IJIRSS
International Journal of Innovative Research & Scientific Studies

**ISSN:** 2617-6548

URL: www.ijirss.com


# Logistic regression for predicting the location of vegetable vendors in the city of Raipur, Chhattisgarh, India

Sushmita Chakraborty[1*], Abir Bandyopadhyay[2], Swasti Sthapak[3]

[1,2,3]*Department of Architecture, National Institute of Technology Raipur, India*.

Corresponding author: Sushmita Chakraborty (*Email: schakraborty.phd2018.arch@nitrr.ac.in*)

## Abstract

Urban planning plays a pivotal role in guaranteeing the functionality, accessibility, and adaptability of cities to meet the needs of their diverse population and various official and informal economic activities. The primary objective of this work is to investigate the application of machine learning techniques in the identification of optimal places for vegetable vendors within the urban context of Raipur City, India. While logistic regression has been used in previous studies to address issues such as soil erosion, land susceptibility mapping, and identifying potential sites for health facilities and mining exploration, this model has yet to be applied to determining suitable locations for vegetable vendors. This gap in research could be beneficial if addressed, particularly in India, where many city residents rely heavily on vegetable vendors for their dietary needs. The paper's main focus is on evaluating the reliability of the model and encouraging its implementation in similar scenarios, highlighting its efficiency and adaptability, which are also evaluated in this study. A stratified random sampling technique was implemented to collect data from four different regions of Raipur City. Subsequently, the gathered data was subjected to analysis employing the logistic regression machine learning technique, with the objective of making predictions. The results obtained from the analysis were highly impressive, as the model successfully predicted 44 out of the total 50 locations with an accuracy rate of 88%.

**Keywords:** Location, Logistic regression, Machine learning, Model, Prediction, Python, Vegetable vendor.

**DOI:** 10.53894/ijirss.v6i4.2123
**Funding:** This study received no specific financial support.
**History: Received:** 8 June 2023/**Revised:** 4 September 2023/**Accepted:** 22 September 2023/**Published:** 2 October 2023

**Authors' Contributions:** All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
**Competing Interests:** The authors declare that they have no competing interests.
**Transparency:** The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
**Institutional Review Board Statement:** Not applicable.
**Publisher:** Innovative Research Publishing

## 1. Introduction

The urbanization rate in India has experienced a substantial growth during the past few decades [1]. Consequently, there was a substantial increase in the urban population within Indian cities. The urban population predominantly depends on rural areas for purchasing consumable goods, such as food and vegetables [2, 3]. The vegetables are abundantly consumed by the urban population on a daily basis [2], in addition to other perishable food items. The fresh vegetables are mostly sourced from street vendors [3, 4].

The vegetable vendors procure these goods from nearby rural areas, which makes the vegetables fresh, affordable, and in close proximity to the urban population [4]. These street vegetable vendors are an important informal segment as they address the daily nutritional needs of urban areas by providing fresh vegetables to urban dwellers on a daily basis [5, 6]. Therefore, to sustain this business, the strategic location of these vegetable vendors plays a crucial role. The location not only increases the sales of these vendors but also contributes to their personal well-being, like access to basic facilities and amenities [4, 7].

The term "location" refers to the placement of a business or facility in a specific area, and it has a direct impact on the success of any business or commercial setup, particularly on its sales and profit [7]. Therefore, the location plays an important role for any commercial setup to thrive [5, 6].

This has grabbed the attention of urban planners, policymakers, and economists, as it also influences urban and regional development [5]. Several methods and tools are utilized to identify the best location for large-scale formal setups like shopping malls [8, 9], convenience stores [10, 11], local markets [12], etc. In contrast, informal sectors like hawkers and street vendors received very little attention in this respect [7]. Even though the informal sector is an essential segment of the urban milieu of Indian cities, unfortunately, the locational choices of Indian street vendors have not been extensively researched.

The present study deals with the identification of the location of the street vegetable vendor in the Indian context. The government classifies Indian cities into three categories based on population density: Tier I, Tier II, and Tier III. This research focuses on Raipur, Chhattisgarh, categorized as a Tier II city by the Government of India [10].

## 2. Literature

Data analysis is made more efficient through the use of Machine Learning (ML) technology [11, 12]. ML involves the automatic identification of important patterns in data, which eliminates the need for explicit programming. Arthur Samuel claims that ML is a field of study that enables computers to learn without explicit programming and is considered favorable since machines handle data better using software and algorithms [13].

Logistic regression is one such algorithm that examines the relationship between a categorical dependent variable and a set of independent variables. It is primarily used when the dependent variable only has two values, such as 0 and 1 or Yes and No [14, 15]. This study aims to determine the suitability of the logistic regression model for assessing the importance of vegetable vendors' location in urban planning using an ML algorithm. Furthermore, the study seeks to determine if the results can be replicated in other cities with similar characteristics.

Vetriselvi and Thenmozhi [16] identified machine learning as a type of artificial intelligence that aims to create models capable of gaining knowledge through experience [16]. Machine learning is used to gather experience from the environment and predict new cases based on the analysis of previously handled cases [17, 18].

According to Arthur Samuel, Machine learning is the field of study that allows computers to learn without being explicitly programmed [11].

Logistics regression is a technique for supervised classification [19, 20]. The target variable (output) x can only take discrete values in response to a given collection of features (inputs) in a classification issue. It has been found that logistic regression is a more flexible and suitable approach for modelling various scenarios compared to discriminant analysis [21, 22].

