Milking The Data

UPenn AI Bootcamp Major Project 3

Team Members

Larry Azar
Thomas Wells
Jason Clibanoff
Patrick Nwankwo

Overview

Project Goal: Predicting Milk Yield in Mountain-Pastured Cows using Artificial Neural Networks

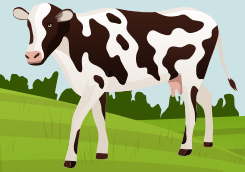
Objective: Design and Build an ANN regression model that accurately predicts the average quantity of milk produced by cows based on environmental conditions and lactation cycle information.

Dataset:

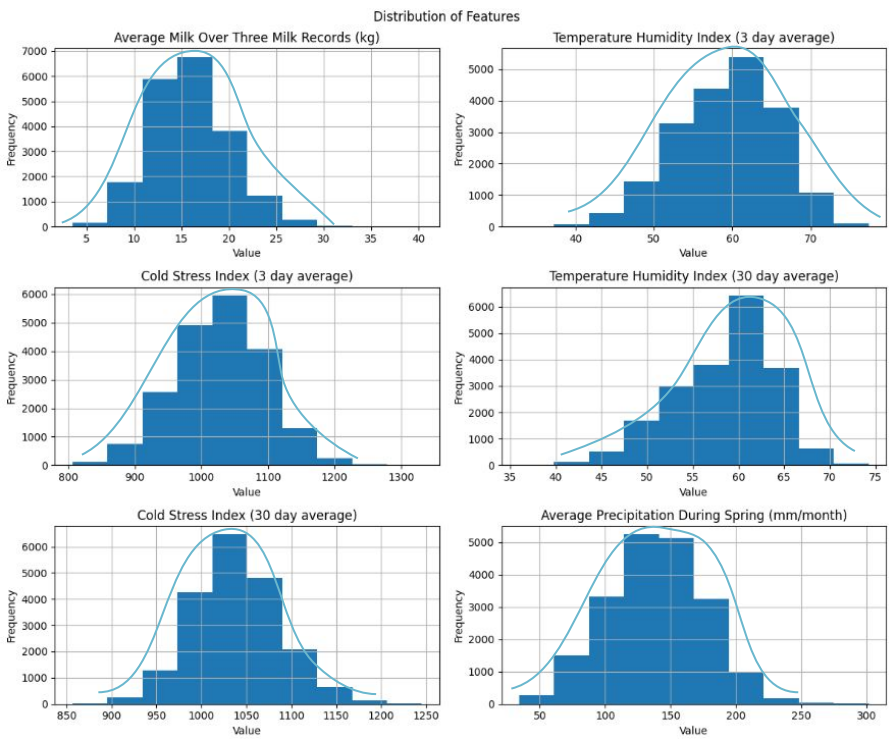
- Target column: Average Milk Production (kgs)
- This dataset has been extracted from the milk record database of the Braunvieh-CH breeding organisation.
- https://zenodo.org/records/3962046

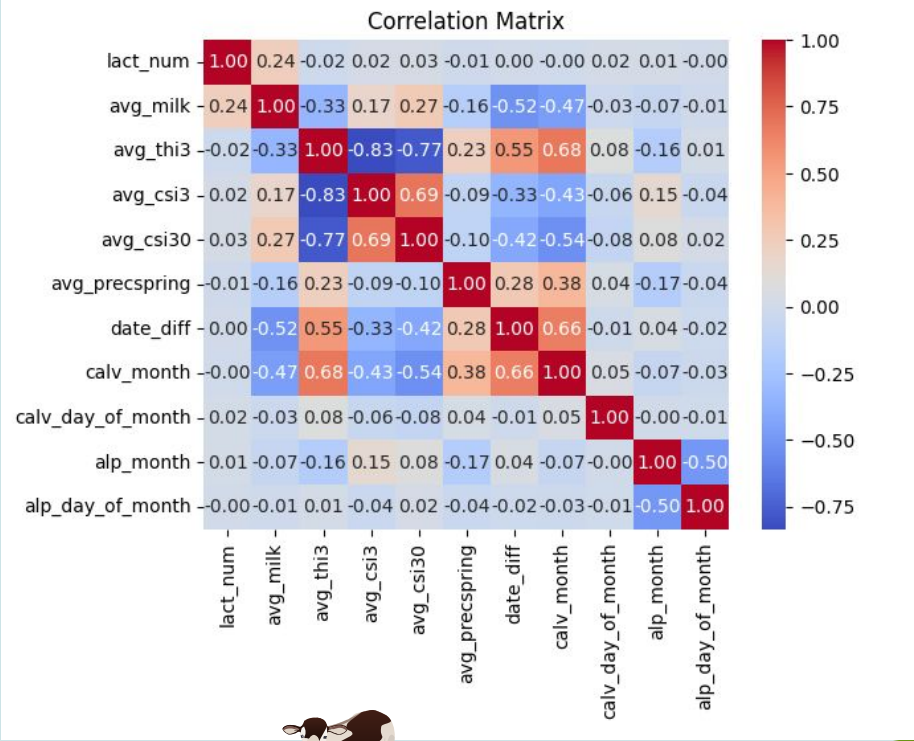
Key variables & Features:

