

## **Time Series Prediction for Food sustainability**

### **Abstract**

With exponential growth in the human population, it is vital to conserve natural resources without compromising on producing enough food to feed everyone. This process can improve people's livelihoods, health, and ecosystems for the present and future generations. Sustainable development, a paradigm of the United Nations, is rooted in food, crop, livestock, forest, population, and even emission of gases. By understanding the overall usage of natural resources in different countries in the past, it is possible to forecast the demand in each country. The proposed solution consists of implementing a machine learning system using a statistical regression model that can predict the top-k products that would endure a shortage in each country in a specific period in the future. The prediction performance in terms of absolute error, root mean square error show promising results due to its low errors. This solution could help organizations and manufacturers understand the productivity and sustainability needed to satisfy the global demand.

### **Introduction**

According to the first principle of the 1992 Rio Declaration on Environment and Development [5], "Humans are at the center of concern for sustainable development. They are entitled to a healthy and productive life that is in harmony with nature." With over 7.9 billion humans, it is getting harder for the majority of the population to lead a healthy life. Around 9.9% of the population, which accounts for 811 million people, still go to bed on an empty stomach. On the contrary, over 1.3 billion tonnes of food are wasted every year. The world's population is rapidly growing, and it is estimated that there will be around 10 billion people on Earth by the year 2050.

Environmentalists have been trying to find solutions to reduce the numbers in terms of hunger and food wastage. Sustainable food development ensures that the current and future human population has enough food to eat and access high-quality, nutritious foods.

Transitioning to a sustainable food system involves understanding the demand for a particular crop/food product required in each country. The EAT-Lancet Commission [6] proposed five strategies as general starting points for national, regional, city, and local change as part of sustainable food development,

- Seek international and national commitment to shift toward healthy diets
- Reorient agricultural priorities from producing high quantities of food to producing healthy food
- Sustainably intensify food production to increase high-quality output.
- Strong and coordinated governance of land and oceans
- At least halve food losses and waste, in line with UN Sustainable Development Goals

With these strategies in mind, a means of knowing the demand for a particular crop/food product required in each country would be beneficial to humans.

The problem can be addressed by developing a machine learning system that could predict the food products that would endure shortages in a particular area at a specific time. The solution uses a vector autoregressive (VAR) statistical model that captures the influence of the production of crops and livestock due to the emission of greenhouse gases. The result of the system is quite simple to comprehend since it lists out the top-k products and its production trend graphs. From the result, agriculturists would. The limitation of the proposed approach is that the solution only accounts for a single environmental factor - the emission of gases to forecast the production of crops, livestock, or forestry products.

## **Related Work**

The research [1] by K. Lutoslawski et al. uses a nonlinear autoregressive neural network for food demand prediction. The solution consists of data collected from over 3.5 years trained with a non-linear auto regressive exogenous neural network (NARXNN). The optimal models provided forecasts with the lowest mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE) values. The proposed prediction model [2] by Jianqiang Ren et al. is used to predict the region yield for wheat-based on crop biomass estimation. It uses a net primary production model to estimate crop biomass. The paper on price prediction for agricultural products [3] uses a long short-term memory (LSTM) model using wavelet analysis for prediction. The experiments show that this model achieved better performance and accuracy. The papers mentioned above use neural networks to solve food demand prediction. However, not much research has been done using statistical machine learning models like ARIMA or VAR. The proposed solution aims to solve this challenge by using the vector autoregressive regression model to predict the production of crops and livestock.

## **Data**

### **Datasets:**

The dataset for the solution is provided by Food and Agriculture Organization Corporate Statistical Database (FAOSTAT). The database provides free access to food and agriculture data for over 245 countries and territories and covers all FAO regional groupings from 1961 to the most recent year available (2019). The three datasets chosen from this database are as follows,

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

#### *1. Climate Change: Emissions Totals:*

The dataset summarizes the emissions generated from agricultural and forest land. The gas emissions consist of methane (CH<sub>4</sub>), nitrous oxide (N<sub>2</sub>O), and carbon dioxide (CO<sub>2</sub>) emissions

from sources like crops, livestock, forest management, and land use. Emissions data in the units of kilotonnes are available for each country for the period 1961-2019. The dataset consists of over a million instances.