According to Sangchini, et al. [23], researchers have utilized the Analytical Hierarchy Process (AHP) and logistic regression for weight calculations and prediction, and it has been applied to land susceptibility mapping [24-29], and healthcare facilities [30, 31], groundwater potential sites [28, 32], bankruptcy [33], soil erosion [34], clinical use [35, 36], and mine exploration [37].

However, its applicability to the locational preferences of vegetable vendors has not been previously studied.

The main objective of this work is to employ Logistic regression in order to forecast the geographical placement of vegetable sellers.

To achieve this goal, Analytic Hierarchy Process (AHP) weights were computed and incorporated into the Logistic Regression Model for predicting vegetable vendors' locations in Raipur.

## 3. Study Area

Raipur, the capital of Chhattisgarh, India, is a II-tier city [10] and a significant regional center for administration, commerce, education, health, and culture. It was established as the capital of the newly formed state of Chhattisgarh in 2000, after its separation from Madhya Pradesh, Refer to Figure 1.

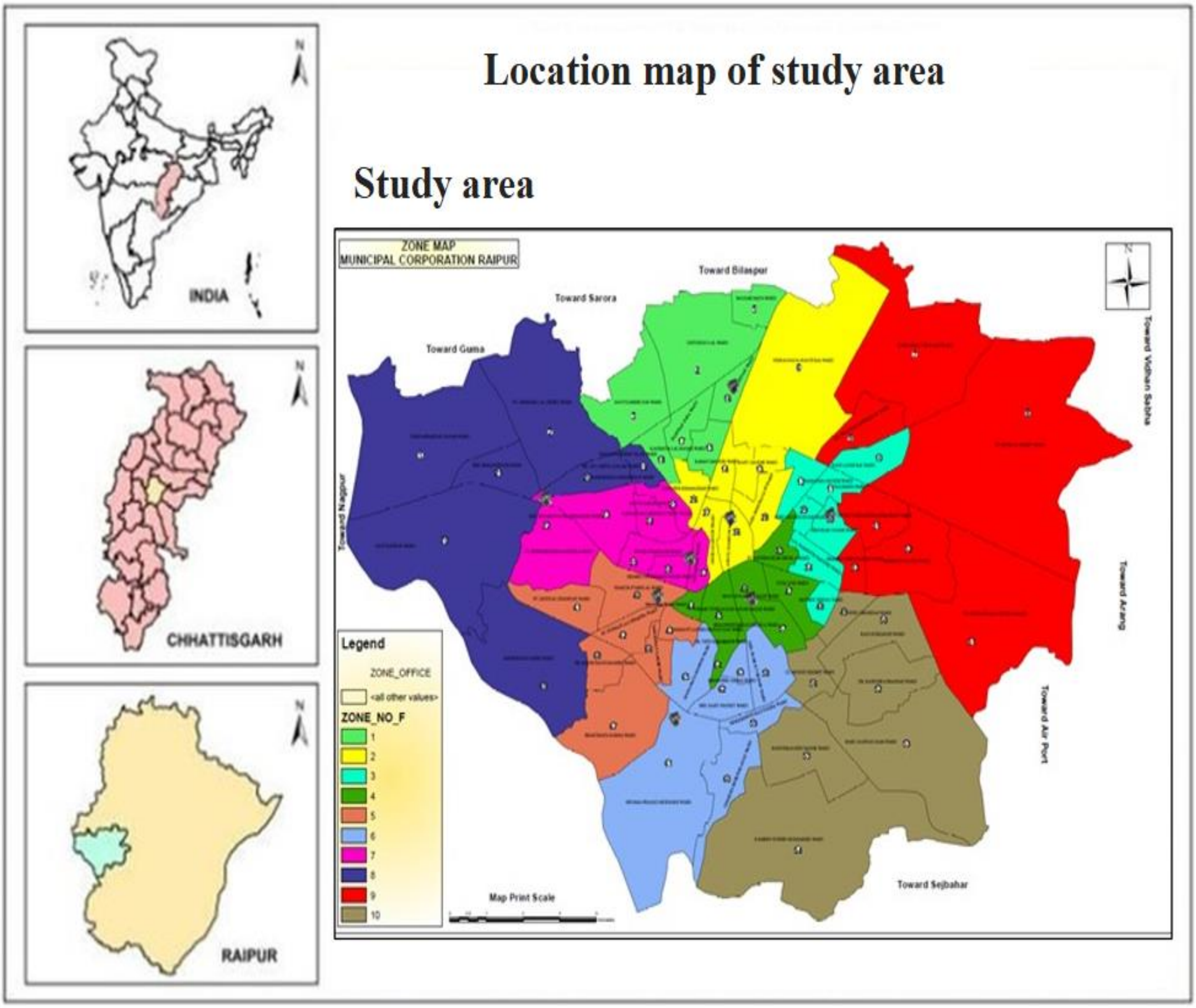


**Figure 1.**
The map shows the (A) Location of Raipur in India map (B) Research area, Raipur city, Chhattisgarh.
**Source:** Das and Swarnakar [38].

