Correlational Analysis to Predict Impact of Weather and Population on Crop Production

Eswaran Badrinarayanan Venkateswaran
Janani Suresh Kumar
Zannatul Ferdoush

COMP 8700: Introduction to Artificial Intelligence

Project Report
Submitted to: Dr. Pooya Moradian Zadeh
Term: Winter 2022
School of Computer Science
University of Windsor

Correlational Analysis of Weather and Population on Crop Production

Eswaran Badrinarayanan Venkateswaran, Janani Suresh Kumar, Zannatul Ferdoush

Graduate Student, School of Computer Science, University of Windsor, Windsor ON N9B 2E7

badrina@uwindsor.ca, sureshkj@uwindsor.ca, ferdous@uwindsor.ca

***Abstract* - Agriculture is heavily reliant on the weather [1]. Rising global temperatures as a result of increased carbon dioxide (CO_2) emissions may have a significant impact on agricultural production in some areas. However, in order to understand these phenomena, elements such as nutritional levels, soil moisture, and water availability must be examined as well [2]. This project intends to develop a system that analyses historical data on crop production, weather, and population in Ontario, Canada, in order to do predictive analysis and identify factors that influence agricultural production rates. The system employs regression analysis to determine which variables have the most impact on agricultural productivity and to forecast future yields. The data obtained are from reputable sources such as Statistics Canada [3], Agriculture Canada [4], and Data Ontario [5]. The research will look at statistics for specific food industries, such as fruits and vegetable production, over 20 years (2002 - 2021) to see how these crops have grown in Canada based on the above-mentioned criteria. The findings of the study might be utilized to create a recommendation system that explains how various elements like weather and population can boost or reduce agricultural productivity, as well as alternative crops to plant based on the observed patterns.**

***Keywords:* Correlational analysis, Correlation matrix, Regression models, Crop Production, Pearson Correlation Coefficient.**

I. INTRODUCTION

A. GENERAL OVERVIEW:

Agriculture is vital to the economic prosperity of every country. Meeting the present population's food demands has become a tough issue due to population expansion, frequent changes in climatic conditions, and limited resources. Precision agriculture, often known as smart farming, has risen to prominence as a cutting-edge solution for tackling current agricultural sustainability challenges. The main force behind this cutting-edge technology is machine learning, which allows machines to learn without needing to be explicitly taught. By 2050, the world's population will have grown by about 34%, reaching 9.1 billion people.

Food consumption will increase by 70% in the next several years, whereas area suitable for agriculture will decrease due to rising urbanization [6]. By 2050, Ontario will be the world's most populous country, yet it is already lagging behind in terms of domestic food production. Lower food yield is caused by a lack of preparation, unexpected weather circumstances, poor harvesting and irrigation processes, and animal mismanagement. Thus, it becomes significantly important to leverage between production and export.

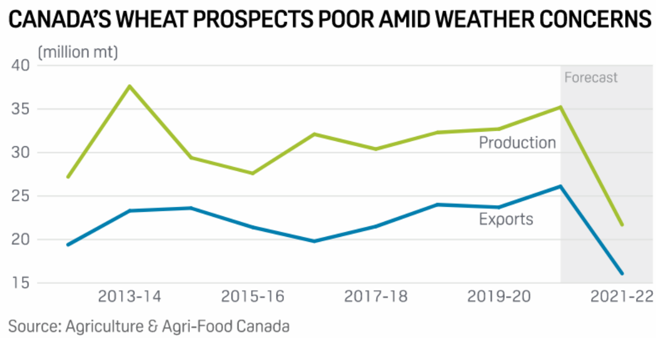

Figure 1. Canada's Wheat Production over the years [7]

B. MOTIVATION:

As a result of global warming, nature has experienced a drastic change in weather conditions in recent years. The Earth's average temperature has increased, putting climatic conditions at jeopardy. The most tough obstacles for disadvantaged farmers are droughts and heavy rainfall. Poor weather conditions reduce farmer revenue by 20-25 percent, according to a yearly economic assessment done by the Canadian government. With rising temperatures, longer growing seasons, changing precipitation patterns, and an increase in the frequency and intensity of extreme events, climate change will present both challenges and opportunities for Canada's agricultural business. The consequences of climate change will not be uniform across Canada or constant across seasons.

