

# Election Result Prediction using Spatial Statistical Method

R. Vinodha, R. Parvathi

**Abstract:** Since 17th Century, Modern representative democracy in all forms is being made possible by having a formal way of collecting people's overall opinion on who is to be chosen as their representative. Election is a most important event in any democratic country. Confronting on predicting the winning person/party in the election is the biggest challenge here. In this paper, a study of spatial statistics method to predict the election result using the exploratory data analysis is carried out by applying spatial statistical methods. The paper presents, prediction of the election result with respect to the income level of the citizen and also a comparison of results obtained from the different statistical methods.

**Index Terms:** Geographic Information System, Residual Auto-Correlation, Spatial Regression, Spatial Statistical Methods

## I. INTRODUCTION

Election is a democratic process which helps people of a country; select an individual or group of individuals from a party to hold positions in government bodies and public ministries. In general 'To elect' refers to choosing someone or something, it can either be a person or a decision Election are commonly held to fill positions in public bodies, however it is also conducted in business and private sectors. Analyzing such a vital process and predicting who will win can help the people on a greater margin. A good election can make a nation flourish in all aspects or make a nation into a land of suffering. In this paper, election results are analyzed and predicted using spatial statistical method.

Spatial statistics is a study of statistical methods that can be used in spatial relationship and spaces such as area, distance, length, height, volume, orientation, etc. It contains different types of analysis such as pattern analysis, surface prediction, surface modeling, spatial auto-correlation, spatial regression, statistical modeling and prediction of spatial interaction. Many types of spatial statistics exist comprising of geo-statistics, ecology, climatology, economics, earth science, environment, epidemiology and so on. This paper covers two types of spatial data models such as Geo-statistical data (location-time model, exploratory data analysis, spatial regression) and provincial data (spatial auto-correlation and spatial auto-regression). Using these techniques an approximate election result can be calculated. Prediction of election results is most important in terms of finding the chances of winning along with chances of losing of candidate in all the places. The prediction of election result using spatial statistical methods is carried forward using R tool coding. The rest of the paper is structured into 6 sections. Section 2

discusses about related work, section 3 about spatial statistics techniques used, section 4 presents work carried out by us, section 5 shows our implementation and result. Finally, section 6 holds our conclusion about the work.

## II. RELATED WORK

The vital role of 'numbers' spreads its influence in all places. The essence of usage of statistical method comes in whenever and wherever handling with 'huge' numbers. Prediction of the election result is not only featured on finding the 'probability' of who will win, but also in 'huge numbers' which represents the required data to find the probability. The dataset here holds parameters like income level, distance, longitude and latitude of the region using which we used to actually predict the election result. This section holds the work done on spatial data which is wide ranging and also very varied. The reason for choosing only specific papers for discussion is quite simple since the analysis is constrained to papers drafted on spatial statistics for prediction and its methods. The most common approaches of spatial statistics which have been widely used are Linear Regression (LR) [6,10], Geographic Information System (GIS) [3,9,14,16, 17], Moran's [13], Local Indicators of Spatial Association (LISA) and Spatial Empirical Bayes (SEB) [22] for prediction. The reason for choosing LR is to inference to be more significant than the correlation coefficient [8]. A similar result was given by [6] using geographical parameters as a latent variable with spatial LR. Spatial LR is also found to be independent of the spatial coordinates used, which makes the results even more reliable. A wise choice is to option for model whose parameters are predictable and reliable [6,10,21], hence SLR is definitely a good option for election result prediction. Six specific categories of GIS-based spatial analysis, such as basic measurement, features, classification of quantitative data, density analysis, overlay polygons, overlay lines are calculated primarily involving measurements such as shape, length, distance and quantifying features have been successfully obtained in [16].

