Time Series Prediction for Food sustainability

Amrit Mohapatra-222IT004
Information Technology
National Institute of Technology,
Karnataka,Surathkal, India 575025
amrtimohapatra.222it004@nitk.edu.in

Natasha Jain-222IT023
Information Technology
National Institute of Technology,
Karnataka,Surathkal, India 575025
natashajain.222it023@nitk.edu.in

Dr. Biju R Mohan
Information Technology
National Institute of Technology,
Karnataka,Surathkal, India 575025
biju@nitk.edu.in

***Abstract*— Maintaining the balance between the demand for natural resources and the need to produce enough food to feed everyone is extremely difficult due to the world population's rapid growth. To improve people's livelihoods, health, and ecosystems for present and future generations, it is imperative to conserve natural resources while ensuring adequate food production. The United Nations' paradigm for sustainable development emphasizes the need to balance environmental protection, social progress, and economic growth. It includes a wide range of factors, such as those relating to food, crops, livestock, forests, population, and gas emissions. Understanding how natural resources have been used historically in various nations is essential for sustainability. It is now possible to predict the demand for natural resources in each country by looking at historical usage patterns. With this knowledge, organizations, manufacturers, and policymakers can take proactive steps to meet future demand and ensure resource sustainability. One suggestion is to create a machine learning system that uses statistical regression models to identify the top k products that are most likely to experience shortages in particular nations over the course of the next x number of years. Absolute error and root mean square error measurements of the machine learning model's prediction performance have yielded encouraging results with few errors. This solution can assist businesses and producers in comprehending the productivity and sustainability required to satisfy global demand.**

Keywords*— *food production, machine learning, statistical regression models, absolute error, root mean square error, productivity, and sustainability

I. INTRODUCTION

The first principle of the Rio Declaration on Environment and Development [5] from 1992 states that "The well-being of humans should be the primary focus of efforts to promote sustainable development. People have a right to a life that is both healthy and fruitful, and that is lived in harmony with nature " Because there are now more than 7.9 billion people on Earth, it is becoming increasingly difficult for the vast majority of people to live healthy lives. Even now, there are still approximately 9.9% of people worldwide, or 811 million individuals, who go to night without having eaten anything. On the other hand, each and every year we throw away more than 1.3 billion metric tonnes of food. It is projected that there will be approximately 10 billion people living on Earth by the year 2050, as the population of the planet is expanding at an alarmingly high rate.

Those concerned with the environment have been looking for ways to cut down on both the number of people who go hungry and the amount of food that is wasted. The production of food in a sustainable manner assures that the current and future human population will not only have sufficient food to consume but also access to food of a high quality and adequate nutrition. Understanding the demand for a specific crop or food product that is required in a country is a necessary step in the process of transitioning to a sustainable food system. As a part of the process of achieving sustainable food development, the EAT-Lancet Commission [6] suggested five strategies as general beginning points for change on the national, regional, city, and local levels.

- Obtain international and national commitment to transition towards healthy diets.
- Reorient agricultural priorities away from generating high amounts of food and towards creating nutritious food.
- Reduce food losses and waste by at least half in accordance with the United Nations Sustainable Development Goals.
- Intensify food production in a sustainable manner in order to raise the quality of the output.
- In light of these tactics, it would be to the advantage of humans to have a method for determining the demand for a specific crop or food product that is required in each country.

The issue can be remedied by designing a machine learning system that has the capacity to forecast the types of food items that will become in short supply in a particular region at a certain point in time. In order to solve the problem, a vector autoregressive (VAR) statistical model has been utilised. This model accounts for the impact that the emission of greenhouse gases has on the output of crops and cattle. Because it graphs the production trend of the top k items and lists them out, the outcome of the system is very easy to understand. Agriculturists would, as a result of the outcome. The forecasting of the production of crops, animals, or forestry products can only take into consideration a single environmental component using the method that has been proposed, which is a limitation of the approach.

II. RELATED WORK

In the study [1] that K. Lutoslawski and his colleagues conducted, a nonlinear autoregressive neural network was used to anticipate future food demand. The solution is made up of information gathered over the course of more than three and a half years and trained using a non-linear auto regressive exogenous neural network (NARXNN). The ideal models generated forecasts that had values for mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE) that were the lowest possible. Based on the calculation of crop biomass, the proposed prediction model [2] by Jianqiang Ren et al. is applied in order to make a prediction regarding the region's yield for wheat. In order to calculate crop biomass, a model of net primary production is applied. In the study [3] that attempts to forecast the prices of agricultural goods, the authors make use of a long-term memory (LSTM) model that is informed by wavelet analysis. According to the results of the studies, this model attained higher levels of performance and accuracy.

In order to solve the problem of food demand forecast, the studies described above use neural networks. On the other hand, statistical machine learning models such as ARIMA and VAR haven't been used in nearly as much research as they could.

The vector autoregressive regression model is going to be utilised in order to make projections regarding the amount of crops and livestock that will be produced by the solution that has been proposed.