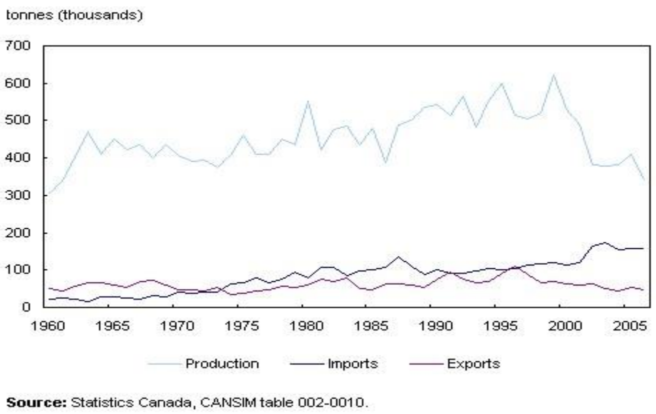

Figure 2: Crop Import & Export between 1960 - 2005

In terms of production, certain regions may have the opportunity to plant warmer-weather crops and benefit from a longer growing season with fewer cold weather occurrences that might destroy crops.

Water stress (flooding or drought), heat stress, wind damage, increased pest and disease loads, and the influence of these many stressors on soil health will all hamper production, lowering productivity, profitability, and competitiveness for farmers across the country. As a result, farmers must be aware of the dynamic elements that might have a significant impact on productivity [8]. Henceforth, the connection of influential elements such as weather and population may assist farmers in predicting the influence on their crop development, estimating harvest time, and determining the best crop sowing for a certain land topography.

C. IMPORTANCE OF THE PROBLEM:

Using traditional agricultural practices, farmers face obstacles every day. It is difficult to foresee climate change as a result of increased global warming, which has an impact on agricultural productivity. Machine learning regression algorithms are being used to evaluate and understand agricultural growth patterns in Canada during the previous 20 years.

The forecast will aid farmers in monitoring the health of their land and crops in order to produce more productive harvests. Customers as well as businesses are part of the farmers' target market. Farmers markets and companies help take out the middlemen, which is one of its benefits. They do not have to worry about transporting food from the farm to the shop, which reduces resource use.

Correlational analysis can be used to determine prevalence and relationships among variables, and to forecast events from current data and knowledge. Correlational analysis can be applied in agricultural research where researchers can effectively establish dependencies and identify potential mismatches between supply and demand. The goal of our project is to evaluate importance of key metrics in each of considered correlational variables that can affect crop production, thereby providing valuable insights for future analytical study.

D. CONTRIBUTION:

The proposed system involves a predictive analysis of environmental factors and population affecting Canada's food production industry. We plan to focus on a particular category under the food industry, e.g., fruits, vegetables, and observe the data across 20 years, e.g., 2002 -2021, to understand the trends affecting the growth pattern. Relevant categorical datasets from Statistics Canada, Agriculture Canada, and Data Ontario will be gathered individually and merged based on common field aspects.

The project will employ regression algorithms to predict continuous data values of the crop production over the next few decades based on the training dataset. Linear regression, Support Vector Regressor and Decision Tree Regressor are used to predict the value of crop production. We intend to apply the Correlation matrix and scatter plots to visualize data better and comprehend behavioral trends. The models have been trained and tested to improve data performance efficiency and accuracy.

Although cutting-edge technologies and modern developments in precision agriculture show a positive trend there has always been a degree of uncertainty in understanding behavioral patterns of crop produce.

Our project involves a correlational analysis of environmental factors and the trend in which affect crop production (food production industry) in Canada.