The city of Raipur, located at Latitude 21.250000 and Longitude 81.629997, has been selected for the current study. The document of the Government of India [10] includes several tier II cities in India, and Raipur, the capital of Chhattisgarh, is among them. According to the Municipal Corporation of Raipur, there are 7336 street vendors, including fish, vegetable, fruit, jewelry, and fast-food vendors [39]. There are around 1700 vegetable vendors in the category of Raipur's vendors.

## 4. Materials and Method

For the present study, a systematic and structured approach to the research process was followed to ensure the accuracy and reliability of the findings. Overall, the methodology employed in this study is presented in the flow chart in Figure 2.

### *4.1. Data Collection*

The stratified random sampling technique was used based on its effectiveness in heterogeneous populations, where the population is divided into homogeneous groups, or strata [40]. In accordance with Singh and Masuku [40], stratified sampling entails the random selection of units with varying sample sizes in each stratum, determined by their relative significance in the population.

Sampling is then carried out independently in each stratum, with strata or subgroups chosen based on their correlation to the outcome.

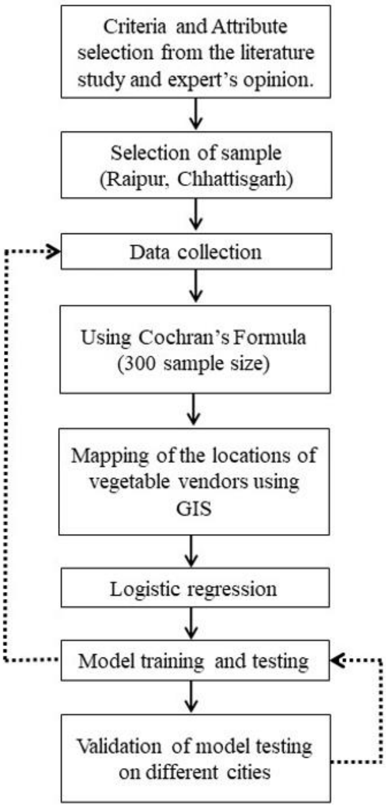


**Figure 2.**
Flow chart of research methodology.

*4.2. Population and Sample Size*

For the study, Raipur City was divided into four pockets with two orthogonal, imaginary lines in the cardinal direction passing through the central business district CBD of the city, with every pocket having its own set of wards (refer to Figure 3).

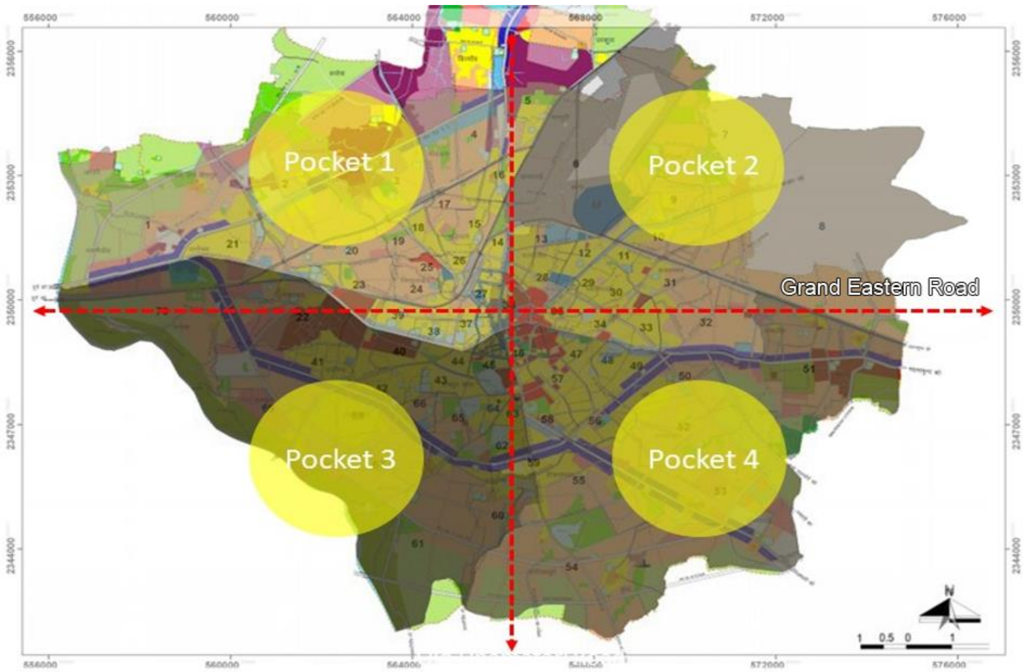


**Figure 3.**
Map displaying the boundaries of Raipur Municipal, with the Grand Eastern Road running horizontally and the Old Dhamatari Road running vertically.

To determine the sample size for the population, Cochran's formula was used, providing a sample size of 300 for the present study. The clear geographical division of the city made it a favorable location for the present context, as it allowed for a better comprehension of the needs and requirements of each section. Additionally, the approach led to a more structured and efficient boundary for survey points. The selection of vendors in Raipur is based on the city's land use [39].

**Table 1.**
The table shows the various categories of pockets, along with corresponding population statistics.

| Pocket | No. of wards | Population [13] | No. of vendors (Sample collection) [13] |
|---|---|---|---|
| Pocket 1 | 19 | 85413 | 56 |
| Pocket 2 | 19 | 92339 | 87 |
| Pocket 3 | 18 | 40189 | 83 |
| Pocket 4 | 14 | 21511 | 74 |
| Total no. of wards = 70 | | Total no. of vendors = 300 | |

**Source:** Government of India [39].

The data from the sample collection is presented in Table 1. The first three pockets have a relatively equal number of wards, whereas pocket 4 has a lesser number of wards. The population data were collected from the municipal corporation of Raipur. Based on the information provided, valuable insights can be offered into the distribution of wards in the area and their demographic profiles.