Paper	**Author**	**Year**	**Work Done**	**Drawbacks**
"Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network," in IEEE Access, vol. 9, pp. 146123-146136, 2021,	K. Lutoslawski, M. Hernes, J. Radomska, M. Hajdas, E. Walaszczyk and A. Kozina	2021	They used nonlinear autoregressive neural network. The solution is made up of data that was gathered through NARXNN training sessions spanning over three and a half years.	The main limitation of the developed models is the lack of the possibility of analysing small datasets
"Machine Learning based Prediction of GDP using FAO Agricultural Data Set for Hungary,"IEEE 15th International Symposium on Applied Computational Intelligence andInformatics	Adedeji Charles Adeyemo; Bence Bogdandy; Zsolt Toth	2021	This paper presents experimental results on training various recurrent neural networks for modeling the changes of Agricultural and Gross Domestic Products.	Although the results from GRU are promising, it poses the challenge of training the network appropriately. Specifically, how the speed and accuracy can be balanced and made optimum.
Regional yield prediction for winter wheat based on crop biomass estimation using multi-source data	Jianqiang Ren; Su Li; Zhongxin Chen; Qingbo Zhou; Huajun Tang	2017	Judged from the results of validation, the method was effective and practical in yield estimation of winter wheat in a larger region	Regionalizing and optimizing the relationship between fPAR and MODIS-NDVI is the next work which should be done better.
Machine Learning for Price Prediction for Agricultural Products	Sussy Bayona-Oré, Rino Cerna Eduardo, Tirado Hinojoza	2021	The results show that the mostly commonly used research paradigm is positivism, the research is quantitative and longitudinal in nature and neural networks are the most commonly used algorithms.	Machine learning algorithms can be complex and difficult to interpret, making it challenging for farmers to understand the reasoning behind the predictions.
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems.	Willett W, Rockström J, Loken B, Springmann M, Lang T, Vermeulen S, et al	2019	According to the results of the studies, this model attained higher levels of performance and accuracy.	The report emphasizes a universal approach to healthy and sustainable diets, without taking into account cultural and traditional dietary patterns of different regions and populations.
Predicting Future Food Security in Sub-Saharan Africa: A Time Series Approach	R. M. J. Kadigi and S. G. Mwakaje	2022	The authors identify key factors that impact food security, such as climate change and population growth, and use statistical techniques to make predictions about future trends.	One potential drawback of these papers is that the accuracy of the predictions relies heavily on the quality and availability of the data used
Food Sustainability Index: A Time-Series Analysis	L. A. Zaccarelli, G. Pellegrini, and R. P. H. Peters	2019	The authors analyze a variety of factors, such as agricultural productivity, food waste, and food affordability, and use statistical techniques to identify trends and patterns.	Use complex models or statistical techniques that may be difficult to understand for individuals who are not familiar with these methods
Predicting Food Prices and Availability Using Machine Learning: Evidence from Brazil.	L. N. R. Chagas, M. C. Lopes, and M. C. S. Barreto	2021	The authors analyze a variety of data sources, such as weather patterns and economic indicators, and use machine learning algorithms to identify patterns and make predictions.	Have assumptions and limitations that impact the validity of the predictions the models used may assume that historical patterns will continue into the future, which may not always be the

				case in a rapidly changing world.
Forecasting Future Food Security in Asia: A Time Series Analysis.	R. K. Singh, A. Kumar, and P. B. Shyam	2021	The authors analyze a variety of factors, such as agricultural productivity, food prices, and population growth, and use statistical techniques to make predictions about future trends.	It may not be possible to generalize the findings to other regions or time periods, as different factors may impact food sustainability in different ways.

Table. Literature Survey

III. DATASET

The Food and Agricultural Organization Corporate Statistical Database is providing the dataset that will be used for the solution (FAOSTAT). Access to the database is free of charge, and it covers all FAO regional groupings from 1961 up to the most current year for which data is available. The database contains information pertaining to food and agriculture for over 245 countries and territories (2019). The following is a list of the three datasets that were selected from this database:

- Climate Change: Emissions Totals
- Forestry: Forestry Production and Trade
- Production: Crops and livestock products

A. *Change Climate Change: Emissions Totals*

The emissions that were created from both agricultural and forest land are summed up in the dataset. Emissions of methane (CH4), nitrous oxide (N2O), and carbon dioxide (CO2) are included in the gas emissions. These emissions come from sources such as agricultural production, livestock husbandry, forest management, and land usage. For the years 1961 through 2019, statistics on a country's yearly emissions may be found in the unit of measure known as kilotonnes. The dataset contains more than one million individual records.

	Area Code_x	Area	Item Code_x	Item	Element Code_x	Element	Year Code_x	Year	Value	Flag_x	...	EmissionItem	Element Code_y	EmissionElement	Year Code_y	Source Code	Source
0	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	7225	Emissions (CH4)	1961	3050	FAO TIER 1
1	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	724413	Emissions (CO2eq) from CH4 (AR5)	1961	3050	FAO TIER 1
2	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	723113	Emissions (CO2eq) (AR5)	1961	3050	FAO TIER 1
3	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Manure Management	7225	Emissions (CH4)	1961	3050	FAO TIER 1
4	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Manure Management	7230	Emissions (N2O)	1961	3050	FAO TIER 1