We focus on particular category under food industry and observe the data across a period span of 20 years eg; 2002-2021 to understand the pattern growth of the selected crops - fruits and vegetables

II. LITERATURE REVIEW

Puyu Fenga,b , Bin Wangb , De Li Liub,c , Cathy Watersd , Qiang [9] have created a hybrid model by combining the APSIM model outputs with growth stage-specific ECEs indicators (such as frost, drought, and heat stress) in the Random Forest (RF) model, using the multiple linear regression (MLR) model as a benchmark. The APSIM + RF hybrid model explained 81% of observed yield fluctuations in the New South Wales wheat belt in south-eastern Australia, with a 33% improvement in modelling accuracy over the APSIM model alone and a 19% improvement over the APSIM + MLR hybrid model.

Rehman, Abdul, Deyuan, Zhang; Hussain, Imran; Iqbal, Muhammad Shahid [10] used econometric analysis to look into the relationship between agricultural gross domestic product (AGDP) and variables like apple, citrus, pears, grape, and banana in Pakistan from 1980 to 2015. They used time series data from secondary sources such as the Pakistan Bureau of Statistics, Statistical Year Books, and the Economic Survey of Pakistan. The Ordinary Least Square (OLS) technique and the Augmented Dickey Fuller (ADF) test were used to analyse the data, and the Johansen co-integration test was used to interpret the results. The method of machine learning was applied to investigate and forecast future agricultural productivity in Pakistan. The authors discovered that banana, citrus, and pear output had a positive and considerable impact on AGDP, but apples and grapes had a negative but minor impact. Andrew Crane-Droesch [11] implemented a yield modelling method that employs a semiparametric form of a deep neural network to account for complicated nonlinear interactions in high-dimensional datasets, as well as known parametric structure and undiscovered cross-sectional variability. They show that our strategy outperforms both traditional statistical approaches and fully-nonparametric neural networks in forecasting yields of years withheld during model training using data on corn yield from the US Midwest. They found out that there are strong negative consequences of climate change on maize production using scenarios from a suite of climate models, although they are less severe than impacts anticipated using traditional statistical approaches. In particular, in the hottest locations and scenarios, our approach is less gloomy.
Gerald C. Nelson, Mark W. Rosegrant, Jawoo Koo, Richard D. Robertson, Timothy Sulser [12] used the SPEI drought index to assess the size, intensity, and duration of predicted drought. A proposed non-linear ensemble of Random Forest (RF) and Gradient Boosting Machines was used to quantify the influence of future drought on agricultural output using the ISI-MP (Inter-Sectoral Impact Model Inter-comparison Project) crop model (GBM). According to the findings, high drought magnitude is predicted in some parts of South Asia over a longer period of time, although high drought intensity is predicted for a shorter period of time.

It was also discovered that droughts in Afghanistan, Pakistan, and India will last longer in the future. When compared to the stand-alone approaches of RF and GBM for yield loss risk projection, our suggested ensemble machine learning (EML) methodology exhibited a good prediction skill with a minimum value of RMSE (0.358-0.390), MAE (0.222-0.299), and a maximum value of R2 (0.705-0.918). The impact of drought on agricultural output reveals a substantial probability of yield loss during extreme drought events, with yield losses of 54.15%, 29.30%, and 50.66% in the future for rice, wheat, and maize crops, respectively.

Monteith, J. L analyzed the climate and efficiency of crop production in Britain by computing the efficiency of crop production as a ratio of energy output (carbohydrate) to energy input (solar radiation) in thermodynamic terms. He also considered two major factors of climatic constraints – temperature and water supply for which that impacted the efficiency score.[13]

On the other hand, Evangelista, J. C., Escalona, J. A. S., & Pigao, K in their research work conducted correlational analysis between industrial and agricultural sector towards economic development. In this paper, the authors employed Pearson Coefficient and multiple regression models to establish and identify relationships between Manufacturing-Agriculture & Construction-Agriculture Industry.[14]

III. PROJECT DETAILS AND METHODOLOGY

A. DEFINITIONS:

Correlational Analysis: Correlation or dependence in statistics refers to any statistical relationship, whether causal or not, between two random variables or bivariate data. Although the term "correlation" can refer to any type of association, in statistics it usually refers to the degree to which two variables are linearly related. The correlation between the height of parents and their offspring, and the correlation between the price of a good and the quantity of goods that consumers are willing to purchase, as depicted in the so-called demand curve, are two well-known examples of dependent phenomena [15].

