

## ASSIGNMENT 1: Feed forward neural network prediction model

**Due: October 17<sup>th</sup>**

1. Question: Develop a prediction model to fit a prediction model for food insecurity.
  - a. Food insecurity is influenced by socioeconomic factors, physical environment, discriminative policies, and health infrastructure.
  - b. We have data at the county-level for these features
  - c. Developing a predictive model can be useful for various reasons, e.g., predicting at a more granular level of local community to inform community-specific decisions.

Data Source: [County Health Rankings](#)

Dataset: Use the [2025 County Health Release National Data](#)

### Instructions:

2. You will be evaluated on the steps you see to build this model
  - a. Data preparation and processing
  - b. Data analysis -understanding the data through data statistics and visualization
  - c. Data handling –missing data, dimensionality reduction, feature selection
  - d. Model setup and prediction – hyperparameter tuning in optimizer, NN architecture selections, validation.s
3. You can use AI chatbots to generate the code-but determining what is needed for each of above should be your own. You can use examples in the open literature on other datasets as reference to identify some common steps, but when your sequence of step should be informed by analyses of outcomes on previous steps
4. What to submit: Your final code with clear step-by-step analyses, including prompts used for AI chatbots, e.g.,
  - a. Reviewed the CSV file to understand the layout of the sheets and the data features.
  - b. My decision was to extract features from sheets 'sheet1' and 'sheet2', that I preselected, and added as Labels in a new sheet.
    - i. Gemini prompt: Create a code to read file from GoogleDrive folder.
    - ii. Gemini prompt: Create code to create dataframe using features from uploaded file. There are two sheets of data '..1' and '..2', the feature names are in row 2. The features I want to extract from '..1' are in listed in sheet 'label1', and the features to extract from '..2' are listed in

‘label2’. Append both to same dataframe, make sure row matching is done by FIPS which is in column one of both data sheets.

- c. Understand the data , how many missing variables, data statistics.
  - i. Gemini prompt:..
- d. Visualize the data
  - i. Gemini prompt:
- e. (I don’t need to know prompts for every minor edits you did to main prompts, but just the main prompts)

### Some sources for data processing

Missing value handling techniques:

- Guide: <https://www.kaggle.com/code/parulpandey/a-guide-to-handling-missing-values-in-python>
- Python packages: <https://scikit-learn.org/stable/api/sklearn.impute.html#module-sklearn.impute>

Feature selection techniques:

- MIE 590 - Focus on the ‘filter’ methods (MIE690 students see below)  
<https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/>
- Reference code on small dimension dataset: [https://github.com/Bakari01/House-Price-Prediction-using-DNN/blob/main/house\\_price\\_prediction\\_DNN.ipynb](https://github.com/Bakari01/House-Price-Prediction-using-DNN/blob/main/house_price_prediction_DNN.ipynb)

### MIE 690D component:

1. Read the following articles and incorporate a feature selection technique  
Cherepanova, Valeriia et. al., A Performance-Driven Benchmark for Feature Selection in Tabular Deep Learning, NeurIPS 2023.  
<https://openreview.net/pdf?id=v4PMCdSaAT>
2. Read: Ahmed, K.R., Ansari, M.E., Ahsan, M.N. et al. Deep learning framework for interpretable supply chain forecasting using SOM ANN and SHAP. *Sci Rep* **15**, 26355 (2025). <https://doi.org/10.1038/s41598-025-11510-z>