5 rows × 23 columns

Fig1. Sample Emission Dataset

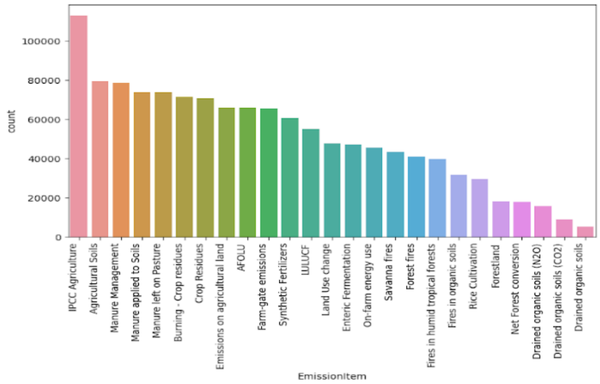

Fig2. Data Visualisation of Climate Change: Emission Totals

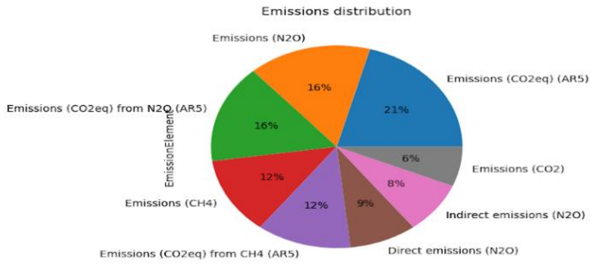

Fig3. Data Visualisation of Climate Change: Emission Totals

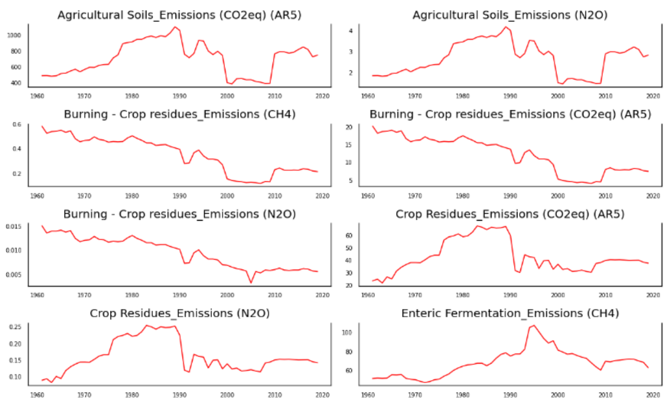

Fig4. Time series graph of Climate Change: Emission Totals

B. *Forestry: Forestry Production and Trade*

The collection consists of information regarding the production and trade of primary wood and paper products for all countries and territories across the world. This database contains a variety of primary forest products, including roundwood, sawnwood, wood-based panels, pulp, paper and paperboard, and paper and paperboard goods. There are more than two million cases included in the collection.

	Area Code_x	Area	Item Code_x	Item	Element Code_x	Element	Year Code_x	Year	Value	Flag_x	...	EmissionItem	Element Code_y	EmissionElement	Year Code_y	Source Code	Source
0	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	7225	Emissions (CH4)	1961	3050	FAO TIER 1
1	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	724413	Emissions (CO2eq) from CH4 (AR5)	1961	3050	FAO TIER 1
2	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	723113	Emissions (CO2eq) (AR5)	1961	3050	FAO TIER 1
3	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Manure Management	7225	Emissions (CH4)	1961	3050	FAO TIER 1
4	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Manure Management	7230	Emissions (N2O)	1961	3050	FAO TIER 1

5 rows × 23 columns

Fig5. Sample of Forest Dataset

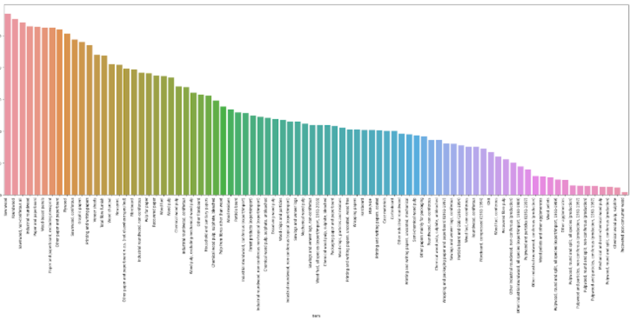

Fig6. Data Visualisation of Forestry Production and Trade

C. *Production: Crops and livestock products*

Statistics on crop and livestock production are included in this collection for more than 173 different commodities, covering all countries and regions. These products are referred to as Crops Primary, Fibre Crops Primary, Cereals, Coarse Grain, Citrus Fruit, Fruit, Jute Jute-like Fibres, Oilcakes Equivalent, Oil Crops Primary, Pulses, Roots and Tubers, Tree Nuts and Vegetables and Melons. The information is presented in terms of harvested area, total production quantity, and total yield. In total, the database has almost 3.8 million unique occurrences of this.

	Area Code_x	Area	Item Code_x	Item	Element Code_x	Element	Year Code_x	Year	Value	Flag_x	...	EmissionItem	Element Code_y	EmissionElement	Year Code_y	Source Code	Source
0	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	7225	Emissions (CH4)	1961	3050	FAO TIER 1
1	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	724413	Emissions (CO2eq) from CH4 (AR5)	1961	3050	FAO TIER 1
2	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Enteric Fermentation	723113	Emissions (CO2eq) (AR5)	1961	3050	FAO TIER 1
3	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Manure Management	7225	Emissions (CH4)	1961	3050	FAO TIER 1
4	3	Albania	515	Apples	5312	Area harvested	1961	1961	NaN	M	...	Manure Management	7230	Emissions (N2O)	1961	3050	FAO TIER 1