Correlations are useful because they can show a predictive relationship that can be used in practice. Based on the correlation between electricity demand and weather, an electrical utility, for example, may produce less power on a mild day. There is a causal relationship in this example because extreme weather causes people to use more electricity for heating or cooling. In general, the presence of a correlation does not imply the presence of a causal relationship.

Correlation Coefficient: The values range from -1.0 to 1.0. A calculated number greater than 1.0 or less than -1.0 indicates that the correlation measurement was incorrect. A correlation of -1.0 indicates that there is a perfect negative correlation, whereas a correlation of 1.0 indicates that there is a perfect positive correlation [16].

A correlation of 0.0 indicates that there is no linear relationship between the two variables' **movements.**

The correlation coefficient is a statistical measure of the strength of the relationship between two variables' relative movements. Correlation statistics have applications in finance and investing.

A correlation coefficient, for example, could be calculated to determine the level of correlation between crude oil prices and the stock price of an oil-producing company, such as Exxon Mobil Corporation. Because oil companies make more money as oil prices rise, the correlation between the two variables is strong.

Size of Correlation	**Interpretation**
.90 - 1.00	Very high positive correlation
.70 - .90	High positive correlation
.50 - .70	Moderate positive correlation
.30 - .50	Low positive correlation
.00 - .30	Negligible correlation

Table 1. Interpretation of Correlational Values

Pearson Correlation Coefficient: Pearson correlation coefficient, often known as Pearson's r, the Pearson product-moment correlation coefficient (PPMCC), the bivariate correlation, or simply the correlation coefficient, is a measure of linear correlation between two sets of data [17]. It is the ratio between the covariance of two variables and the product of their standard deviations; consequently, it is basically a normalized measurement of covariance, with the result always falling between 1 and -1. As with covariance, the measure can only indicate a linear correlation of variables and ignores many other forms of link or association.

When applied to a population, Pearson's correlation coefficient is often symbolized by the Greek letter (rho) and is also known as the population correlation coefficient or the population Pearson correlation coefficient. The formula for is: given a pair of random variables (X,Y) [18].

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

where:

- cov is the covariance
- σ_X is the standard deviation of X
- σ_Y is the standard deviation of Y

Figure 3. Calculation of Pearson Coefficient

The following regression models are used in the project for calculating mean square error and covariance score. Multiple regression models are used to verify model's performance and correctness (measure to extent of difference between actual and analyzed outcome).

Linear Regression: A linear technique to modelling the connection between a scalar response and one or more explanatory factors is known as linear regression (also known as dependent and independent variables). Simple linear regression is used when there is only one explanatory variable; multiple linear regression is used when there are more than one [19]. Relationships are modelled using linear predictor functions whose unknown model parameters are derived from data in linear regression. These kinds of models are known as linear models.

Decision Tree Regressor: Decision tree learning or induction of decision trees is one of the predictive modelling approaches used in statistics, data mining and machine learning [20]. It uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves).

KNN regressor: KNN regression is a non-parametric approach that approximates the relationship between independent variables and continuous outcomes by averaging data in the same neighborhood [21]. The KNN algorithm predicts the values of new data points based on 'feature similarity.' This means that a value is assigned to the new point depending on how closely it resembles the points in the training set.

Support Vector Regression (SVR): It is a regression function generalized by Support Vector Machines; a machine learning model used for continuous data categorization. Support Vector Regression is a form of Support Vector Machine that allows for both linear and nonlinear regression [22]. In our project, this is useful in analyzing weather data as it is dynamically changing from time and time and in categorizing importance of a particular feature with crop production (feature importance).

Null Values and Outliers: The process of removing outliers involves excluding non-pertinent or irrelevant fields present in the dataset so as to improve model performance during training and testing.

The values of fields that are null are identified and replaced with mean values.

