



Regression as Potential Tool of City Ranking

P. Srujana, J. Dayanika

CSE, CMR Technical Campus,
Kandlakoya, Hyderabad, India

Abstract - *Cities are shaped by the future, so they too can influence the future. Improving governance at the city level improves the quality of life in cities, making them better places in the aspect of quality of life. In order to improve cities performance, there is an urgent need for a single comprehensive system for measuring and monitoring city performance and urban quality of life. Performance measurements are performed based on various city indicators. Ranking programs intends to help cities maximize their investment in performance monitoring by providing a framework to facilitate consistent and comparative collection and sharing of city indicators. In this paper, we propose that regression is an efficient technique to be used for city ranking.*

Keywords: *Regression, City ranking, City indicators, tools of ranking, regression analysis*

I. INTRODUCTION

1.1 Need for Ranking

Population in cities is increasing rapidly. Along with population, city size, pollution, environment is also changing rapidly. Hence, there comes the need for city ranking which can indicate the overall performance of the city and the areas to be developed [1]. People compare cities with one another in order to analyse the facilities in various cities masters as well as lacks. Comparison of cities also allows government to identify the facilities those city lack to provide and improve city further. Hence, to improve overall city development according to time, comparison between cities is essential.

1.2 City Indicators

In order to compare cities, common factors of those cities should be considered. Such factors of cities which indicates the development, environment, education, employment, and other facilities, hence these factors are referred as city indicators. Each city has its own set of indicators to measure city performance. Improving governance at the city level improves the quality of life in cities, making them better places to live, conduct business, and visit. There is an urgent need for a single comprehensive system for measuring and monitoring city performance and urban quality of life that would [1]:

- Enable elected officials, city managers, and the public to monitor the performance of cities over time;
- Facilitate comparisons across cities; and
- Provide enhanced government accountability demanded by policy makers and the public.

1.3 Data Mining and Its Use in City Ranking

Data mining is defined as process of analyzing data from different perspectives and summarizing data into useful information. The data is gathered from various sources. In order to analyse collected city indicators, data mining is used. Mining techniques provide efficient ways to collect and analyse data and retrieve meaning information. Hence to measure city performance using indicators, data mining provides efficient techniques.

II. EXISTING RANKING METHODS

There are many ranking systems and agencies that ranks cities periodically based on city indicators. These systems or agencies generally considers one or two indicators and assigns ranks accordingly. For example, Healthy City Indicators (HCI) [2] [4] initiated by WHO European Healthy Cities Network (WHO-EHCN) considers indicators those related to health, and ranks cities according to their health status and facilities for health maintenance of people. This system does not consider any other factors of cities. The main drawback of such systems is, a city which is rich in providing health facilities can be poor in providing education facilities. Hence such ranking is not efficient to consider overall development of cities.

In order to overcome this drawback, World Bank in collaboration with Japanese Trust Fund, initiated Global City Indicators Facility program [5].

A. Global City Indicators Facility (GCIF)

The proposed Global City Indicator Program focuses on cities with populations over 100,000. The indicators are initially proposed for the first, most direct level of municipal government – usually cities. A mechanism to accommodate

and aggregate metropolitan urban agglomerations still needs to be developed. This program consists of the 86 core, supporting, and desirable future Global City Indicators and seven indices. More detailed indices are proposed to address some emerging municipal issues such as energy use, competitiveness, or “subjective wellbeing.”

The Global City Indicators are a suite of 93 indicators and indices which are divided into 22 “themes” [6] which are organized into two broad categories: city services and quality of life aspects.

The **city services** [6] include- education, finance, governance, recreation, social services, transportation, wastewater, energy, fire and emergency services, health, safety, solid waste, urban planning, and water. The **Quality of life services** include- civic engagement, culture, economy, environment, shelter, social equity, shelter, technology and innovation and subjective well-being. These indicators are selected based on various screening techniques which include evaluation surveys and various screening criteria. The proposed Global City Indicator Program is envisioned as a decentralized, city-driven initiative

B. Limitations

The major limitations of GCIF system are:

- Number of city indicators considered is high in number.
- To gather all the city indicators is highly costly and complex.
- Data is not easily available and may not be reliable.
- Maintenance such large set of parameters is complex task.
- Updation of these data is always an extra burden and time consuming.
- Continuous monitoring is required for such systems which includes additional efforts.
- As data is not freely available only few organizations/ government bodies can use this system.
- As the cost is more, many cities did not join GCIF program.
- All cities may not gather all the indicator values; hence it doesn't serve the purpose of standardized indicator set used for ranking purpose.