5 rows × 23 columns

Fig7. Sample of Production Dataset

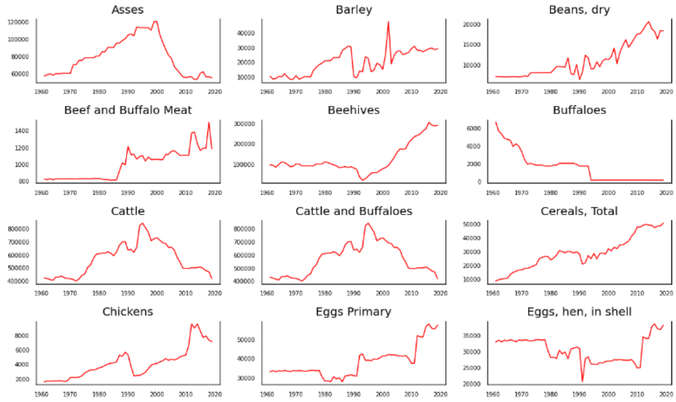

Fig8. Time series graph of Crops and livestock products

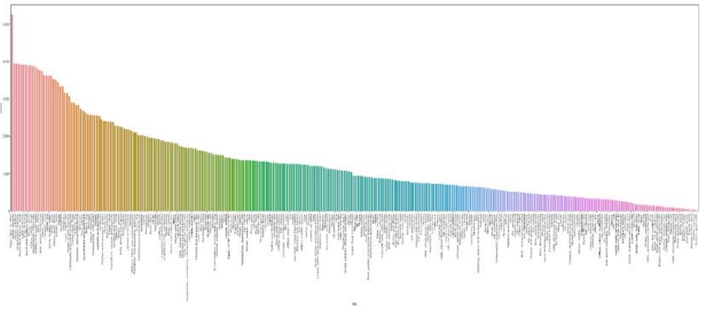

Fig9. Data Visualisation of Crops and livestock products

IV. DATA PREPROCESSING

The dataset was investigated further so that it could better handle the comprehensive data because its size is too enormous. It is clear that the data collected in each nation is independent of the data collected in the other countries; for example, the amount of grapes produced in India would not have an effect on the amount of apples produced in China. Because of this, it is able to train the data for each country independently, which drastically reduces the amount of data required from 300 million to 1 million. The data from each country is trained individually using that country's production, emission, and forestry datasets. The following is a list of the steps involved in the preparation of data:

1) Disabling aspects that aren't needed, such as Area, Source, and Emission Unit. This feature is removed because the values in these rows, such as "FAO TIER 1," do not provide any useful information that is required for forecasting;

   a. Area - This feature is removed because the country that is being used for prediction is already known.

   b. Source - This feature is removed because the values in these rows, such as "FAO TIER 1," do not provide any useful information.

   c. Emission Unit - This feature is removed because the emission unit in terms of kilotonnes.

2) Eliminating occurrences that are redundant

   Case 1:The production crop dataset includes three sample occurrences, which are presented in Table 1. Although there are three instances for the production of carrots, namely yield, area harvested, and production, each component may be estimated from the other element. These examples are yield, area harvested, and production. Production equals 800 tonnes, which multiplied by 10,000 equals 8,000,000 hg Yield can be calculated by dividing total production by the area harvested, which equals 66667 hg per hectare.

Case 2: Comparatively, in Table 2, Indirect emission in addition to Direction emission equals Emission (CH_4) 0.005100 in addition to 0.015700 equals 0.020800. Because of this, we have a better understanding that the indirect and direct instances, as well as the total emission, might be deleted from the model because they do not contribute anything of value to it. The existence of redundant instances results in an increase in the size of the data. Through the use of domain knowledge, one is able to increase the number of examples in the datasets that are of use.

The existence of redundant instances results in an increase in the size of the data. It is possible, with the application of domain knowledge, to cut down on the amount of usable examples that are contained inside the datasets.

3) Eliminating characteristics that have NaN values

During this stage of the process, features from the dataset that contain at least one NaN value will be extracted in their entirety. After carrying out the two procedures listed below, we arrived at this conclusion:

Experiment 1: In the beginning, testing consisted solely of omitting features that included approximately thirty percent NaN values. The missing data were filled in by the mean imputation method. Nevertheless, this method was flawed since it did not take into consideration the fact that the data was presented as a time series, and the values that were imputed did not make a lot of sense.

In the second experiment, the missing values were imputed by using the linear interpolation approach, and either forward or backward fills were utilised. This solution did not correspond to the problem statement because many features contained data from the years 1961 to 2000, and several characteristics contained values that were absent for the years 2000 to 2019. Due to the variance in the dataset, using a single form of linear interpolation in different situations would not produce accurate results.

4) Pivot Table

A data table can have its contents reorganized with the use of something called a pivot table. The data, which mostly comprises of aspects such as "Emission Element," "Emission Amount," "Production crop," and "Year," amongst others, were translated to tables with respect to the Year and Emission/Production crop. Figures 5 and 7 present the updated dataset, respectively. A production crop database and a forestry database have been combined to form a concatenated dataset

Features (for every country and product) (for every country and product)

1. Year (from 1961 to 2019) (from 1961 to 2019)

2. Characteristics of emanation (in kilotonnes)

a. The management of manure

b. The cultivation of rice

c. The use of synthetic fertilisers

d. The residues of crops

e. Up to the total number of emission features that are determined by the dataset for the country

3. One Productive crop or Forestry product, expressed in hundredweights per hectare

Exploratory Data Analysis

EDA is a strategy that is utilised in this project to analyse huge data sets to gain a better understanding of how various features behave over time. This analysis frequently makes use of statistical graphics and other types of data visualisations. EDA is helpful in many situations for understanding outliers or class imbalances, assuming there are any. Since 1961, the trend graphs for Asia's agricultural and livestock production are displayed in Fig. 7.

