Project Title: Data-analysis tool to ingest high-velocity and high-fidelity system metrics and sensor data from the field-deployed Waggle Sensors

Abstract: The main goal of this project is to design and build a machine learning model and to train the model to make hardware improvements on the temperature reading capabilities of field-deployed Waggle sensors. While the sensors measure temperature, humidity, and pressure from the environment, analysis from the sensors’ data indicated that the sensors’ values deviate from the actual values of the environment, provided by an Argonne meteorological tower in the area, as they are corrupted by the heat that is given off from the system inside the sensor. Based on this fact, the goal of this project is to redesign the sensors’ systems using a machine learning model to accurately predict the correct temperature. Before the data is used, it is processed and cleaned using the Python library pandas. An array of sensor values (~7 parameters) is collected from a node every 15 minutes. The model is a linear regression model that uses this array of parameters as features, or characteristics used for analysis, to predict the correct temperature. The project is part of a larger initiative called Sage, which is a project funded by the National Science Foundation to design and build a new kind of national-scale reusable cyberinfrastructure to enable AI at the edge.