GIS are combined with spatial statistics to predict risk factors of disease pattern of disease and to identify clusters [3,9,14,17] and have been successful too. Auto core is another significant statistical method as they help in finding global and local index of spatial auto core thought the global and local indicators obtained from the indexes. Hence, both local and global indicators obtained helps to identify the risks within a region and also among all regions [12,17,22]. Each technique has its own advantage and disadvantage; hence sometimes more than one technique might be of essence to determine the main factors.

**Revised Manuscript Received on July 22, 2019.**

**R.Vinodha**, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.

**R.Parvathi**, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.

In this paper, a few statistics techniques have been explored to predict election results based on varying parameters as discussed in section 3.

III. SPATIAL STATISTICS

Spatial data mining is the concept of identifying interesting relationships and feature aspects which exists implicitly in spatial databases. It is more difficult to identify and extract useful patterns in the dataset that correspond to the purpose of existence of complex aspects such as spatial relationship, spatial data and spatial autocorrelation. Spatial statistics like prediction, regression, pattern analysis, shape analysis, statistical modeling and prediction of spatial communication are utilized for different classes of analysis.

Spatial autocorrelation can calculate as well as evaluate degree of dependence amongst observation in an environmental space or examine either the observed value in one location is independent of the neighboring locations. It consists of three types such as positive, negative and null spatial autocorrelation as shown in Fig. 1. Positive spatial autocorrelation specifies that related values are grouped as one cluster map. Negative spatial autocorrelation specifies that unrelated values are grouped as another cluster map and related values are distributed as another. Null spatial autocorrelation specifies that randomness exists in spatial pattern.

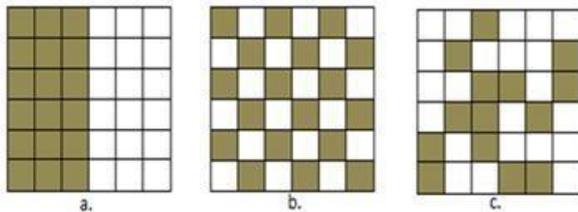


Fig. 1 Patterns of Spatial Autocorrelation (a. Positive spatial autocorrelation, b. Negative spatial autocorrelation and c. Null spatial autocorrelation)

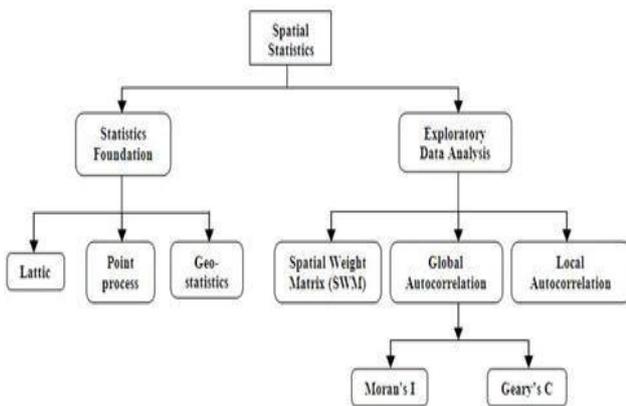


Fig. 2 Spatial Statistics Techniques

Few statistical techniques are used in this paper such as statistics foundation and exploratory data analysis techniques as shown in Fig. 2.

A. Statistics Foundation

Mainly statistical models are required for approximation, prediction and description depending upon probability theory. Few types of problems are discussed in the spatial statistical model such as lattice, point process and geo-statistics.

Lattice – Lattices data are observations associated with spatial regions. It mainly depends on defining a neighbourhood structure, testing for spatial autocorrelation, fitting spatial linear models and model selection.

Point process– Spatial point process gives the locations of objects in a spatial region. It mainly depends on nearest-neighbour techniques, intensity estimation and simulating point process. Geo-statistics – It involves data and non-random sample from one realization of a random function.