V. Experiments

Suggested model - Vector Autoregression model (VAR model) and Vector Error Correction Model (VECM model)

When there are multiple variables that influence one another, a VAR model and VECM model, which are a technique for multivariate forecasting and regression, are utilised. Before a dataset can be educated using the statistical model, it must first be able to successfully complete two tests, which are as follows:

1. The Granger Causality Test is a statistical test that checks to see if the prediction of one variable is influenced by another one. It examines whether or not the values of x in the past can help in predicting yt.

If this is the case, x is considered to be the "Granger cause" of y. The Granger causality matrix of one producing crop that has been impacted by emission gases. If a p-value is lower than the crucial threshold of 0.05 or 0.01, the associated X series (column) causes the Y series to change (row).

	Year	Asses_forecast	Asses	Asses_Unscaled	Asses_Forecast_Unscaled
0	2011	0.000000	0.333333	56000.0	91594.31
1	2012	0.232105	0.333333	56000.0	127298.67
2	2013	0.055085	0.000000	53000.0	100067.92
3	2014	0.458100	0.055556	53500.0	162063.11
4	2015	0.415385	0.777778	60000.0	155492.35
5	2016	0.595122	1.000000	62000.0	183140.99
6	2017	1.000000	0.333333	56000.0	245422.64
7	2018	0.507096	0.333333	56000.0	169600.10
8	2019	0.425815	0.222222	55000.0	157096.79

Fig 10. Granger Causality Test

2. The Augmented Dickey–Fuller (ADF) Test is a statistical test that is used to determine whether or not a particular time series is stationary. A stationary time series is one in which the statistical features or moments

being measured (such as the mean and variance) do not change over the course of the series.

VAR models are developed with consideration given to their order, which indicates the amount of previous time periods that will be incorporated into the model. For instance, a 4th-order VAR would model the quantity of apricots produced in each year as a linear combination of the amounts produced in the preceding four years. The equation that describes a pth-order VAR model can be found here.

$$yt = \gamma + A1yt-1 + ... + Apyt-p + ut, t \in Z \qquad (1)$$

According to the equation, yt represents the value of that variable I time periods earlier. This value is referred to as the "ith lag" of yt. The intercept of the model is denoted by the variable denoted by the vector of constants denoted by the letter. Ap is a time-invariant (k k)-matrix, while ut is a k-vector of error terms. Both matrices have the same size.

```
ADF test:
    Augmented Dickey-Fuller Test on "EmissionValue"
    -----------------------------------------------
 Null Hypothesis: Data has unit root. Non-Stationary.
 Significance Level    = 0.05
 Test Statistic        = -10.2756
 No. Lags Chosen       = 19
 Critical value 1%     = -3.437
 Critical value 5%     = -2.864
 Critical value 10%    = -2.568
 => P-Value = 0.0. Rejecting Null Hypothesis.
 => Series is Stationary.
```

Fig.11 The Augmented Dickey–Fuller (ADF) Test

VI. TRAINING AND INFERENCE

The flow chart of our methodology is given below.

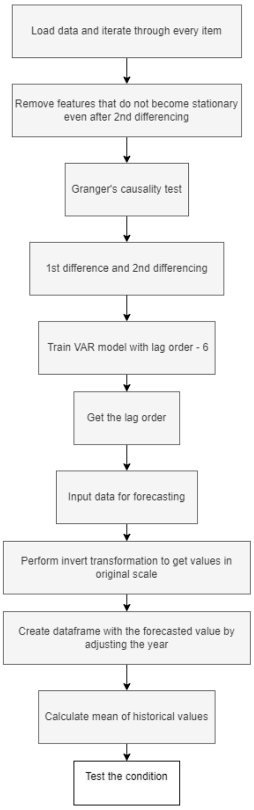

Fig12.Flow Chart

Following the input from the user comes a sequence of events for the datasets that include both the training and the inference processes. The user has the ability to specify the region to be forecasted, the number of years ahead, and the top-k products that continue to have a scarcity in the given year. The relevant dataset from the database is retrieved as a result of the information that was provided. After being cleaned and preprocessed, the datasets are then put through two tests (the ADF and the Granger causality test) before being used as input for the VAR model with a lag order of 6. The results of the model are utilised in the process of making forecasts regarding the quantity of all items produced by crops, animals, or forests. The values of the anticipated output that are less than half the mean of the previous values have been separated since it is more likely that they may experience shortages. After applying this filter to the data collection, the values are then sorted by the formula "abs(recent forecast – historical mean)," and the top k items together with their trend graphs are sent to the user as outputs.

Figure 11 presents the results of the calculations.