*4.3. Data Analysis Methodology*

Several supervised ML algorithms like ANN (Artificial Neural Network), Random Forest, Naïve Bayes, and Logistic Regression are utilized for predictions due to their transparency and ease in explaining outcomes [22]. Logistic regression, as a quantitative statistical method for prediction, is employed to study the relationship between a categorical dependent variable and a set of independent variables (explanatory) [29]. A Logistics Regression technique was utilized in this study to create a continuous spatial s-curve model that predicts the placement of vegetable vendors. The model produces values that fall within the range of 0 to 1. If a value is close to 1, it suggests there is a higher probability, whereas a value closer to 0 indicates a greater likelihood of absence [41]. When using logistic regression, the default threshold is set at 0.5. This means that if an object's probability value, given its features and weight set, falls below 0.5, it will be classified as a member of either class 0 or 1 [42]. Hence, in the present study, binary logistics regression can be used to predict a good or bad location for vegetable vendors.

## 4. Result and Findings

To determine the location of the vegetable vendor, a set of attributes and corresponding weights are utilized in a logic function. Table 2 includes all the attributes considered during the selection process.

**Table 2.**
The input data for the logistic regression.

| S. No | Attributes | Data input | Data type |
|---|---|---|---|
| 1. | Accessibility | 1. Very difficult | Categorical |
| | | 2. Difficult | |
| | | 3. Moderate | |
| | | 4. Easy | |
| | | 5. Very easy | |
| 2. | Street light | 1. Yes | Binomial |
| | | 2. No | |
| 3. | Road side dumping | 1. Yes | Binomial |
| | | 2. No | |
| 4. | Water supple | 1. Yes | Binomial |
| | | 2. No | |
| 5. | Type of road | 1. Arterial road | Categorical |
| | | 2. Sub-Arterial road | |
| | | 3. Collector road | |
| | | 4. Local road | |
| 6. | Waste disposal | 1. Yes | Binomial |
| | | 2. No | |
| 7. | Land-use | 1. Residential | Categorical |
| | | 2. Commercial | |
| | | 3. Public/Semi-public | |
| | | 4. Mixed-use | |
| | | 5. Industrial | |
| 8. | Drainage line | 1. Yes | Categorical |
| | | 2. No | |
| 9. | Parking space | 1. Yes | Binomial |
| | | 2. No | |
| 10. | Type of parking | 1. On-street parking | Binomial |
| | | 2. Off-street parking | |
| 11. | Population count | Ward-wise data | Categorical |

A study was conducted using various attributes, including accessibility, street light, roadside dumping, water supply, road type, waste disposal, land use, drainage, parking availability, type of parking, and population count. These attributes were given weights and integrated into the logit function. The method included both training and testing data, which showed underfitting and overfitting with varying scores.

The Google Collaboratory-enabled Jupyter Notebook is where the Python programming code used to create the prediction model is available. To execute the suggested system, the Jupyter Notebook procedures outlined below were followed:

1. To begin, the data is imported into a Pandas data frame using the Load Data method. It is then inspected for any null, missing, or NaN values. Exploratory Data Analysis (EDA) is conducted to analyze and interpret the data [43]. Feature Engineering (FE) is utilized to encode the data and allocate weights based on the input data's scale [44]. FE involves modifying, adding, removing, combining, and transforming the dataset to enhance the model's training and precision.
2. The data is divided into two sets: training data, which comprises 80% of the data, and testing data, which makes up the remaining 20%. A logistic regression model is created using the training data and subsequently verified with the testing data. To check for underfitting or overfitting of data frames, a mock evaluation is conducted specifically for Raipur city. Finally, hyper-parameters are used to optimize the data and remove any underfitting or overfitting.
3. By using the earlier approaches, the final 20% of unlabeled data is projected.