- Temperature Humidity Index (THI)
- Cold Stress Index (CSI)
- Precipitation during Spring
- Lactation number (num births)
- Date of calving
- Date of first record in the alp



Exploring Our Data





Feature Engineering

- Converting date strings to datetime
- Calculating the time between calving and arriving in the alp
- Converting datetime columns to numeric values for the month and day of the month
- Dropping original datetime columns once data this data has been extracted

Other Preprocessing

- Dropped the 41/20,000 rows with missing data
- Removed ID and avg_thi30 to address data leakage
- After train test split, we scale all of our data with standard scaler

Creating Our ANN Regression Model

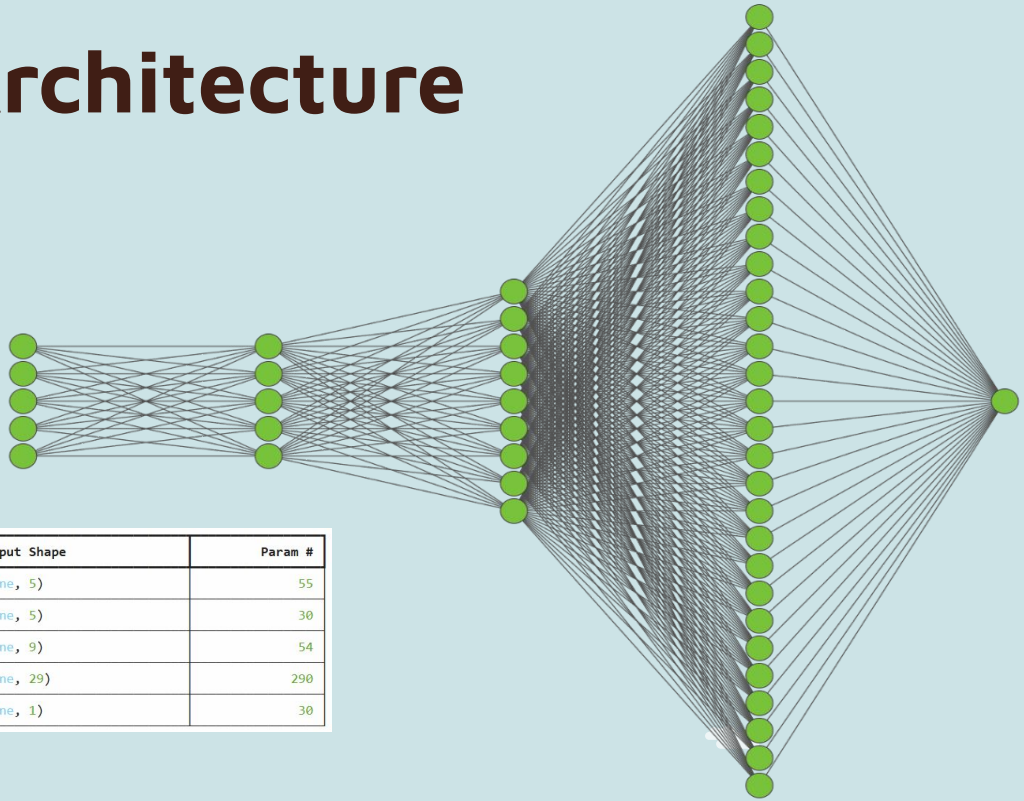
Initial ANN Model Creation

- **Input layer**: 10 neurons (one for each feature).
- **Two hidden layers**: 32 neurons each, ReLU activation.
- **Output layer**: 1 neuron for predicting milk production

Hyperparameter Tuning & Optimization

- Used **KerasTuner** to find the best:
  - **Activation function**
  - **Number of layers**
  - **Number of neurons** per layer
  - **Learning rate** & **optimizer**
  - **Hyperparameter Tuning**

Model Architecture



Layer (type)	Output Shape	Param #
dense (Dense)	(None, 5)	55
dense_1 (Dense)	(None, 5)	30
dense_2 (Dense)	(None, 9)	54
dense_3 (Dense)	(None, 29)	290
dense_4 (Dense)	(None, 1)	30

Results

The Mean Absolute Error of our tuned model is **2.439**

This means our predictions are roughly 2.439 kg of from the actual values. Given our range of average milk values (roughly 3 to 40), this error is relatively small but still leaves some room for improvement.

Lets Try Our Gradio App!

Thanks!

Any questions?