	Area	EmissionItem	EmissionElement	Year	Source	EmissionUnit	EmissionValue
0	Afghanistan	Enteric Fermentation	Emissions (CH4)	1961	FAO TIER 1	kilotonnes	240.6831
1	Afghanistan	Enteric Fermentation	Emissions (CH4)	1962	FAO TIER 1	kilotonnes	245.3106
2	Afghanistan	Enteric Fermentation	Emissions (CH4)	1963	FAO TIER 1	kilotonnes	255.8285
3	Afghanistan	Enteric Fermentation	Emissions (CH4)	1964	FAO TIER 1	kilotonnes	259.0650
4	Afghanistan	Enteric Fermentation	Emissions (CH4)	1965	FAO TIER 1	kilotonnes	265.5980

**Fig 1. Emissions dataset**

## 2. Forestry: Forestry Production and Trade

The dataset contains information on the production and trade in primary wood and paper products for all countries and territories worldwide. The primary forest products included in this database are Roundwood, sawnwood, wood-based panels, pulp, and paper and paperboard. The dataset consists of over two million instances.

	Area	Item	Element	Year	Value
0	Afghanistan	Forest products (export/import)	Import Value	1964	301.0
1	Afghanistan	Forest products (export/import)	Import Value	1965	321.0
2	Afghanistan	Forest products (export/import)	Import Value	1966	435.0
3	Afghanistan	Forest products (export/import)	Import Value	1967	466.0
4	Afghanistan	Forest products (export/import)	Import Value	1968	297.0

**Fig 2. Forestry production dataset**

## 3. Production: Crops and livestock products

The dataset contains crop and livestock production statistics for all countries and regions for over 173 items. The items include 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 data is expressed in area harvested, production quantity, and yield. There are over 3.8 million instances found in the database.

	Area	Item	Element	Year	Value
0	Afghanistan	Almonds, with shell	Area harvested	1975	0.0
1	Afghanistan	Almonds, with shell	Area harvested	1976	5900.0
2	Afghanistan	Almonds, with shell	Area harvested	1977	6000.0
3	Afghanistan	Almonds, with shell	Area harvested	1978	6000.0
4	Afghanistan	Almonds, with shell	Area harvested	1979	6000.0

**Fig 3. Crops and livestock products dataset**

### **Preprocessing:**

Since the data size is too large, the dataset was explored more to deal with extensive data. It is evident that each country's data is independent of the other, i.e., the production of grapes in India would not affect the production of apples in China. This makes it possible to train each country's data separately, significantly reducing the data size from 300 million to ~1 million. Each country's data is individually trained with its emission, production, and forestry datasets.

The steps involved in preprocessing of data are as follows:

#### *1. Removing unwanted features like Area, Source and Emission Unit*

- a. Area - The feature is removed since the country that is being used for prediction is already known
- b. Source - This feature is removed since the values in these rows like 'FAO TIER 1' provides no valuable information that is required for forecasting
- c. Emission Unit - The feature is removed since the emission unit in terms of kilotonnes is common throughout the dataset

#### *2. Removing redundant instances*

##### Case 1:

Table 1 shows three sample instances from the production crop dataset. Although there are three instances for the production of carrots, namely, yield, area harvested, and production, each element can be calculated from the other.

Production = 800 tonnes =  $800 * 10,000 = 8,000,000$  hg

Production / Area harvested =  $8,000,000 / 120 = 66667$  hg/ha = Yield

**Table 1. Production crop sample data**

Country	Year	Production crop	Element	Value
Panama	1961	Carrots	Yield	66667 hg/ha
Panama	1961	Carrots	Area harvested	120 ha
Panama	1961	Carrots	Production	800 tonnes

##### Case 2:

Similarly in Table 2, Indirect emission + Direct emission = Emission (CH<sub>4</sub>)  $0.005100 + 0.015700 = 0.020800$ . This helps us understand that the indirect and direct instances (or) total emission could be removed since they add no value to the model. The presence of redundant instances increases the dataset size. Using domain knowledge, the number of useful instances present in the datasets can be improved.

