**DATS 6101**

**INTRODUCTION TO DATA SCIENCE**

***Dr. Darcy Morris***

*Project II Proposal on*

**An Analysis of Food Security in USA - 2021**

*By*,

**TEAM II**

*Akshay Verma, Aveline Mariya Shaji and Uugangerel Bold*

**Research Proposal**

* **RESEARCH TOPIC**

In the first project, we analyzed some demographic and socioeconomic factors’ impact on food security level in the USA in 2021.

For the second part of the project, we will build models using different Logit Regression techniques to find out which combination of social groups are the most at risk for food insecurity.

Our variables for now are states, family size, income level, SNAP usage, race/ethnicity, immigrant status, Hours at Work and education level.

* Our **SMART** questions are below:
* **Specific**: To extend the EDA on the Census data to find out what combinations of socio-economic factors lead to food insecurity.
* **Measurable**: Measure the risk of food insecurity at different combinations of selected factors.
* **Achievable**: Make a prediction model using logistic regression for the food insecurity. Extend the modelling using KNN and Random Forest.
* **Relevant**: Food being the basic requirement of any human, this study can shed light on what the authorities and we ourselves can do in order to eradicate food insecurity.
* **Time-Oriented**: Data set for the month of December 2021 is considered for the study and the study is expected to come-up with interesting results by early December 2022.
* **Understanding Data**

We have obtained the data from the US Census website using CPS data. The CSV file contains approximately a hundred thousand observations.

The link to our dataset is:

<https://www2.census.gov/programs-surveys/cps/datasets/2021/supp/dec21pub.csv>

* **GitHub Repository**

We are a group of three and will be working on this project. Our collaborations can be seen at our GitHub Repository.

The link to our repo is : <https://github.com/ubold01/DATS6101-Project1-2022Fall>