B. Unit Exploratory Data Analysis

Exploratory spatial data analysis is a group of models for examining, envisioning, identifying spatial outliers and patterns of spatial autocorrelation. It depends on two different types of autocorrelation such as global autocorrelation and local autocorrelation methods. In this work, the election result is analyzed with the help of spatial autocorrelation using R-tool to get the Spatial Weight Matrix (SWM) and also to measure the global and local spatial autocorrelation measures using Moran's I, Geary's C and Local Moran's I methods.

1) Spatial Weight Matrix (SWM)

The basic requirement of exploratory data analysis is to build a Spatial Weight Matrix (SWM) to represent the spatial relationships among the attributes as represented in (1). Here, W is the Spatial Weight Matrix, n is region number, while i and j are nearby relationship such as  $W_{ij} = 1$  or  $W_{ij} = 0$ .

$$\begin{matrix} W_{11} & W_{12} & W_{1n} \\ W_{21} & W_{22} & W_{2n} \\ W_{n1} & W_{n2} & W_{nn} \end{matrix} \quad (1)$$

If region  $W_{ij} = 0$  it means no nearby relationship is present with itself. Spatial matrix is mainly used to measure the distance.

2) Global Autocorrelation

Global Autocorrelation finds the presence of spatial trends in the circulation of a geographically variable total space. Moran's I and Geary's C statistic approaches are employed to quantify global spatial autocorrelation.

For Moran's I statistics calculation, cross-product is built on changes from the mean of the two location values. The calculation uses Moran's I index and depending on p-value and z-score the significance of that index is preserved. Both z-score and p-value depend upon standard normal distribution. Moran's I statistic is expressed in (2).

$$I = \frac{N \sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i \sum_j w_{ij} \sum_i (X_i - \bar{X})^2} \quad (2)$$

where N is the number of observations (points or polygons),  $\bar{X}$  is mean of variables,  $X_i$  is value at a specific location,  $X_j$  is value at another location,  $w_{ij}$  is a weight indexing location of i relative to j. Moran's I rules can change between -1 and 1.



A value near 1 depend upon similar attributes to be clustered and a value near -1 depends upon dissimilar attributes not to be clustered. If it is zero, it reflects a random pattern.

$$C = \frac{(N-1)\sum_i \sum_j w_{ij}(x_i - x_j)^2}{2w \sum_i (x_i - \bar{x})^2} \quad (3)$$

For Geary's C statistics calculation, the cross-product makes use of actual values at each location and ranges from 0 (maximal positive autocorrelation) to a positive value for high negative autocorrelation. Geary's statistics is expressed in (3) and also comparison of Moran's I and Geary's C is shown in I.

### I. Comparison of Moran's I and Geary's C

Statistic Methods	Null Hypothesis (Non-Significant) Z - value	Alternative Hypothesis (Significant)	
		Z > 0	Z < 0
Moran's I	No Spatial Auto-correlation	Positive Spatial Auto-correlation	Negative Spatial Auto-correlation
Geary's C	No Spatial Auto-correlation	Negative Spatial Auto-correlation	Positive Spatial Auto-correlation

### 3) Local Autocorrelation

LISA – Local Indicators of Spatial Autocorrelation values are calculated for every observation unit. Moran's I is most commonly used is LISA. Local correlation Moran index is measured as in (4).

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1} w_{ij}(x_j - \bar{x})}{m_2} \quad (4)$$

where  $x_i$  is the value of the variable at jth location,  $w_{ij}$  is indicating weight between i and j,  $m_2$  is average of the mean value. The comparison of global and local autocorrelation is shown in II.

### II. Comparison of Global and Local Autocorrelation

Data Analysis Techniques	Description
Global Autocorrelation (Moran's I)	Moran's I calculates the global spatial autocorrelation. Moreover, local or global autocorrelation is aspect similarity which depends on the attribute and location values to evaluate the pattern whichever it is spatial autocorrelation like positive spatial autocorrelation (Cluster), negative spatial autocorrelation (disperse) or null spatial autocorrelation (random).
Local Autocorrelation (Moran's I)	Local Moran's statistic is adopted as a spatial association local indicator that is calculated for isolated areas around all observations in a characterized neighborhood to recognize similar or dissimilar pattern in adjacent.