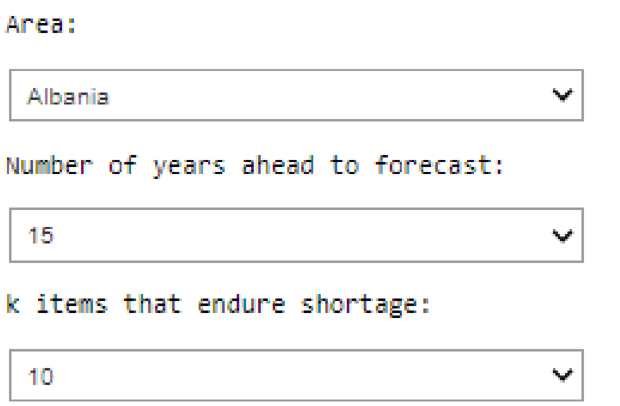

Fig13. User Inputs

```
Top - 10 products that will face shortage by 2034
=============================================
Vetches
Tobacco, unmanufactured
Sugar beet
Sugar Crops Primary
Skins, goat, fresh
Sheep and Goats
Sheep and Goat Meat
Seed cotton
Rye
Pulses, Total
```

Fig14. Predicted Output for Albania

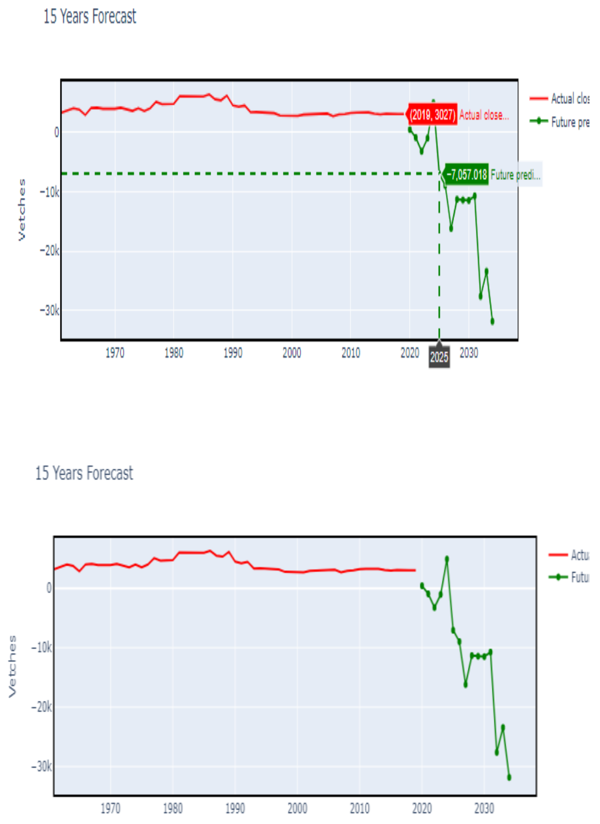

Fig15. Predicted outputs of vetches for Albania

.

VII. Evaluation Results

Experimentation is carried out in a variety of different ways so that the performance of the trained model may be evaluated.

A training set consisting of 50 historical values spanning the years 1961 to 2010 and an evaluation set have been carved out of the dataset (9 historical values - 2011 to 2019). Using the 50 historical value dataset, the data is trained with a VAR model that has a lag order of 6, and this model has a lag order of 6. The model is utilised to make projections regarding the next nine values. In order to compute the evaluation metrics, the anticipated values are judged against the evaluation set.

The evaluation forecasting graphs for two different production crops and livestock are displayed in Figure 14. It is clear from the pattern that the projection is accurate to the actual output quantities because of how they compare.

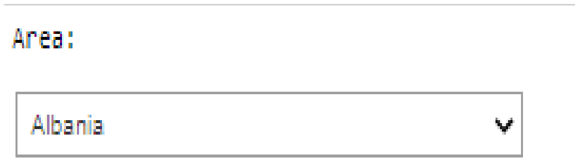

Fig16. User Inputs for Evaluation Results

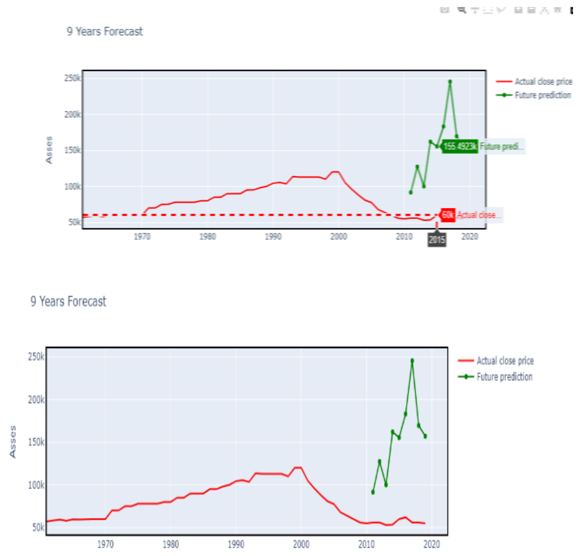

Fig.17: 9 Years Evaluation Forecasting Graph of VAR model

We have also evaluated using the VECM model

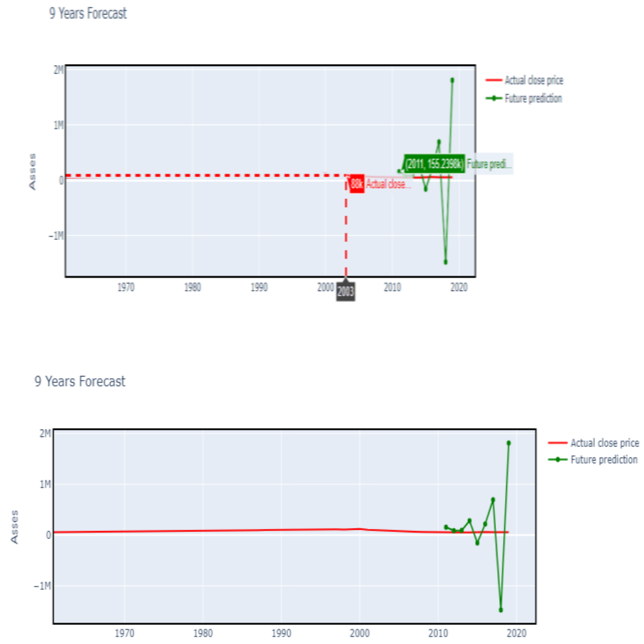

Fig.18: 9 Years Evaluation Forecasting Graph of VECM model

The evaluation metrics for a time series regression are as follows,

- Mean absolute error (MAE)