III. DATA MINING and REGRESSION

In existing systems, indicators are collected randomly through questionnaire, online surveys, discussions etc., these methods of data collection follows opinions of particular group of people, which may not be efficient for different group. Hence data collected may not be efficient. Different cities follow different methods to collect information making it difficult to compare cities. In order to overcome these problems, analysis using data mining techniques is proved efficient.

In data mining, regression is used to analyse indicators. Regression technique is generally used to establish relationship between variables.

Linear regression is a technique which is used to define relationships between independent and dependent variables. Hence it is important to define the dependent and independent variables in the considered data set. Here in the indicator list, **dependent variable** is “**competitive score**” and remaining all are independent variables.

Linear regression is a simple approach to supervised learning. It assumes that the dependence of Y on $X_1, X_2 \dots X_p$ is linear [3].

We assume a model

$$E(Y | X) = \alpha + \beta_1 X_1 + \dots + \beta_p X_p,$$

Where, α is called the intercept and the β_j are called slopes or coefficients.

Before starting the regression procedure, below assumptions are made regarding the data.

They are-

- a. Dependent variable is continuous.
- b. Error terms are independent from each other and identically distributed.

Process of applying regression contains following steps:

- a. **Linearity test:** This test is carried out to check the linearity of the data. It can be done by using PLOT procedure. The output of the PLOT procedure is a graph. By observing the pattern of the graph it can be declared whether data is linear or not.
- b. **Correlation test:** There is the possibility of existence of correlated variables in a cluster. If two variables are correlated it would be sufficient to use any one of the two in the regression model for efficient results. This is done with help of CORR procedure.
- c. **Creating the model:** After selecting the required variables, to create the model REG procedure is used.
- d. **Test for assumptions:** To test whether the assumptions made hold correct or not, DW and SPEC options of REG procedure is used.
- e. **Writing the equation:** REG procedure output is compared and the equation is written.
- f. **Testing for multi co-linearity:** Multi co-linearity between the X terms of regression is checked using VIF (Variance Influence Factor) option.
- g. **Testing for outliers:** Outliers will have a major effect on the output. Hence it is important to detect outliers and handle them correctly. To identify the outliers INFLUNCE and R options of REG procedure is used.

- h. Testing the Fit of a model:** Model fitness is generally checked by observing the F static and R-Square values of the REG procedure option. If dataset contains ‘duplicate’ values then LACKFIT option of REG procedure is used for checking the model fitness.

IV. REGRESSION as POTENTIAL TOOL for CITY RANKING

Regression is an efficient data mining technique, which defines relationship between variables. Here, indicators are considered as variables. By applying regression, indicators which have significant effect on overall development of cities can be identified efficiently. Hence, overhead of collecting large number of indicators and their maintenance cost is reduced largely. As indicator list will be specified analysis performed will be efficient and results will be accurate. Hence ranking of cities will be done efficiently using regression.

V. CONCLUSION

As importance of cities is increasing rapidly, city ranking plays an important role in calculating the overall development of cities periodically. Existing ranking techniques fail to rank cities based on their overall performance. Some rank based on one or two factors efficiently ignoring all other factors. GCIF considers all factors, but the complexity of its maintenance and data collection is very high and complex and it is not accepted widely. Hence, ranking system which is implemented using regression techniques is efficient when compared to other systems. It is cost effective, result oriented and easily maintainable. Hence it can be concluded that regression technique is appropriate tool for city ranking.

REFERENCES

- [1] Daniel Hoornweg and Fernanda Ruiz Nunez, “City Indicators: Now to Nanjing”.
- [2] Premila Webster and Denise Sanderson, “Healthy City Indicators- A sustainable Instrument to Measure Health”, *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 2012.
- [3] Leslie A. Christern, “Introduction to Building a Linear Regression Model”, Akron Ohio.
- [4] Duhl LJ, “The healthy city: its function and its future’, *Heal Promot*, 1: 55–60, 1986;
- [5] Global City Indicators Facility members, “Global City Indicators Facilities briefing document”, Spring 2013.
- [6] Guide to sustainable community indicators, Hart M., Andover, MA., *Sustainable Measures*; 2nd ed. 1999.