Chhetri, K. M. Joshi, B. M. Shrestha, P. Kayastha, *Application of an Analytic Hierarchy Process (AHP) in the GIS interface for suitable fire site selection: A case study from Kathmandu Metropolitan City, Nepal*, Socio-Economic Planning Sciences, 2015, pp. 60-71, Vol 53.

[3] D. Sarkar, S. Shah, "*Evaluating tower crane safety factors for a housing project through AHP*", Journal of Construction Management, 2012.

[4] G. Zhao, T. Liu, X. Qian, T. Hou, H. Wang, X. Hou, Z. Li, "*Antenna placement for landslide monitoring using Analytical Hierarchy Process (AHP) and Geographical Information System (GIS)*," IEEE Symposium on Wireless Technology and Applications, 2013.

[5] P. Avesani, C. Bazzanella, A. Perini, A. Susi," *Supporting the requirements prioritization process. A Machine Learning approach,*" in 16th International /conference on Software Engineering & Knowledge Engineering 2014.

[6] A. Ali, Y. Hafeez, S. Hussain, S. Yang, " *Role of Requirement Prioritization Technique to Improve the Quality of Highly-Configurable Systems,*"in IEEE Access, 2020, pp.27549-27573, Vol 8.

[7] M. H. Satman, M. Altunbey, *Selecting Location of Retail Stores Using Artificial Neural Networks and Google Places API*, Int. J. of Statistics and Probability, 2014, vol. 3.

[8] A. Rikalovic, I. Cosic, D. Lazarevic," *GIS Based Multi-Criteria Analysis for Industrial Site Selection*", 24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013, pp. 1054 – 1063.

[9] R. B. Bhagat, "*Rural – Urban Classification and municipal governance in India*", Singapore Journal of Tropical Geography, 26(1), 2005, pp. 61-73

[10] M. Vlachopoulou*, G. Silleos, V. Manthou," *Geographic information systems in warehouse site selection decisions*", Int. J. Production Economics 71, 2001, pp. 205-212.

ISSN: 2233-7857 IJFGCN