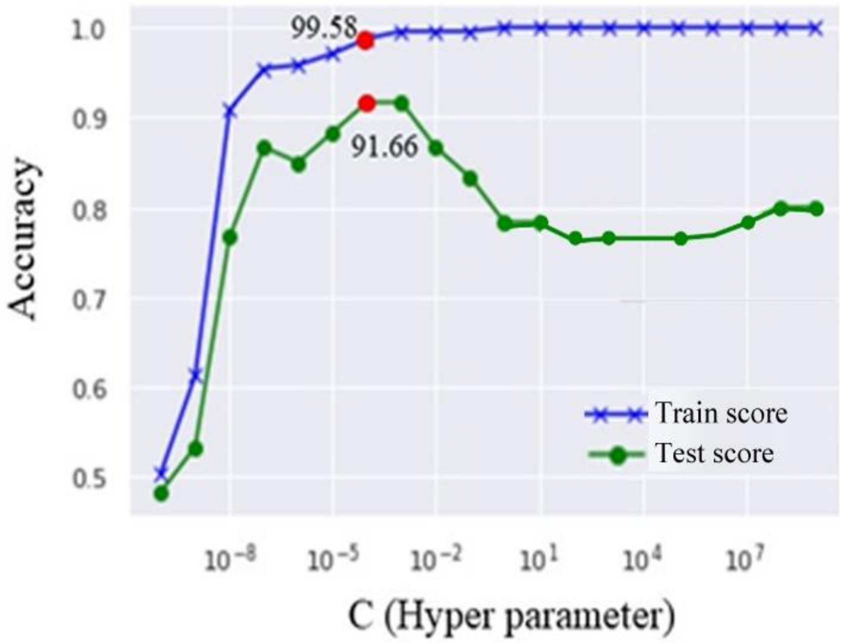


**Figure 4.**
Graph representing the train and test score of the model.

Figure 4 shows Train Score and Test Score for the prediction model. The model was tested using points between these scores, and three regularization parameters (C) were defined with values of $10^{-4}$, $10^{-5}$, and $10^{-6}$ to control the learning process [45]. The best score model is determined by a 5% to 10% difference between the train score and the test score. A 99.16% train score and 90.00% test score were obtained for $10^{-4}$, while $10^{-5}$ gave a 98.33% train score and 90.00% test score. For $10^{-6}$, the train score was 96.66% and the test score was 88.33%, but its prediction score of 78% led to the rejection of the $10^{-6}$ model.

Both the $10^{-4}$ and $10^{-5}$ models were qualified based on their train scores and test scores, as well as their respective prediction accuracy rates. The $10^{-5}$ model was chosen over the $10^{-4}$ model due to its higher precision, achieving a prediction accuracy rate of 88% in 44 out of 50 new locations of vegetable vendors across Raipur City, with a difference of 8.33% between its train and test scores of 98.33 and 90.00, respectively.

In order to validate the model, a study was conducted at 50 new locations of vegetable vendors across Raipur City. Out of these, the model successfully predicted 44 locations with an 88% accuracy rate. Table 3 represents the data used for prediction.

**Table 3.**
Prediction score for the datapoints of the city Raipur.

| Sr. no | Location no. | Class probability | Prediction value (Model)[1] | Prediction (Manually)[2] |
|---|---|---|---|---|
| 1 | 1 | 0.889 | 1 | 1 |
| 2 | 11 | 0.928 | 1 | 1 |
| 3 | 16 | 0.778 | 1 | 1 |
| 4 | 31 | 0.994 | 1 | 1 |
| 5 | 41 | 0.957 | 1 | 1 |
| 6 | 55 | 0.922 | 1 | 1 |
| 7 | 90 | **0.174** | 0 | 1 |
| 8 | 107 | 0.601 | 1 | 1 |
| 9 | 113 | 0.656 | 1 | 1 |
| 10 | 128 | **0.113** | 0 | 1 |
| 11 | 132 | 0.606 | 1 | 1 |
| 12 | 161 | 0.968 | 1 | 1 |
| 13 | 175 | 0.917 | 1 | 1 |
| 14 | 189 | 0.944 | 1 | 1 |
| 15 | 196 | 0.954 | 1 | 1 |
| 16 | 200 | 0.933 | 1 | 1 |
| 17 | 209 | **0.466** | 0 | 1 |
| 18 | 220 | 0.547 | 1 | 1 |
| 19 | 231 | 0.828 | 1 | 1 |
| 20 | 240 | 0.827 | 1 | 1 |
| 21 | 253 | 0.758 | 1 | 1 |
| 22 | 272 | 0.902 | 1 | 1 |
| 23 | 288 | 0.907 | 1 | 1 |
| 24 | 294 | 0.744 | 1 | 1 |
| 25 | 295 | 0.942 | 1 | 1 |
| 26 | 301 | 0.216 | 0 | 1 |
| 27 | 302 | 0.095 | 0 | 1 |
| 28 | 305 | 0.136 | 1 | 1 |
| 29 | 312 | 0.517 | 1 | 0 |
| 30 | 318 | 0.305 | 0 | 1 |
| 31 | 328 | 0.123 | 0 | 1 |
| 32 | 335 | 0.025 | 0 | 1 |
| 33 | 341 | 0.095 | 0 | 1 |
| 34 | 349 | 0.016 | 0 | 1 |
| 35 | 358 | 0.606 | 1 | 0 |
| 36 | 362 | 0.182 | 0 | 1 |
| 37 | 372 | 0.575 | 1 | 0 |
| 38 | 373 | 0.561 | 1 | 0 |
| 39 | 376 | 0.414 | 0 | 1 |
| 40 | 385 | 0.371 | 0 | 1 |
| 41 | 386 | 0.153 | 0 | 1 |
| 42 | 389 | 0.152 | 0 | 1 |
| 43 | 390 | 0.282 | 0 | 1 |
| 44 | 391 | 0.050 | 0 | 1 |
| 45 | 400 | 0.007 | 0 | 1 |
| 46 | 406 | 0.428 | 0 | 1 |
| 47 | 407 | 0.175 | 0 | 1 |
| 48 | 414 | 0.045 | 0 | 1 |
| 49 | 447 | 0.176 | 0 | 1 |
| 50 | 458 | 0.023 | 0 | 1 |
| | | | Total prediction = 50 | Correct prediction = 44 |

In analyzing the data presented in Table 3, it is apparent that the use of logistic regression in predicting location was largely successful. With an accuracy rate of 88%, the model was able to predict the location of 44 out of 50 instances accurately. This suggests that the method can be applied in similar settings to achieve dependable predictions. These results demonstrate the potential of using logistic regression in location prediction and highlight its usefulness in various industries.

