**Project Title**-Spatial Data Analysis for Cellular Signal Strength

**Project description** (describe why you think the project has merit and how it contributes to the field of computing and information \*)

This entire study deals with the Spatial Data Analysis using synthetic data. Spatial Data comprise of geographical information about an entity on Earth. In Spatial data analysis it’s difficult to collect data in many situations. Without the data, training models is not possible and we can’t pull out important trends we wish to study. When this occurs, synthetic data are used for "what if” analysis. A potential scenario is proposed and encoded within a model, and the effect is studied by running the model. Although synthetic data can be very useful, it suffers from the fact we wont know how appropriate the assumed scenario is relative to reality.

This study is the first step in bridging the gap between fitting models to data and "what if" synthetic data practices. The goal is to first specify an assumed scenario for "what if" analysis. The data is generated from that model. Various candidate models are made to fit on the generated data. These candidate models range from simple to complex. I will then identify the best model accounting for complexity. Since I know the exact model which generated the data, I will check if the performance metrics identified the correct model. I will then check if the performance metrics properly identifies the correct model under various sample sizes and various levels of noise.

With the help of this generated framework, we will be able to give recommendations on the required number of samples in order to correctly train models. The work is therefore correcting a framework to help identify how many training points are needed based on assumed spatial features. This spatial data analysis is used to understand the variation of cellular strength in different areas.

Rather than following the usual approach of fitting models on the available data, this study involves generation of synthetic spatial data using a simple and complex process, understand the parameters that generated data, extract the important features, list out candidate models, fit various candidate models from simple to complex and understand if models do better with simple or complex data. Will try to understand how models fit and evaluate which model generalize the best. This involves the evaluation of the effect of sample size and noise on candidate models. This spatial data analysis is used to understand the variation of cellular strength in different areas. With Spatial Analysis we get a better understanding about the characteristics of places and relationship between them for an entity. Spatial analysis gives a new perspective for decision making.

**Learning Objective** (provide a description of what will you learn by doing this project)

From this study I will learn about

1)Spatial Data Analysis

2)Understand the synthetic data generation and learn about the parameters generating data.

3)Understand how candidate models are affected by the simple and complex data

4)Learn about the effect of sample size and noise on the models.

**Project Scope** (determine and document a list of specific project goals, features, functions, tasks)

1)Specify an assumed scenario for “what if” analysis

1)Generating synthetic spatial data using for the model defined above.

2)Learn about the parameters that generated data.

3)Fit the generated data on a list of candidate models ranging from simple to complex.

4)Evaluate which model generalize as best accounting to complexity.

5)Evaluate the effect of sample size and noise on the candidate models.

**Deliverables** (provide a list of what will you deliver / submit at the end of the semester \*)

1) Generate synthetic data for the assumed scenario for “what if” analysis. For this data learn about the parameters that generated data.

2)List candidate models. Fit the candidate models to the above generated data. Evaluate which models generalize the best.

3)Vary sample size and noise to evaluate candidate models.

4) Consider data generation using complex scenario and repeat the above steps 1-3 on this data.

**Timeline and proposed meeting schedule with your faculty advisor**

May-Literature study, understand and try generating the synthetic spatial data for the considered scenario.

June- Generate synthetic simple and complex spatial data and fit candidate models.

July-Understand the effects of sample size and noise on candidate models. Work on report and prepare for deliverable.

August- Submission and deliverable.

Meeting Schedule- Weekly once.