# Data Science for Understanding and Assessing Spatial Justice

James Roberts, Dr. Debzani Deb and Dr. Russell Smith
*Winston-Salem State University, Winston-Salem, NC*

## Purpose

- *Spatial justice* is a relatively new concept of social justice considered in the context of geographical space.
- According to Rocco [1], "Spatial Justice refers to general access to public goods, basic services, cultural goods, economic opportunity and healthy environments".
- Achieving spatial justice would be a means by which to address the inequitable distribution of goods, services and resources. But how is the achievement of spatial justice to be consistently assessed?
- We argue that there is a data-driven means for developing a spatial justice index (SJI) in order to make an objective and accurate assessment of spatial justice/injustice.
- The spatial units considered in this study are the *census tracts* (geographic units), as defined by the US Census Bureau.
- We hypothesized that the census tracts with greater access (measured as the presence/number of a spatial variable within the tract) will have a lesser spatial injustice.

## Acknowledgement

**NSF RIA Award #1600864**

**Feedback @ http://ibigcloud.altl.org**

## Methodology

1. Create necessary data sets by downloading, merging and cleaning data from various sources [2,3.4].
2. Explore features that are most correlated with spatial justice.
3. Build a predictive model that can meaningfully provide insights about SJI.
4. Compare our model-generated SJI with other geographic goodness measures (e.g., per capita income, median house price) to validate our hypothesis.

- WEKA [5], an open source data mining tool, was utilized for classification and prediction.
- This study particularly focuses on the 93 census tracts of Forsyth county in the US state of North Carolina.
- Set of features (26) include: Number of schools, grocery stores, residential property, industrial property, commercial property, brownfields, free charity, gas stations, libraries, museums, and bus stops; Area enclosed by land and water; Presence/absence of parks, ABC stores, hospitals, food dessert, highway access, airports, water and sewer availability, colleges, community centers, farmers market, fire stations; and travel time to work.
- Data regarding to 26 spatial attributes for each of these 93 census tracts are combined in a CSV file.
- For the classification purpose, a SJI column (class column) has been added which has the value "Yes" if the Per Capita Income of the specific tracts is more than the county median value and possesses the value "No" otherwise.
- **Classification**: Machine learning algorithms such as Naïve Bayes (NB), K-nearest neighbors (KNN), Decision Tree (DT), and Random Forest (RF) from WEKA were utilized to classify.
- Principal Component Analysis (PCP) was utilized to identify most correlated (valuable) attributes.
- In all experiments, 10 fold cross validation was utilized and accuracy of the results are recorded as % of correctly classified census tracts in terms of spatial justice/injustice for unlabeled data.

## Results & Discussions

| | % Accuracy | % Accuracy with PCP |
|---|---|---|
| NB | 64.5% | 78.5% |
| KNN | 64.5% | 66.6% |
| DT | 65.6% | 69.9% |
| RF | 66.6% | 73.1% |

- The above table shows that using all 26 attributes, the best accuracy (66.6%) was achieved in case of RF Algorithm.
- Using attribute selection (PCP), resulted in 78.5% accuracy while using Naïve Bayes for classification.
- Our attribute selection experiments show that the most important attributes are Food Dessert, Bus Stops, Gas Stations and Brownfields.
- We hope that this developed SJI can be applied across the entire country to help communities comprehend, accept and potentially combat spatially unjust geographies within their communities.

## References

1.Rocco de Campos Pereira, "Why Discuss Spatial Justice in Urbanism Studies", Atlantis, 24 (4), 2014.
2.United States Census Bureau, https://www.census.gov
3.MapForsyth, Forsyth County Geographic Information Service, http://mapforsyth.org
4.United States Environmental Protection Agency. https://www.epa.gov
5.Weka 3: Data Mining Software, https://www.cs.waikato.ac.nz/~ml/weka/

WINSTON-SALEM STATE UNIVERSITY