### C. Spatial Regression Model

This model deals with predicting one variable (Y is the dependent variable) with additional variable (X - the independent variable). Linear Regression is used to analyze linear relationships among variables.

The residual is the difference between the observed and predicted values the Ordinary Least Squares (OLS) are used to find the line that minimizes the total squared distance from the line to our observed values in (5).

$$Y = a + bX + \epsilon \quad (5)$$

where Y is the dependent variable, a is the intercept (value of Y when X = 0), b is the regression coefficient or slope of the line, ε is the residual (error). A spatial log is a variable that average neighbouring values of location are given in (6).

$$Y = \rho W y + x\beta + \epsilon \quad (6)$$

where ρ is spatial autocorrelation parameter, Y is a vector of observations on dependent variable, x is observation on explanatory variable, W is for neighbourhood matrix, β is a vector of regression co-efficient and ε is a vector of unobservable error.

## IV. PROPOSED WORK

In previous works, election results are predicted using fuzzy logic [18]. Autocorrelation methods can be employed to get more accurate results in voting systems. Such methods make prediction much easier and accurate. Prediction of election results with autocorrelation methods takes place with the help of Distance Conversion such as Creation of polygon centroid matrix, Executing spatial statistical methods, Plotting the local autocorrelation, Spatial Regression model and Spatial Auto-regression model as shown in Fig. 3.

Distance conversion is the process of calculating the distance from one location to another in terms of kilometers. Locations are chosen randomly from the data set and the distance is measured. The distance conversion method of converting kilometer to degrees (500km) and converting the degrees to kilometers (1 degree) is performed and based on the analysis of results nearby locations can be identified.

With the significant locations identified by the distance conversion process, a matrix of polygon centroids is created. This helps to analyze the specific election area and plot the connections. Plotting enables coverage of the entire area and makes the analysis process easier.

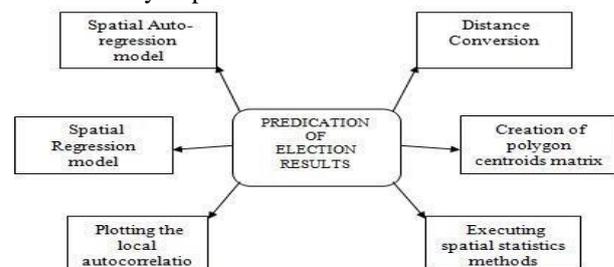


Fig. 3 Predication of Election Result Using Spatial Autocorrelation

The two spatial statistical methods are employed in predicting election results such as Moran's I and Geary's C autocorrelation methods as already discussed. On analysis results clearly depict that Moran's method provides better results compared to Geary's C. Moran's I scatter plot provides a clear idea of the area selected based on the statistical results. Apart from global techniques such as Moran's I and Geary's C methods, local autocorrelation also predicts election results. One of the best statistical measurements, namely Z-score has been employed to predict the level of voting (high, medium or low) in locations.

By the usage of the linear regression method and with variable area (Bust\_pct) and income status (pcincome), results are derived. The regression model is used to find the co-efficient between two variables like estimate, standard error; t-value is used to analyze these values. To overcome the issues of spatial autocorrelation, the spatial autoregressive model is used instead of OLS models to derive better results. By employing such a method, efficient results were arrived with variable area (Bust\_pct) and income status (pcincome) and coefficient values are predicted as shown in VI.

**V. IMPLEMENTATION AND RESULTS**

In this paper election result [23] have been analyzed and predicted by spatial statistical techniques in R tool. R tool is a well known statistical tool for analyzing data; an effective data handler and has storage facility. It is an open source (available at www.r-project.org) and provides a large collection of statistical functions (time-series analysis, graphical techniques, classical statistical tests, linear and nonlinear modelling classification, clustering, etc.) and is also highly extensible. Election dataset is taken with the attributes of the NAME, STATE\_NAME, STATE\_FIPS, CNTY\_FIPS, AREA, FIPS and FIPS\_NUM. It contains 7 rows and 69 columns.