The MAE is the average of the absolute difference that exists between the values that were predicted and those that were actually observed. According to equation 2, yi represents the expected value, xi stands for the actual value, and n stands for the total number of values that make up the test set. The value of the MAE should be kept as low as possible for the model to function at its best.

MAE = (1/n) * Σ|yi – xi | (2)

- Mean squared error (MSE)

The mean squared error, or MSE, is the average of the error squares. This metric takes into account both the variance, which refers to the difference between the values that were anticipated, and the bias, which refers to the amount that the predicted value deviates from the actual value.

MSE = (1/n) * Σ(yi – xi) ^2 (3)

- Root mean squared error

The square root of the mean square error is the value that is being measured here. Based on equation 4, in where yi denotes the expected value, xi is the actual value, and n denotes the total number of values contained in the test set. In order for the model to be considered optimal, it is necessary for each of the three error terms to be as small as feasible.

RMSE = √(Σ(yi – xi) 2 \s/ n) (4)

The following is the comparison of the average performance metric for the trained VAR model and VECM model

	VAR	VECM
Mean absolute error	0.3916	0.43433
Mean squared error	0.23006	0.26571
Root mean squared error	0.4622	0.504205

Table: Comparison between Models

The findings of the experiment demonstrate that the VAR model is an adequate model for forecasting crop production when it is subject to the influence of emission of gases.

A. *Residual Analysis of VAR and VECM model*

Residual analysis is an important step in the validation of any statistical model, including VAR (Vector Autoregression) and VECM (Vector Error Correction Model). Residuals are the differences between the predicted values and the actual values of the dependent variable in the model. Analyzing the residuals can help us determine if the assumptions of the model are met and if there is any systematic pattern.

Below attached is the analysis of both VAR (Vector Autoregression) and VECM (Vector Error Correction Model). Figure 19 represents the residual analysis plot of VAR model. Figure 20 represents the residual analysis plot of VECM model.

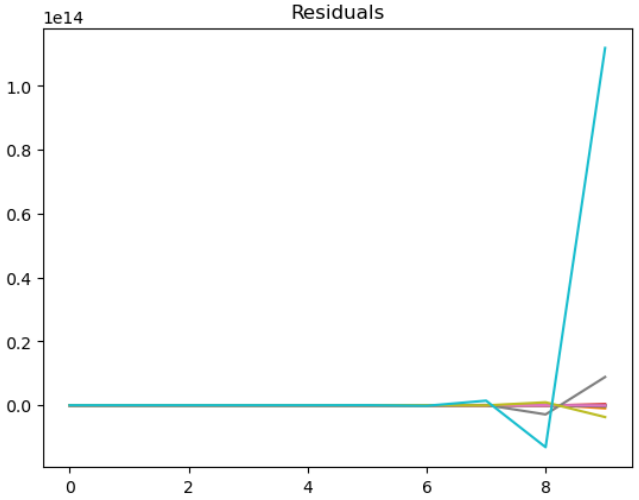

Fig 19. Residual Analysis of VAR model

- Durbin-Watson statistic for Enteric Fermentation: 1.481671915890381
- Durbin-Watson statistic for Manure Management: 1.541377335530611
- Durbin-Watson statistic for Synthetic Fertilizers: 1.3233251989100154
- Durbin-Watson statistic for Manure applied to Soils: 1.598807216980558
- Durbin-Watson statistic for Manure left on Pasture: 1.6590553568395352
- Durbin-Watson statistic for Crop Residues: 0.8560135925064513
- Durbin-Watson statistic for Burning - Crop residues: 1.523093599343073
- Durbin-Watson statistic for IPCC Agriculture: 1.6889211309457455
- Durbin-Watson statistic for Agricultural Soils: 1.5347257751850454
- Durbin-Watson statistic for Asses: 1.2484057959795898