B. SPECIFICATIONS:

In our project, the correlational analysis is performed in python programming language and jupyter notebook is employed as it provides a user-friendly interface for performing data visualization. The dataset considered for the project is taken from Statistics Canada [24] and five different types of fruits and vegetables production data between the period 2002-2021 are chosen for correlational analysis. Five fruits chosen are apple, strawberry, raspberry, apricot and Blueberry and five vegetables chosen are asparagus, carrot, onion, tomato and capsicum.

C. ARCHITECTURE:

Our proposed work can be categorized into three phases.
In the first phase - Collects different crop production, weather and population data of Ontario between 2002 - 2019 – Preprocessing (removing outliers, replacing null values with mean and checking skewness) and cleaning of data is done. In the second phase - Performs correlational analysis to better understand the statistical relationships between the input variables and the output variable - This will help to identify which variables may or may not be relevant as input for developing a model.

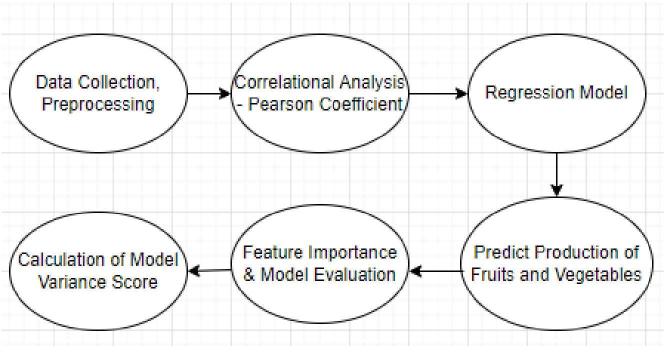

Figure 4. System Architecture

After the data is cleansed and preprocessed [23], the correlational analysis is calculated using Pearson Coefficient and training data (70%) & testing data (30%) are assigned. The input to the regression model is weather population data from the dataset that is fed into to predict production of fruits and vegetables.

Then, the model is evaluated with variance score calculation and finally the feature importance of weather and population are separately identified and its correlation with crop production is analyzed.

IV. EXPERIMENTAL SETUP

The following sections discuss the implementation, testing and results of our work.

A. IMPLEMENTATION AND TESTING:

In this part, we use a real-world dataset to assess the performance of our approach. From 2002 through 2019, our dataset includes 20 years of production data for two important agricultural products in Canada, Apple and Asparagus, as well as weather and demographic data. This information come from reputable sources including Statistics Canada [3], Agriculture Canada [4], and Data Ontario [5]. Figure 5 displays some samples from our dataset.

All experiments were conducted on a machine (64-bit Windows 10) with an up to 4.00 GHz Intel Core i7 CPU and 16 GB RAM. Python programming language was used to develop the experimental model on the IntelliJ IDEA framework. Then, for each crop, we ran a correlational analysis to see what statistical relations existed between the input variables (weather and population data) and the output variable (production amount).

We evaluated the data for empty and null values, as well as missing rows, as part of the data preprocessing. We found that the dataset was clean of such issues, and we then saved it in our local computer storage for future experiments.

Year	Apple Production (tonnes)	Asparagus Production (tonnes)	Annual Snowfall (cm)	Annual Rainfall (mm)	Mean Temperature (*C)	Population
2002	18.8	2,032	115.4	684.6	7.5	11,683,290
2003	24.1	2,234	120.2	632.3	6	11,897,534
2004	26.4	2,457	118.1	656.7	6.5	12,094,174
2005	25.7	3,098	110.3	662.4	8.4	12,245,039.0
2006	23.2	3,493	102.7	629.4	13.1	12,391,421
2007	32.6	3,629	109.5	647.8	17.9	12,528,663.0

Figure 5. Sample of the dataset

We identify these relations using correlation coefficient which is a statistical measure that quantifies the association between two variables. For our experiments, we used Pearson's correlation coefficient calculation method which is a popular test to determine the strength of a linear relationship between two normally distributed variables. For varied weather and population data setups, Table 2 displays the Pearson's correlation coefficient values for both Apple and Asparagus. Finally, as part of the predictive modelling technique, we did regression analysis utilizing four popular regression models (Linear Regression, Decision Tree Regression, *K*-Nearest Neighbors Regression, and Support Vector Regression) to evaluate the relationship between our target variable (apple and asparagus production quantity) and independent factors (weather and population) by predicting crop production data. These algorithms have also been used to calculate feature importance. To conduct our experiments, we used 70% of the data for training and 30% for testing.