**A. Distance Conversion**

The first step in prediction of election results is the calculation of the distance between one location and another location based on distance conversion. Bust\_pct data is used to calculate the distance required to cover the nearby areas.

**B. Creation of polygon centroids matrix**

Next step is the creation of matrix for polygon centroids, to identify the location and to plot the connection. Plotting enables coverage of the entire area and makes the analysis process easier.

**C. Executing spatial statistical methods**

Moran's I spatial statistics method is tested under randomization. The correlation between Bust\_pct and weights is calculated on the election result. Alternative hypothesis are greater in Moran's method. To Geary's method also same data are provided to calculate the measure.

The alternative hypothesis provides greater expectation than a statistic in this method. Based on statistics obtained Moran scatter plot has been plotted as shown in Fig. 4.

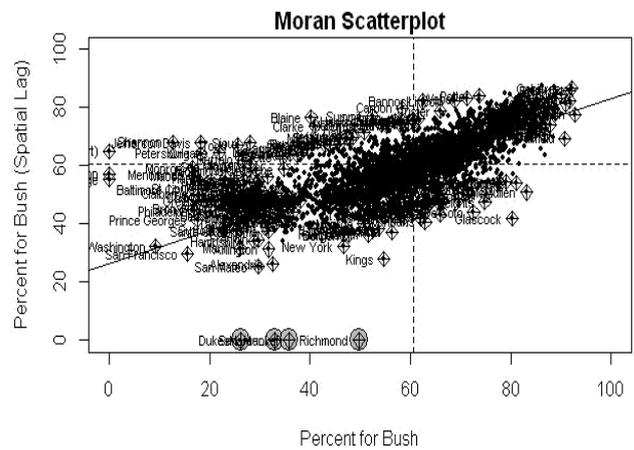


Fig. 4 Moran's Scatter plot

III. Represents the autocorrelation statistics table consisting of Moran's I and Geary's C coefficients. In global Moran's I average value is 51.7308, that indicates a high positive spatial autocorrelation such as highly clustered pattern.

**III. Spatial Autocorrelation Statistics**

Spatial Autocorrelation Statistics					
Coefficient	Statistics	Expected	Variance	Std. dev	p-value
Moran's I	0.55651	-0.0003	0.00011	51.7	2.2e-16
	742	219575	58676	308	
Geary's C	0.42040	1.00000	0.00013	50.3	2.2e-16
	685	00000	24606	593	

**D. Plotting the local autocorrelation**

Local Moran's autocorrelation extracts z-score with the help of statistic methods. Level of voting can be identified clearly from the results.

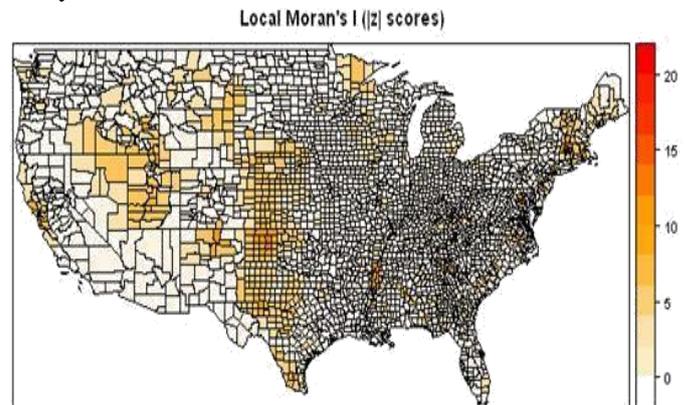


Fig. 5 Local autocorrelation: local Moran's

From the Fig. 5 it is clear that the level of voting is higher in regions of (15% to 20%), the regions with (5% to 15%) depicts that the voting level is medium. The regions with (0% to 5%) are those with low voting levels.