**Table 2. CH4 Emission sample data**

Country	Year	Emission	Value (kilotonnes)
Panama	1961	Indirect (CH4)	0.005100
Panama	1961	Direct (CH4)	0.015700
Panama	1961	Total emission (CH4)	0.020800

The presence of redundant instances increases the datasize. Using domain knowledge, the number of useful instances present in the datasets can be reduced.

### 3. Removing features having NaN values

In this step, features consisting of at least one NaN value are completely scraped from the dataset. This decision was made after performing the following two steps,

*Experiment 1:* Initially, experimentation was done by only removing features that are about 30% of NaN values. The mean imputation imputed the missing values. However, this method was wrong since it did not account for the factor of being a time series, and the imputed values did not make much sense.

*Experiment 2:* Using forward or backward fills, the linear interpolation method was used to impute the missing values. This method did not fit the problem statement since multiple features had data from 1961-2000, and several features had missing values for 2000 - 2019. The inconsistency in the dataset meant that one type of linear interpolation would not work for all circumstances.

### 4. Pivot Table

A pivot table is used to reorganize the data stored in a table. Data that essentially consists of features like ‘Emission Element’, ‘Emission Amount’, ‘Production crop’, ‘Year’ etc. was converted to tables with respect to the Year and Emission/Production crop. The modified dataset is shown in Figure 4 and Figure 5. Figure 4 is a pivot table of a concatenated dataset consisting of a production crop and forestry databases.

Item	Agave fibres nes	Asses	Avocados	Bananas	Bastfibres, other	Beans, dry	Beans, green	Beef and Buffalo Meat	Beehives	Beeswax	...	Taro (cocoyam)	Tobacco, unmanufactured	Tomatoes	Turkeys
Year															
1961	11100.0	330400.0	946662.0	100204.0	10000.0	4837.0	72672.0	1642.0	234000.0	15.0	...	91185.0	10763.0	128162.0	171.0
1962	8800.0	335950.0	969305.0	98399.0	10000.0	4833.0	70177.0	1614.0	224500.0	16.0	...	91648.0	8063.0	123857.0	172.0
1963	10000.0	341400.0	966337.0	93520.0	10000.0	4690.0	70034.0	1671.0	230600.0	16.0	...	92685.0	8890.0	114796.0	173.0
1964	10000.0	342550.0	956757.0	94221.0	10000.0	4731.0	65792.0	1651.0	222100.0	15.0	...	93716.0	9131.0	128956.0	174.0
1965	10000.0	345100.0	1086117.0	102901.0	10000.0	4714.0	62987.0	1649.0	233800.0	15.0	...	94700.0	8210.0	132700.0	175.0

**Fig 4. Production crop + forestry database**

Emission	Agricultural Soils_Emissions (CO2eq) (AR5)	Agricultural Soils_Emissions (N2O)	Burning - Crop residues_Emissions (CH4)	Burning - Crop residues_Emissions (CO2eq) (AR5)	Burning - Crop residues_Emissions (N2O)	Crop Residues_Emissions (CO2eq) (AR5)	Crop Residues_Emissions (N2O)
Year							
1961	4529.5436	17.0926	4.5391	158.2788	0.1177	94.0621	0.3550
1962	4910.2800	18.5294	4.3200	150.6390	0.1120	96.2801	0.3633
1963	5101.4486	19.2507	4.0331	140.6362	0.1046	94.2885	0.3558
1964	5383.2701	20.3142	3.8378	133.8264	0.0995	83.9561	0.3168
1965	5561.1966	20.9856	3.8132	132.9664	0.0989	79.8558	0.3013

**Fig 5. Emissions database**

The preprocessed dataset that is fed into the VAR (Vector Auto Regressive) model is of the structure as shown in Figure 6.