A total of 50 data points were used to validate the prediction model, as shown in Figure 5. In Figure 5, there are two data points, one in orange and the other in blue. The orange data point represents correct predictions, while the blue data point represents incorrect predictions. The accuracy in predicting 44 of the data points proves the efficiency of the model.

[1]Prediction value (Model) = The values "Presence" and "Absent" at a particular location are represented by the numbers 1 and 0, respectively.
[2] Prediction (Manually) = The values of Correct and Wrong predicted by the model are represented by the numbers 1 and 0.

The graph displays the different data points, with those above 0.5 reflecting a more desirable location prediction. Conversely, those situated below 0.5 suggest a less favorable location [42]. Hence, the value of 0.5 serves as the threshold line for the logistic model. Overall, the model was precise and efficient in its predictions.

## 5. Conclusion

According to a recent study, when choosing a location, vegetable vendor consider a number of aspects, such as waste management, land use, drainage, parking accessibility, population size, accessibility, closeness to services, and road condition. The model utilized in the research region demonstrated a high level of accuracy, correctly predicting 88% of the observed results. It is recommended that more testing be conducted in comparable regions to enhance the generalizability of this approach, since it has the potential to yield benefits for similar urban locations.

The utilization of model-generated data analysis can offer significant advantages to those involved in urban planning, development, and decision-making processes. It can particularly benefit those seeking to propose more effective locations for vegetable vendors, especially in densely populated areas like Raipur. This approach benefits both the vendors and the communities they serve, ultimately leading to more sustainable and equitable urban development. Moreover, this tool can be adapted for other city types or informal sectors, providing further opportunities for future development. Research in these areas can assist in the advancement of such tools.

Research has the potential to enhance the advancement of similar technologies in the domains of urban planning and market analysis. The model that has been previously examined demonstrates efficacy as a planning tool in the identification of appropriate vendor sites, particularly within the context of urban development. This can provide significant advantages for vendors and other stakeholders, including consumers, suppliers, local authorities, and the healthcare industry. The utilization of this technology can provide valuable information that can assist businesses and organizations in remaining informed about market trends, improving the comfort of customers throughout the buying process, and enhancing general accessibility. The practical use of this instrument is evident, and its incorporation has the potential to provide substantial advancements in the realms of market analysis and urban planning.

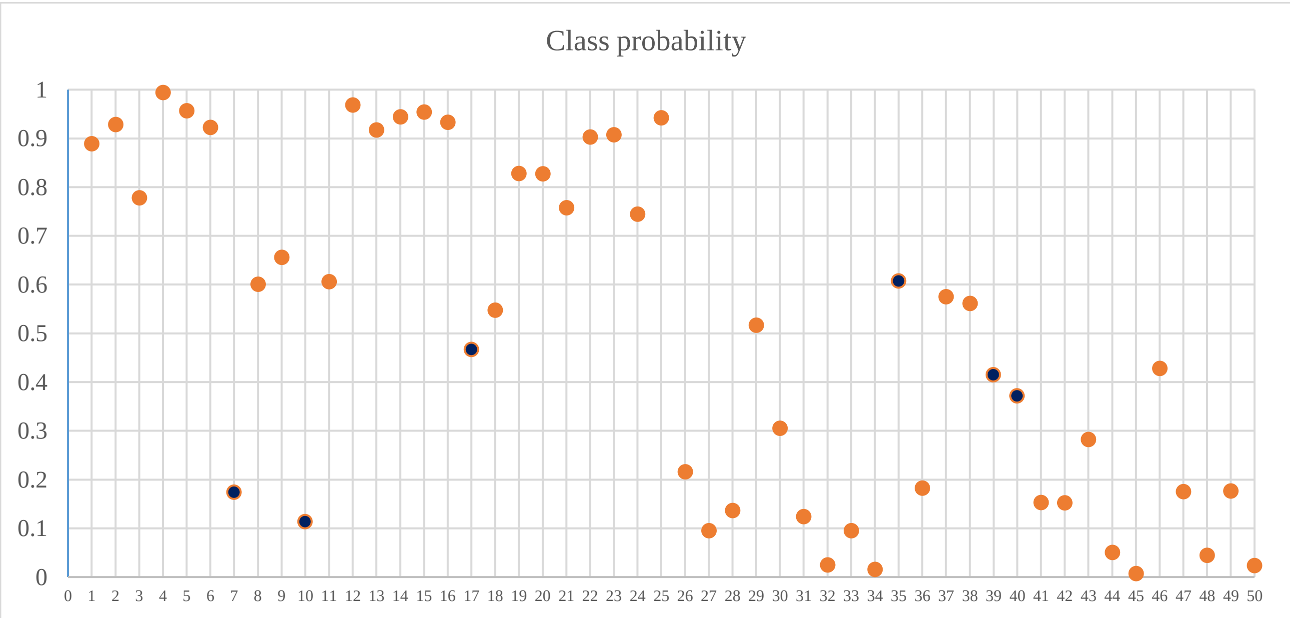


**Figure 5.**
Graph representing the prediction values with the point in two different colors.
**Note:** 1. Orange (Correct prediction).
2. Blue (Wrong prediction).