**E. Spatial Regression model**

Income status analysis in Bust\_pct is carried out using spatial regression model. Two variables: Bust\_pct and pcincome are used to predict the income level of the particular region.



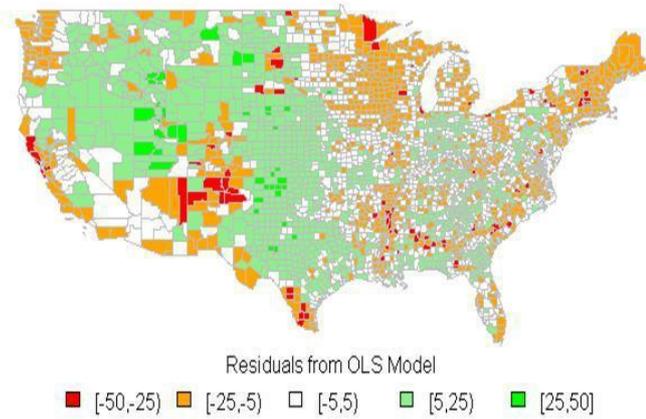


Fig. 6 Residuals from OLS Model

In Fig. 6, residuals from the OLS model is used to classify the result into five class intervals with two predictive variables (Bush\_pct and pcincome). [-50,-25] class depicts the not linearly mapping of predicted variable. [-25,-5], [-5,5], [5,25] and [25,50] are closely moving to high linear mapping with respect predictive variable. The results of these regions under analysis are predicted using the spatial regression model consisting of Moran's and Geary's coefficients and are shown in IV.

IV. Spatial Regression Model

Spatial Regression Model ( Residual Autocorrelation)					
Coefficient	Statistics	Expected	Variance	Std. dev	p-value
Moran's I	0.587534 54	-0.0003 219575	0.000112 2927	55.4 748	<2.2e-1 6
Geary's C	0.413443 01	1.000000 0000	0.000744 359	21.4 99	<2.2e-1 6

F. Spatial Auto-regression (SAR) Model

The spatial autocorrelation with the maximum likelihood approach of the fitted SAR model gives the statistical result. As an additional indicator a spatial weight matrix is prepared during the comparison of SAR parameters with the OLS model. In asymptotic standard error is used to measure the uncertainty. Large sample size gives small standard error. In our work, the number of observations is considered as 3111 and the error value is 0.014316 which shows that the uncertainty is less. Therefore, all the income level people would go to polling. The p-value is a positive value and highly significant, indicating strong spatial autocorrelation on the dependent variable. An eminent difference is there in the variable EI's significance level. The income values are predicted, followed by analyzing the different notations of residuals in the SAR model as shown in Fig. 7.

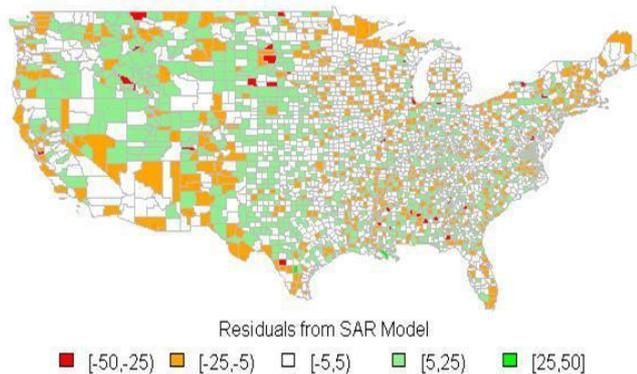


Fig. 7 Residuals from SAR model

The results of spatial auto-regressive model consisting of Moran's and Geary's coefficients for the region under analysis are shown in V.