Year	Emissions							Food / forestry product	
	Manure Management	Rice Cultivation	Synthetic Fertilizers	Manure applied to Soils	Crop Residues	Burning - Crop residues	IPCC Agriculture	Agricultural Soils	Apples
1961	67177.5348	403962.0793	11280.5787	23567.8362	30683.8352	11030.2572	1.260293e+06	166204.7331	64850.0
1962	68832.9585	484990.7646	13368.3162	23471.8359	31958.1838	11038.1735	1.274900e+06	169497.3492	68007.0
1963	69327.2798	485127.6957	14978.9507	24246.1919	32856.3047	11216.0781	1.287168e+06	174006.3209	73024.0
1964	73622.0039	504259.4378	17328.0169	25545.9343	34191.8198	11657.4114	1.331906e+06	181064.9502	73526.0
1965	76861.9901	522776.5788	21187.0413	26513.2726	33829.7953	11632.8168	1.349149e+06	187350.8749	74238.0
1966	80132.8544	510131.2001	28117.1922	27536.4445	34870.2047	11941.0290	1.388311e+06	199144.2170	60878.0
1967	83271.2493	513056.3450	27449.6202	28504.3298	36093.1879	11869.2183	1.404688e+06	201854.0872	70078.0
1968	84932.3585	519288.9250	30715.1862	29136.9448	36830.7156	12096.1355	1.433394e+06	209341.1732	65877.0
1969	84574.0437	524966.2144	35282.6598	29133.8925	37424.1058	12252.3220	1.447795e+06	214895.1806	70532.0

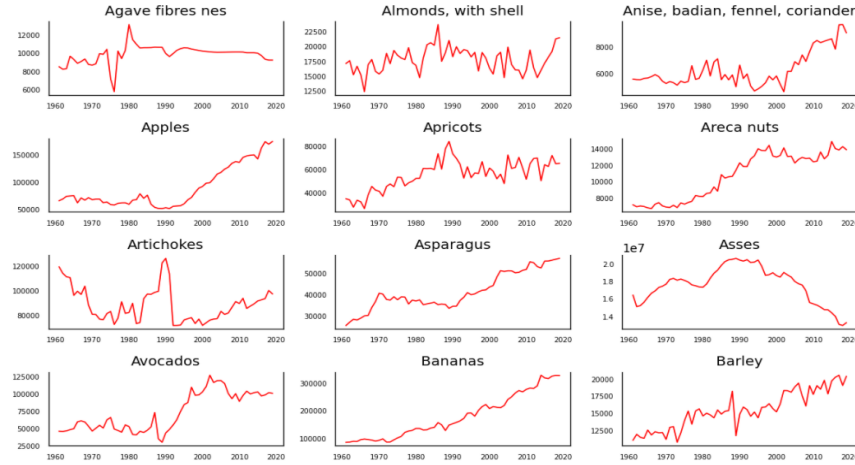
**Fig 6. Preprocessed data**

### Features (for every country and product)

- Year (from 1961 to 2019)
- Emission features (in kilotonnes)
  - Manure management
  - Rice cultivation
  - Synthetic fertilizers
  - Crop residues
  - Upto total number of emission features that is based on country's dataset
- One Production crop/Forestry product (in hg/ha)

### Exploratory Data Analysis

EDA in this project is an approach used to analyze large data sets to understand how different features perform over time, often using statistical graphics and other data visualizations. EDA often helps us understand outliers or class imbalance, if any. Fig 7 shows the trend graphs for production of crop/livestock in Asia since 1961.



**Fig 7. Time series graphs for crop/livestock production**

## Experiments

Proposed model - Vector Autoregression model (VAR model)

VAR model is a multivariate forecasting regression algorithm used when two or more variables influence each other. Before a dataset is trained with the statistical model, the dataset needs to pass two tests which are as follows,

1. *Granger Causality Test* - A statistics test to check whether one variable influences the prediction of another variable. It tests whether past values of  $x$  aid in the prediction of  $y$ . If so,  $x$  is said to “Granger cause”  $y$ . Figure 8 shows the granger causation matrix of one production crop affected by emission gases. From the figure, if a  $p$ -value is less than the critical threshold of 0.05 or 0.01, the corresponding  $X$  series (column) causes the  $Y$  (row).