B. DICSUSSION OF FINDINGS:

Table 2 represents the Pearson's correlation coefficient values for both Apple and Asparagus for varied weather and population data setups. Correlation analysis can reveal meaningful relationships between different metrics or groups of metrics. Information about those connections can provide new insights and reveal interdependencies. From the values represented in Table 2, it is evident that population and annual rainfall have high positive correlations with both Apple and Asparagus production. It signifies that if the people and annual rainfall increase, the output variable (Apple or Asparagus production) will increase, and if one variable decreases, the other decreases equivalently. Other factors such as annual snowfall and mean temperature showed low positive correlation signifying weak dependencies between the target and mentioned variables. We utilized the variance score to assess the effectiveness of our regression models, which is a useful tool for determining how far data deviates from their mean value during regression analysis.

The variance score of the regression models we used is shown in Table 3. The Decision Tree Regression model gets the highest variance score, indicating that it performs the best in terms of crop production forecasting.

The values for feature importance of different input variables are listed in Table 4, with population and annual rainfall being the most important elements.

This supports our findings from the correlational analysis of our data indicating population and annual rainfall being the most effective parameters in predicting future crop yields.

Pearson correlational coefficient	**Apple Production**	**Asparagus Production**
Population	0.875	0.855
Annual Rainfall	0.923	0.830
Annual Snowfall	0.320	0.255
Mean temperature	0.413	0.369

Table 2. Pearson Correlational Coefficient Values

Variance Score	**Apple Production**	**Asparagus Production**
Linear Regression	0.937	0.942
Decision Tree Regression	0.965	0.966
K-Nearest Neighbors Regression	0.919	0.915
Support Vector Regression	0.954	0.952

Table 3. Variance Score of different Regression Models

Feature Importance	**Apple Production**	**Asparagus Production**
Population	0.914	0.902
Annual Rainfall	0.923	0.875
Annual Snowfall	0.322	0.215
Mean Temperature	0.109	0.150

Table 4. Feature Importance of different Input Variables

V. CONCLUSION

- In this paper, we have put forth a solution for the problem that concerns with the understanding statistical analysis, behavior patterns of crop production by taking into two major associated factors such as weather and population for evaluation. One of the main contributions of our work is to solve this task efficiently, by building multiple regression models to compute variance score and make the data more reliable.
- In order to increase system throughput and efficiency of the system, iterative implementation of regression models is carried out to check for precision and correctness of the analyzed outcome.

• Additionally, the study combines extensive crop growth modelling under climate change with insights from an incredibly thorough global agricultural model, all while simulating future climate using two climate scenarios.

VI. FUTURE WORK

• Population analysis: With the aid of a population data set, we may examine the food intake of people of various ages.
• Soil and crop health monitoring: We can use AI/ML to track crop health and detect pests and nutrient deficiencies in the soil. Because the kind of soil and its nutrients have an impact on the type of crop grown and its quality. Because of rising deforestation, soil quality is diminishing, and it is difficult to determine the state of the soil.
• Disease diagnosis: Plant diseases may be easily classified using historical crop growth data, which aids farmers in disease management with the proper technique. Additionally, correlation between multiple variables on multiple metrics (many-to-many relationships) can be studied effectively and meaningful relationships can be derived.

VII. REFERENCES

[1] Carlson, S., & Stockwell, R. (2013). Research priorities for advancing adoption of cover crops in agriculture-intensive regions. Journal of Agriculture, Food Systems, and Community Development, 3(4), 125-129.