## References

[1] K. Mouratidis, "Urban planning and quality of life: A review of pathways linking the built environment to subjective well-being," *Cities*, vol. 115, pp. 1–12, 2021. https://doi.org/10.1016/j.cities.2021.103229

[2] Government of India, "Household consumption of various goods and services in India, 2011-12," Retrieved: http://164.100.161.63/sites/default/files/publication_reports/Report_no558_rou68_30june14.pdf. 2014.

[3] G. Turner, R. Green, C. Alae-Carew, and A. D. Dangour, "The association of dimensions of fruit and vegetable access in the retail food environment with consumption; a systematic review," *Global Food Security*, vol. 29, pp. 1-11, 2021. https://doi.org/10.1016/j.gfs.2021.100528

[4] Y. Li *et al.*, "Assessing the role of access and price on the consumption of fruits and vegetables across New York city using agent-based modeling," *Preventive Medicine*, vol. 106, pp. 73-78, 2018. https://doi.org/10.1016/j.ypmed.2017.10.014

[5] H. Erbıyık, S. Özcan, and K. Karaboğa, "Retail store location selection problem with multiple analytical hierarchy process of decision making an application in Turkey," *Procedia - Social and Behavioral Sciences*, vol. 58, pp. 1405–1414, 2012. https://doi.org/10.1016/j.sbspro.2012.09.1125

[6] A. A. El Karim and M. M. Awawdeh, "Integrating GIS accessibility and location-allocation models with multicriteria decision analysis for evaluating quality of life in Buraidah city, KSA," *Sustainability Switzerland*, vol. 12, no. 4, pp. 1-28, 2020. https://doi.org/10.3390/su12041412

[7] M. Płaziak and A. I. Szymańsk, "Importance of personal factor in decisions on locating enterprises," *Procedia-Social and Behavioral Sciences*, vol. 110, pp. 373-380, 2014. https://doi.org/10.1016/j.sbspro.2013.12.881

[8] M. Fusková, D. Hanáčková, and M. Gubáňová, "Location factors and their importance in location decision making of enterprises in conditions of slovak republic case study," *Economic Management Innovation*, vol. 10, no. 1, pp. 5-18, 2018.

[9] N. Shrestha, "A study on the street vendors of Kathmandu Municipality," *Pertanika Journal of Social Sciences & Humanities*, vol. 21, no. 2, pp. 725–734, 2013.

[10] Government of India, *Classification of cities*. New Delhi: Government of India, 2015.

[11] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research*, vol. 9, no. 1, pp. 381-386, 2020.

[12] I. Ivanov, B. Toleva, and V. J. Hooper, "A fast and effective approach for classification medical data sets," *International Journal of Innovative Research and Scientific Studies*, vol. 6, no. 3, pp. 545-552, 2023. https://doi.org/10.53894/ijirss.v6i3.1580

[13] G. Wiederhold and J. McCarthy, "Arthur samuel: Pioneer in machine learning," *IBM Journal of Research and Development*, vol. 36, no. 3, pp. 329–331, 2010.

[14] C. Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, 2002.

[15] S. Park, S. Jeon, S. Kim, and C. Choi, "Prediction and comparison of urban growth by land suitability index mapping using GIS and RS in South Korea," *Landscape and Urban Planning*, vol. 99, no. 2, pp. 104-114, 2011. https://doi.org/10.1016/j.landurbplan.2010.09.001

[16] D. Vetriselvi and R. Thenmozhi, "A systematic literature review on deep learning based medical image segmentation," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 5s, pp. 519-526, 2023.

[17] H. Tan, "Machine learning algorithm for classification," presented at the In Journal of Physics: Conference Series, 2021.

[18] S. S. Dash, S. K. Nayak, and D. Mishra, "A review on machine learning algorithms," *Smart Innovation, Systems and Technologies*, vol. 153, pp. 495–507, 2021.

[19] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: Classification and comparison," *International Journal of Computer Trends and Technology*, vol. 48, no. 3, pp. 128-138, 2017.

[20] M. Praveena and V. Jaiganesh, "A literature review on supervised machine learning algorithms and boosting process," *International Journal of Computer Applications*, vol. 169, no. 8, pp. 32-35, 2017. https://doi.org/10.5120/ijca2017914816

[21] T. Haifley, "Linear logistic regression: An introduction," IEEE International Integrated Reliability Workshop Final Report, 2002.

[22] G. Tripepi, K. Jager, F. Dekker, and C. Zoccali, "Linear and logistic regression analysis," *Kidney International*, vol. 73, no. 7, pp. 806-810, 2008.

[23] E. K. Sangchini *et al.*, "Assessment and comparison of combined bivariate and AHP models with logistic regression for landslide susceptibility mapping in the Chaharmahal-e-Bakhtiari Province, Iran," *Arabian Journal of Geosciences*, vol. 3, no. 9, pp. 1-15, 2016. https://doi.org/10.1007/s12517-015-2258-9

[24] S. Park, C. Choi, B. Kim, and J. Kim, "Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea," *Environmental Earth Sciences*, vol. 5, no. 68, pp. 1443-1464, 2013. https://doi.org/10.1007/s12665-012-1842-5

[25] H. Shahabi, S. Khezri, B. Bin Ahmad, and M. Hashim, "Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models," *Catena Amst*, vol. 115, pp. 55–70, 2014.

