AGRICULTURAL DATA COLLECTION, ANALYSIS AND FORECASTING

Submitted by:

Shekhar Thapa

Santosh Koirala

Keith Schneider

05/10/2022

INFO8000 Project 2022

University of Georgia, College of Engineering

597 D. W. Brooks Drive, Athens, GA 30602

Abstract

Machine learning is an incredibly valuable tool for data analysis, and as a group where our research is focused on agricultural engineering, it seemed like a good opportunity to use some agricultural data and tie in methods learned in this class to analyze trends within a sector of the agricultural industry. The final product we have is a web server (https://getweatherinfo.ga) running a web API that accepts weather data from a sensor connected to an Arduino Uno and publishes the data on the server in real time. Then, we chose an existing API that had good documentation and relatively complete data and performed data visualizations to explore the raw dataset and to determine a proper method of machine learning. The method of machine learning we decided to use was a Seasonal Autoregressive Integrated Moving Average with Exogenous regressors (SARIMAX); this was selected based on the seasonality displayed in the decomposition plot of the price data.

Our first goal was designing a web server running a Web API that could store sensor data (temperature and humidity of various locations) and publish them instantaneously so that it could be accessible from any location. In precision agriculture, various sensor data like temperature, humidity, moisture content, pH value of soil, sunlight intensity are used to monitor real-time field conditions so that the on-field management systems can act accordingly for better production of crops. So, this project tries to build an agricultural-sensor data storage and management system via web API.

Firstly, a database (sqlitesensordata.db) was created using SQLiteDatabaseBrowser64, which could store location information, temperature & humidity, and timestamp. Then, using FastAPI, a web API code (Task1.py) was written and hosted through a web server with the domain name https://getweatherinfo.ga. Several "GET" endpoints were included on the web API to access and filter the required data from the API's database. For persistent data storage, an Arduino Uno microcontroller was used to read analog temperature and humidity sensor values and communicate with a local computer through serial ports. A python program with serial library import was run in the local computer to get the sensor data through a serial port that was then called to store data into the server's database. Finally, a detailed documentation page for the web API was written in html format and published through the server's domain (i.e., https://getweatherinfo.ga)
Due to the unavailability of temperature and humidity sensors for actual measurement, the sensor readings were randomly generated in the microcontroller. The web server with web API was able to store the sensor data continuously and made the real-time sensor data available simultaneously.

Further, we used data visualization to first observe the trend of cost of rice in the EU member states over a period of two decades and then to create a heatmap for the quantity of rice produced in those countries in 2020. The data for the cost of rice in those EU countries was requested using the instructions from a web API documentation and the quantity of rice produced in those countries in 2020 was obtained manually from a webpage. Pandas functions were used to process the raw data and convert the data types. Furthermore, regex was used for data filtering and matplotlib and geopandas functions were used to visualize graphs and maps. Our line graph showed that the cost of rice in recent years has been higher as compared to the early 2000s.

Moreover, the cost of rice has been consistently high in Italy as compared to other EU countries for most years, whereas the cost of rice has been low in Portugal in those years.

Comparison of the quantity of rice produced in these countries in 2020 resulted in a warmer color map for Italy indicating that rice production was highest in Italy. In contrast, Romania had the lowest rice production in 2020 as shown by a colder color map. The high cost of rice in Italy despite the highest production could not be explained by our study. Further study could be done to analyze the aspect of demand and supply of rice and its effect on the cost of rice in Italy.

In case of machine learning, the goal was to ultimately forecast the future price trends for two years following the conclusion of the available data. A SARIMAX model was utilized due to the seasonality trend noticed in the time-series decomposition plot. The SARIMAX model was promising with an order of 1,0,1 and seasonal order of 0,1,1,12 showing the best Akaike Information Criteria (AIC). The model statistics were calculated with the important statistic of the null hypothesis all having a value of 0 indicating that the variables are all statistically significant. The forecasted model looks promising in being able to accurately predict future price trends. With the forecasted model looking as promising as it does, similar models should be able to be used for more important information within the agricultural industry. Specifically, we hope to be able to use sensor data, such as we collected from the Arduino Uno to be able to forecast future trends in crops specifically in the yield for a given crop within a given season and weather conditions.

Ultimately, this project may not have the most practical applications in terms of the specific data we analyzed. However, the methods used can be applied in the future to several promising ends. One possible way the methods implemented could be used is regionally a farmer could implement the microcontroller with a temperature and humidity sensor in their fields. The data collected could then be used as input into a program that uses machine learning to assist in the forecasting for crop yield. A more complex algorithm that takes into account growing conditions with crop yield would have to be used but unfortunately that type of data is not available from a basic API. This study further emphasizes the importance of data visualization during data exploration and data analysis. Furthermore, sound data visualization is critical in elucidating the outcome of the study to audiences from diverse backgrounds.