	Manure Management_x	Rice Cultivation_x	Synthetic Fertilizers_x	Manure applied to Soils_x	Crop Residues_x	Burning - Crop residues_x	IPCC Agriculture_x	Agricultural Soils_x	Fibre Crops Primary_x
Manure Management_y	1.0000	0.2376	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.000
Rice Cultivation_y	0.0000	1.0000	0.0167	0.0000	0.0002	0.0014	0.1021	0.2804	0.000
Synthetic Fertilizers_y	0.0091	0.0066	1.0000	0.0240	0.0000	0.0006	0.0000	0.0000	0.001
Manure applied to Soils_y	0.0017	0.0001	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.000
Crop Residues_y	0.0000	0.0001	0.0010	0.0000	1.0000	0.0000	0.0000	0.0008	0.000
Burning - Crop residues_y	0.0000	0.0003	0.0000	0.0000	0.0000	1.0000	0.0334	0.0082	0.000
IPCC Agriculture_y	0.0099	0.0368	0.0000	0.0002	0.0000	0.0000	1.0000	0.0000	0.000
Agricultural Soils_y	0.0001	0.0033	0.0001	0.0002	0.0000	0.0000	0.0000	1.0000	0.000
Fibre Crops Primary_y	0.0422	0.0000	0.0002	0.0006	0.0296	0.0000	0.0006	0.0102	1.000

**Fig 8. Granger causation matrix for Fibre crops**

2. *Augmented Dickey–Fuller (ADF) Test* - Statistical test used to test whether a given time series is stationary or not. A stationary time series has statistical properties or moments (e.g., mean and variance) that do not vary in time

```

Augmented Dickey-Fuller Test on "Agricultural Soils"
-----
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level = 0.05
Test Statistic = 5.2806
No. Lags Chosen = 0
Critical value 1% = -3.566
Critical value 5% = -2.92
Critical value 10% = -2.598
=> P-Value = 1.0. Weak evidence to reject the Null Hypothesis.
=> Series is Non-Stationary.

Augmented Dickey-Fuller Test on "Apricots"
-----
Null Hypothesis: Data has unit root. Non-Stationary.
Significance Level = 0.05
Test Statistic = -3.0946
No. Lags Chosen = 0
Critical value 1% = -3.566
Critical value 5% = -2.92
Critical value 10% = -2.598
=> P-Value = 0.027. Rejecting Null Hypothesis.
=> Series is Non-Stationary.

```

	Apricots	Apricots_diff_1	Apricots_diff_2
Year			
1961	34471.0	NaN	NaN
1962	33381.0	-1090.0	NaN
1963	27217.0	-6164.0	-5074.0
1964	33213.0	5996.0	12160.0
1965	31811.0	-1402.0	-7398.0
1966	26002.0	-5809.0	-4407.0
1967	37699.0	11697.0	17506.0
1968	45030.0	7331.0	-4366.0
1969	41856.0	-3174.0	-10505.0

**Fig 9. ADF Test and differencing results**

Figure 9 shows the ‘apricots’ series that is non-stationary. Hence it is subject to first and second differencing which results in a stationary series.

VAR models are constructed based on their order, which refers to the number of historical time periods the model will use. For example, a 4th-order VAR would model each year's apricot production quantity as a linear combination of the last four years of apricot quantities. A pth-order VAR model is written as shown in equation 1

$$(1) \quad y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad t \in Z$$

From the equation,  $y_t$  indicates that variable's value  $i$  time periods earlier and is called the "ith lag" of  $y_t$ . The variable  $v$  is a vector of constants which is the intercept of the model.  $A_p$  is a time-invariant  $(k \times k)$ -matrix and  $u_t$  is a  $k$ -vector of error terms.