[26] A. M. Youssef, "Landslide susceptibility delineation in the Ar-Rayth area, Jizan, Kingdom of Saudi Arabia, using analytical hierarchy process, frequency ratio, and logistic regression models," *Environmental Earth Sciences*, vol. 73, no. 12, pp. 8499-8518, 2015. https://doi.org/10.1007/s12665-014-4008-9

[27] W. Chen, X. Xie, J. Peng, J. Wang, Z. Duan, and H. Hong, "GIS-based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models," *Geomatics, Natural Hazards and Risk*, vol. 8, no. 2, pp. 950-973, 2017. https://doi.org/10.1080/19475705.2017.1289250

[28] D. Josephs-Afoko, S. Godfrey, and L. C. Campos, "Assessing the performance and robustness of the UNICEF model for groundwater exploration in Ethiopia through application of the analytic hierarchy process, logistic regression and artificial neural networks," *Water Sa*, vol. 44, no. 3, pp. 365-376, 2018. https://doi.org/10.4314/wsa.v44i3.04

[29] H. M. Do, K. L. Yin, and Z. Z. Guo, "A comparative study on the integrative ability of the analytical hierarchy process, weights of evidence and logistic regression methods with the flow-R model for landslide susceptibility assessment," *Geomatics, Natural Hazards and Risk*, vol. 11, no. 1, pp. 2449-2485, 2020. https://doi.org/10.1080/19475705.2020.1846086

[30] M. P. LaValley, "Logistic regression," *Circulation*, vol. 117, no. 18, pp. 2395-2399, 2008.

[31] S. Khanmohammadi and M. Rezaeiahari, "AHP based classification algorithm selection for clinical decision support system development," *Procedia Computer Science*, vol. 36, pp. 328-334, 2014. https://doi.org/10.1016/j.procs.2014.09.101

[32] O. Rahmati, A. Nazari Samani, M. Mahdavi, H. R. Pourghasemi, and H. Zeinivand, "Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS," *Arabian Journal of Geosciences*, vol. 8, no. 9, pp. 7059–7071, 2015. https://doi.org/10.1007/s12517-014-1668-4

[33] R. P. Hauser and D. Booth, "Predicting bankruptcy with robust logistic regression," *Journal of Data Science*, vol. 9, no. 4, pp. 565-584, 2011. https://doi.org/10.6339/jds.201110_09(4).0006

[34] S. Kachouri, H. Achour, H. Abida, and S. Bouaziz, "Soil erosion hazard mapping using analytic hierarchy process and logistic regression: A case study of Haffouz Watershed, Central Tunisia," *Arabian Journal of Geosciences*, vol. 8, no. 6, pp. 4257–4268, 2015. https://doi.org/10.1007/s12517-014-1464-1

[35] E. Christodoulou, J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel, and B. Van Calster, "A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models," *Journal of Clinical Epidemiology*, vol. 110, pp. 12-22, 2019. https://doi.org/10.1016/j.jclinepi.2019.02.004

[36] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, "Developing prediction models for clinical use using logistic regression: An overview," *Journal of Thoracic Disease AME Publishing Company*, vol. 11, no. S4, pp. S574–S584, 2019. https://doi.org/10.21037/jtd.2019.01.25

[37] F. Hoseinali, R. Farajloo, M. Rajabi, F. Hosseinali, and M. Ali RAJABI, "Weighting information layers using logistic regression for mine exploration applications," *Shaping the Change*, vol. 23, pp. 8–13, 2006.

[38] B. Das and A. K. Swarnakar, "Mobility of heavy metals in soils and wastewater from landfill, Sarona, Raipur, Chhattisgarh, India," *Journal of University of Shanghai for Science and Technology*, vol. 22, no. 10, pp. 372–381, 2020.

[39] Government of India, "Raipur municipal corporation. Street Vendors," Retrieved: https://raipur.gov.in/district-statistics-book-2020/. [Accessed Jul. 19, 2023], 2022.

[40] A. S. Singh and M. B. Masuku, "Sampling techniques & Determination of sample size in applied statistics research: An overview," *International Journal of Economics, Commerce and Management*, vol. 11, no. 2, pp. 1–22, 2014.

[41] O. Cokluk, "Logistic regression: Concept and application," *Educational Sciences: Theory and Practice*, vol. 10, no. 3, pp. 1397-1407, 2010.

[42] S. Handoyo, Y.-P. Chen, G. Irianto, and A. Widodo, "The varying threshold values of logistic regression and linear discriminant for classifying fraudulent firm," *Mathematics and Statistics*, vol. 9, no. 2, pp. 135-143, 2021. https://doi.org/10.13189/ms.2021.090207

[43] K. Sahoo, A. K. Samal, J. Pramanik, and S. K. Pani, "Exploratory data analysis using python," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 4727–4735, 2019.

[44] T. Verdonck, B. Baesens, M. Óskarsdóttir, and S. vanden Broucke, "Special issue on feature engineering editorial," *Mach Learn*, 2021. https://doi.org/10.1007/s10994-021-06042-2

[45] D. S. N. Ulum and A. S. Girsang, "Hyperparameter optimization of long-short term memory using symbiotic organism search for stock prediction," *International Journal of Innovative Research and Scientific Studies*, vol. 5, no. 2, pp. 121-133, 2022. https://doi.org/10.53894/ijirss.v5i2.415