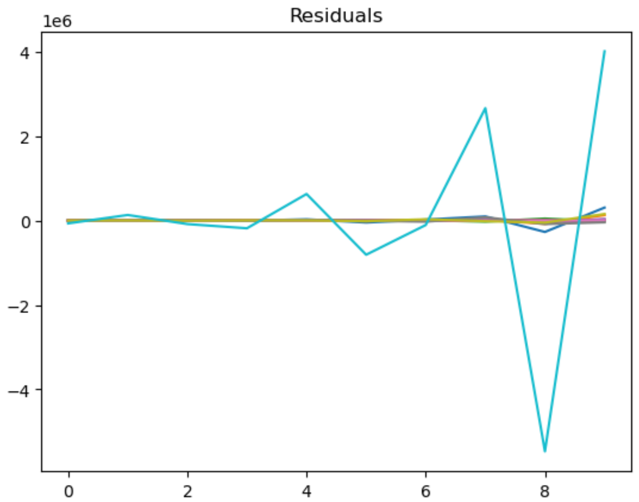

Fig 20. Residual Analysis of VECM model

- Durbin-Watson statistic for Enteric Fermentation: 2.69939037650087
- Durbin-Watson statistic for Manure Management: 2.29003824365793
- Durbin-Watson statistic for Synthetic Fertilizers: 3.2528381898002094
- Durbin-Watson statistic for Manure applied to Soils: 2.935032029495343
- Durbin-Watson statistic for Manure left on Pasture: 2.9350060468537054
- Durbin-Watson statistic for Crop Residues: 2.9258448862987865
- Durbin-Watson statistic for Burning - Crop residues: 2.376988605651661
- Durbin-Watson statistic for IPCC Agriculture: 2.5078864678359007
- Durbin-Watson statistic for Agricultural Soils: 1.6980089862229761
- Durbin-Watson statistic for Asses: 3.0803183310980033

VIII. CONCLUSION

Under the effect of gases, the goal of the research project is to develop a system that is able to make accurate forecasts regarding the output of crops, animals, and forestry products. It helps humans recognise the direct influence that food production has on environmental pollution by highlighting the fact that more than 30 percent of the world's greenhouse gas emissions are directly emitted owing to food production. In order to solve the problem, the proposed approach trains a vector autoregressive model with data from the United Nations' database on food and agriculture. The model of forecasting is able to anticipate trends in production amounts as well as the top-k goods that will continue to experience shortage in the future. The findings of the solution contribute to a better understanding of the significance of sustainability, which is important in order for developed nations to make the necessary adjustments to their lifestyles before it is too late. The findings may be of use to governments in less developed nations as well as those in nations that are still in the process of developing in terms of assisting agriculturalists in the production of an adequate quantity of food products to meet the needs of the nation.

IX. LIMITATIONS AND FUTURE WORK

The following is a list of some of the restrictions that apply to the current work:

- The solution that has been suggested takes into account only one environmental component, and that is the emission of gases in order to make projections for the production of crops, animals, or forestry products. As a result, the values that have been anticipated could not be correct in real time because the model does not take into consideration other factors such as rainfall, drought, population growth, and so on.
- It is likely that the model will not learn the influence of some petrol emissions in an effective manner if numerous features that contain NaN values are omitted. This could result in the model's accuracy being negatively impacted.

The work that will be done in the future on this project will centre on collecting additional records on changing temperatures, rainfall, population, and trade flows. These datasets will add complicated aspects that will be used to determine the production of crops in the future. To convert the system into a reliable model for the prediction of sustainability, the solution can also be expanded to incorporate social and economic sustainability. Technology corporations such as Microsoft and Amazon Web Services are increasing their investments in the development of sustainable solutions to lower their companies' overall carbon footprints. The findings of this research shed light on the considerations that should be given by an organisation when developing environmentally friendly strategies. This research could be further enhanced to become an in-built library to generate sustainable factor predictions on social, environmental, and economic aspects if the appropriate dataset is used.

X. REFERENCES

[1] K. Lutoslawski, M. Hernes, J. Radomska, M. Hajdas, E. Walaszczyk and A. Kozina, "Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network," in IEEE Access, vol. 9, pp. 146123-146136, 2021, doi: 10.1109/ACCESS.2021.3123255.

[2] Jianqiang Ren, Su Li, Zhongxin Chen, Qingbo Zhou and Huajun Tang, "Regional yield prediction for winter wheat based on crop biomass estimation using multi-source data," 2007 IEEE International Geoscience and Remote Sensing Symposium, 2007, pp. 805-808

[3] Q. Chen, X. Lin, Y. Zhong and Z. Xie, "Price Prediction of Agricultural Products Based on Wavelet Analysis-LSTM," 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), 2019, pp. 984-990, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00142

[4] A. C. Adeyemo, B. Bogdandy and Z. Toth, "Machine Learning based Prediction of GDP using FAO Agricultural Data Set for Hungary," 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2021, pp. 000473-000478, doi: 10.1109/SACI51354.2021.9465608.

[5] Willett W, Rockström J, Loken B, Springmann M, Lang T, Vermeulen S, et al. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. The Lancet. 2019 Jan 16.

[6] Report of the United Nation Conference on Environment and Development, Rio de Janeiro,June 1992

[7] Westerveld, J. J., van den Homberg, M. J., Nobre, G. G., van den Berg, D. L., Teklesadik, A. D., & Stuit, S. M. (2021, September). Forecasting transitions in the state of food security with machine learning using transferable features. *Science of the Total Environment*, *786*, 147366. https://doi.org/10.1016/j.scitotenv.2021.147366

[8] Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., Husain, A., & Omodei, E. (2022, September 15). Machine learning can guide food security efforts when primary data are not available. Nature Food, 3(9), 716–728. https://doi.org/10.1038/s43016-022-00587-8

[9] Foini, Pietro & Tizzoni, Michele & Paolotti, Daniela & Omodei, Elisa. (2021). On the forecastability of food insecurity. 10.1101/2021.07.09.21260276.

[10] Stanley Jothiraj, F. V. (2022, September 14). *Time Series Prediction for Food sustainability*. arXiv.org. https://arxiv.org/abs/2209.06889v1