## Training and Inference

Area:

Asia

Number of years ahead to forecast:

10

k items that endure shortage:

5

**Fig 10. User Inputs**



Both the training and inference of the datasets take place sequentially after the user input. The user can provide the area to forecast, the number of years ahead, and the top-k items that endure shortage in the specific year. The provided information pulls up the relevant dataset from the database. The datasets are cleaned, preprocessed, subject to two tests (ADF and Granger causality test), and fed as input to the VAR model with a lag order of 6. The model output is used to forecast the production quantities of all crops, livestock, or forestry products. The forecasted output values 50% less than the mean of historical values are separated as they are more likely to have shortages. From this filtered set, the values are sorted by ‘abs(recent\_forecast – historical\_mean),’ and the top-k items and their trend graphs are presented as outputs to the user. The outputs are shown in Figure 11.

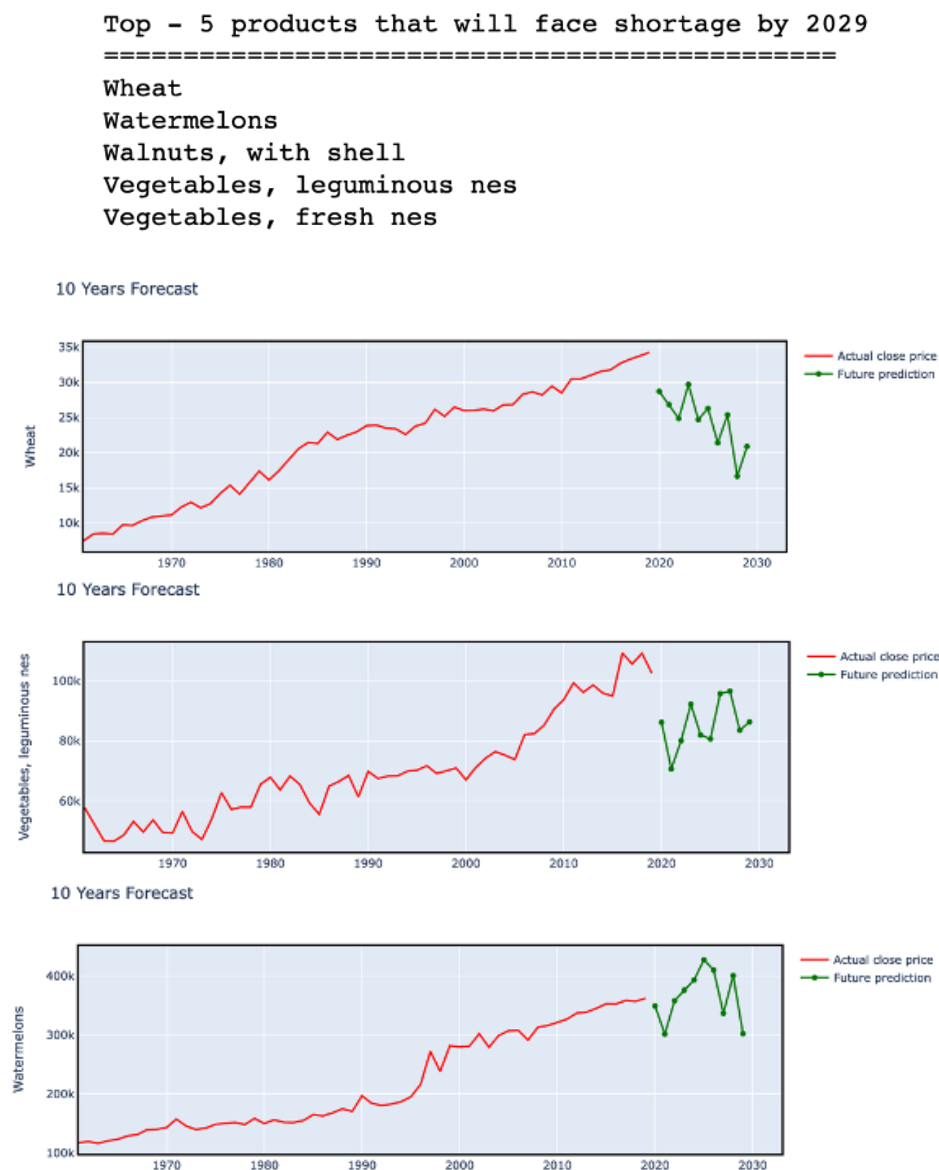
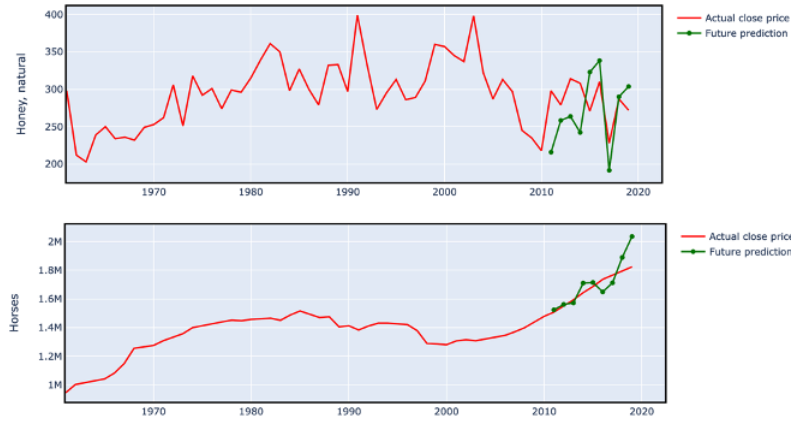