V. Spatial Auto-Regression Model

Spatial Auto-Regression Model ( Residual Autocorrelation)					
Coefficient	Statistics	Expected	Variance	Std.dev	p-value
Moran's I	-0.0409 754776	-0.0003 219575	0.000115 739	-3.7788	0.999 9
Geary's C	1.00688 23305	1.000000 0000	0.000154 467	-0.5538	0.710 1

This regression model checks the goodness of fit using the R-squared values, analyses the pattern of residuals and hypothesis testing. Statistical significance can be checked with F-test of overall fit, followed by t-test of individual parameters. The predictor variable of income and voting level are significant and the p-value is less than the common alpha level of 0.05, which indicates that it is statistically significant.

VI. Comparison of Co-Efficient Between SR and SAR

	Co-efficient (Intercept)		Co-efficient (pcincome)	
	SR (Spatial Regression)	SAR (Spatial Auto-Regression)	SR (Spatial Regression)	SAR (Spatial Auto-Regression)
Estimate	6.343e+01	1.4073e+01	-1.592e-04	5.4600e-05
Std. Error	8.893e-01	1.0572e+00	4.832e-05	3.3758e-05
t value	71.329	13.3118	-3.294	1.6174
Pr (> t )	< 2e-16	<2e-16	0.000997	0.1058

P-value represents the probability that the observed spatial pattern was created by some random process. If p-value is small means it is very unlike (observed pattern is randomized). The relation between z-value and p-value is as follows. 1) Very high or very low z-value with very small p-value has the significance clustering of data. 2) Very high or very low z-value with very high p-value has randomness in the data. In our result, z-score values are positive, which reveals the spatial distribution of the highest values and or low values in the dataset is more spatially clustered according to the income level verses voting.

VI. CONCLUSION

The election data are analyzed using the following factors: Income level, Distance, Latitude and Longitude. The polygon centroid matrix is constructed to find the contiguity neighbors. Prediction results are analyzed by Moran's I and Geary's C methods. Result of Moran's I and Geary's C methods is presented in the scatter plot to identify the locations where the probability of winning or losing in the election is found. In addition to that, regression and correlation analysis is also being carried out to find the correlation between the voting levels with respect to the income level of the citizens. From the above observations, it is found that the income level of the people and their election polling status are positively correlated to each other.



This election result prediction can help a person/political party/organization to change their election plans beforehand once they sense the prediction. If the results are in their favor they can work even more determined towards carrying forward their election mottos. If the predicted results are not in their favor the candidate/party can make some changes in their plan to coax the people to vote for them. An election can either make or break a nation, hence it is very essential to analyze and predict the election results.