[2] Végh, K. R. (1991). Effect of soil water and nutrient supply on root characteristics and nutrient uptake of plants. In Developments in Agricultural and Managed Forest Ecology (Vol. 24, pp. 143-148). Elsevier.

[3] The Daily Fruit and vegetable production, 2019. (2020, February 11). Statistics Canada. https://www150.statcan.gc.ca/n1/dailyquotidien/200211/dq200211a-eng.htm

[4] Agriculture and Agri-Food Canada, agriculture.canada.ca. (2020, September 23). Agriculture Canada. https://agriculture.canada.ca/en

[5] Ontario Data Catalogue. (2021, August 31). Data Ontario. https://data.ontario.ca/

[6] “Measuring canadian export diversification.” https://www150.statcan.gc.ca/n1/pub/13-605-x/2017001/article/54890-eng.htm.

[7] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, “Machine learning applications for precisionagriculture: A comprehensive review,” IEEE Access, vol. 9, pp. 4843–4873, 2020.

[8] “Commodities 2022: Domestic supply tightness may boost canada’s production, import, export rate wheatprices.”https://www.spglobal.com/ketinsights/latest-news/agriculture/commodities-2022-domestic-supply-tightness-may-boost-canadas-wheat-prices.

[9] “Climate change impacts on agriculture.” https://agriculture.canada.ca/en/agricultureandenvironment/climate-change-and-air-quality/climate-scenarios-agriculture. (accessed 18 Febru-ary, 2022).

[10] Feng, P., Wang, B., Li Liu, D., Waters, C., & Yu, Q. (2019). Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. Agricultural and Forest Meteorology, 275, 100-113.

[11] Rehman, A., Deyuan, Z., Hussain, I., Iqbal, M. S., Yang, Y., & Jingdong, L. (2018). Pre-diction of major agricultural fruits production in Pakistan by using an econometric analysis and machine learning technique. International Journal of Fruit Science, 18(4), 445-461.

[12] Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environmental Research Letters, 13(11), 114003.

[13] Prodhan, F. A., Zhang, J., Sharma, T. P. P., Nanzad, L., Zhang, D., Seka, A. M., & Mohana,H. P. (2022). Projection of future drought and its impact on simulated crop yield over South4 Asia using ensemble machine learning approach. Science of The Total Environment, 807,151029.

[14] Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 281(980), 277-294.

[15] Evangelista, J. C., Escalona, J. A. S., & Pigao, K. (2022). The Correlational Analysis between the Industrial Sector and Agriculture Sector towards Economic Development. Journal of Economics, Finance and Accounting Studies, 4(2), 44-54.

[16] Wright, S. (1921). "Correlation and causation". Journal of Agricultural Research. 20 (7): 557–585

[17] Prematunga, R. K. (2012). Correlational analysis. Australian Critical Care, 25(3), 195-199.

[18] Asuero, A. G., Sayago, A., & Gonzalez, A. G. (2006). The correlation coefficient: An overview. Critical reviews in analytical chemistry, 36(1), 41-59.

[19] Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In Noise reduction in speech processing (pp. 1-4). Springer, Berlin, Heidelberg.

[20] Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. Wiley Interdisciplinary Reviews: Computational Statistics, 4(3), 275-294.

[21] Luo, H., Cheng, F., Yu, H., & Yi, Y. (2021). SDTR: Soft Decision Tree Regressor for Tabular Data. IEEE Access, 9, 55999-56011.

[22] Kudraszow, N. L., & Vieu, P. (2013). Uniform consistency of kNN regressors for functional variables. Statistics & Probability Letters, 83(8), 1863-1870.

[23] Awad, M., & Khanna, R. (2015). Support vector regression. In Efficient learning machines (pp. 67-80). Apress, Berkeley, CA.

[24] Brys, G., Hubert, M., & Struyf, A. (2004). A robust measure of skewness. Journal of Computational and Graphical Statistics, 13(4), 996-1017.

[25] Brisbane, J., & Mohl, C. (2014). The Potential Use of Remote Sensing to Produce Field Crop, Statistics at Statistics Canada. In Proceedings of Statistics Canada Symposium.