### REFERENCES

1. A. Anselin, "Local indicators of spatial association LISA", *Geogr Anal* 27: 97 – 115, 1995.
2. L. Anselin, A.K. Bera, R. Florax, M.J. Yoon, "Simple diagnostic tests for spatial dependence", *Reg Sci Urban Econ* 26: 77 –104, 1996.
3. M. I. Bakker, M. Hatta, A. Kwenang, M.P. Van, W.R. Faber, P.R. Klatsner, L. Oskan, "Risk factors for developing leprosy a population-based cohort study in Indonesia", *Lepr Rev* 77: 48 – 61, 2006.
4. J.S.Cramer, "Efficient grouping, regression and correlation in Engel curve analysis", *J am Stat Assoc*, 59, 233–250, 1964.
5. N.A.C. Cressie, *Statistics for Spatial Data*, rev. Ed. New York: John Wiley & Sons, Wiley Ser Prob Stat, 1993.
6. N. Cressie, S.N Lahiri, "Asymptotics for REML estimation of spatial covariance parameters", *J Stat Plan Infer* 50, 327–341, 1996.
7. M.C De Souza Dias, G.H Dias, M.L Nobre, "The use of Geographical Information System (GIS) to improve active leprosy case finding campaigns in the municipality of Mossoró", *Rio Grande do Norte State, Brazil, Lepr Rev* 78: 261 – 269, 2007.
8. G. Firebaugh, "A rule for inferring individual-level relationships from aggregate data", *Am Sociol Rev* 43, 557–572, 1978.
9. E. Fischer, D. Pahan, S. Chowdhury, J. Richardus, "The spatial distribution of leprosy cases during 15 years of a leprosy control program in Bangladesh: An observational study", *BMC Infect Dis* 8: 126, 2008.
10. A.S. Fotheringham, "Scale-independent spatial analysis", Taylor & Francis, pp.221-228,1989.
11. M. Goodchild and S. Gopal (Eds.), *The Accuracy of Spatial Data Bases*, pp. 221–228, London: Taylor & Francis, 1989.
12. J. Goncalves Barreto, D. Bisanzio, et.al, "Spatial Analysis Spotlighting Early Childhood Leprosy Transmission in a Hyperendemic Municipality of the Brazilian Amazon Region", *PLOS Neglect Trop D*, Volume 8, Issue 2, 2014.
13. D. Gujarati D, "Basic Econometrics", Fourth edition. New York: McGraw-Hill, *Economet J*, 2014.
14. L. Huang, M. Kulldorff, D. Gregorio, "A spatial scan statistic for survival data", *Biometrics* 63: 109 –118, 2007.
15. M. Kulldorff, N. Nagarwalla, "Spatial disease clusters: detection and inference", *Stat Med* 14: 799 – 810, 1995.
16. P.A. Longley, M. F Goodchild, D.J. Maguire, D.W. Rhind, "Geographic Information Systems and Science", *Int J Remote Sens*, pp.560, 2011.
17. A. Odoi, S.W. Martin, P. Michel. J. Holt, D. Middleton, et al, "Geographical and temporal distribution of human giardiasis in Ontario, Canada", *Int J Health Geogr* 2: 5, 2003.
18. Harmanjit Singh, Gurdev Singh, Nitin Bhatia, "Election Results Prediction System based on Fuzzy Logic", *Int J Comput Appl*, Volume 53-No.9, 2012.
19. S.J. Prais, J. Aitchison, "The grouping of observations in regression analysis", *Revue de l'Institut International de Statistique*, 1–22, 1954.
20. J. Queiroz, G. Dias, M. Nobre, M. De Sousa Dias, S. Araujo, et al, "Geographic information systems and applied spatial statistics are efficient tools to study Hansen's disease (leprosy) and to determine areas of greater risk of disease", *Am J Trop Med Hyg* 82: 306–314, 2010.
21. S. Richardson, C. Guihenneuc, V. Lasserre, "Spatial linear models with autocorrelated error structure", *J R Stat Soc, The Statistician*, 41, 539–557, 1992.
22. W. Tobler, "Frame independent spatial analysis", In M. Goodchild and S. Gopal, *The Accuracy of Spatial Data Bases*, pp. 115–122. London: Taylor & Francis, 1989.
23. L.A. Waller, C.A. Gotway, "Applied Spatial Statistics for Public Health Data", Wiley Ser Prob Stat, 520p, 2004.

### AUTHORS PROFILE



**R. Vinodha** received the B.E degree for Computer Science and Engineering from Dr Pauls Engineering College, Anna University in 2011 and M.Tech degree in Networking from Sri Manakula Vinayagar Engineering college, Pondicherry University in 2013. Now Currently working in NIT Puducherry karaikal She has 6 years of working experience which includes 4 years as research associate in Vellore Institute of Technology, Chennai and 2 year as Assistant Professor in NIT Puducherry Karaikal. She pursues her interest in Data Mining for her Ph.D. degree from Vellore Institute of Technology, Chennai.

**Dr. R. Parvathi** completed her doctoral degree from Anna University, Chennai, India, by contributing his ideas to the field of Spatial data mining. She has a teaching experience of over 20 years in the field of computer applications. Her research interests include data mining, Recommendation Systems and social network analysis. She has authored articles in Bigdata analytics for renowned publications.