Fig 11. Predicted outputs for Asia

## Evaluation Results

In order to evaluate the performance of the trained model, the experiment is repeated differently. The dataset is split into a training set (50 historical values - 1961 to 2010) and an evaluation set (9 historical values - 2011 to 2019). The data is trained with a VAR model having a lag order of 6 using the 50 historical value dataset. The model is used to forecast the future nine values. The predicted values are compared against the evaluation set to calculate the evaluation metrics. Figure 12 shows the evaluation forecasting graphs for two production crops/livestock. From the trend, it is evident that the forecast is accurate to the actual production quantities.



**Fig 12. Evaluation forecasting graphs**

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

- *Mean absolute error (MAE)*

MAE is the average of the absolute difference between forecasted and true values. From equation 2,  $y_i$  is the expected value,  $x_i$  is the actual value and  $n$  represents the total number of values in the test set. The lower the MAE value, the more optimal the model.

$$MAE = (1/n) * \sum |y_i - x_i| \quad (2)$$

- *Mean squared error (MSE)*

MSE is the average of the error squares. This metric takes into account the variance (the difference between anticipated values) and bias (the distance of predicted value from its true value).

$$MSE = (1/n) * \sum (y_i - x_i)^2 \quad (3)$$

- *Root mean squared error*

This measure is the square root of mean square error. From equation 4, Where  $y_i$  is the expected value,  $x_i$  is the actual value and  $n$  represents the total number of values in the test set. All three error terms need to be as low as possible for the model to be called optimal.

$$RMSE = \sqrt{(\sum(y_i - x_i)^2 / n)} \quad (4)$$

The average performance metric for the trained VAR model with lag order of 6 is as follows,

- Mean absolute error - 0.359
- Mean squared error - 0.203
- Root mean squared error - 0.434

The experiment results show that the VAR model is a sufficiently good model to forecast the production of crops when it is under the influence of emission of gases.

## **Conclusion**

The research project focuses on creating a system that can predict the production of crops, livestock and forestry products under the influence of gases. With over 30% of global Greengrass gas emissions directly released due to food production, it helps humans realize the direct impact of food production and environmental pollution. The proposed solution uses the food and agriculture database from the United Nations to train a vector autoregressive model. The forecast model can project trends of production quantities and the top-k items that would endure shortage in the future. The results of the solution help in understanding the importance of sustainability to make the necessary lifestyle changes in developed countries before it is too late. In underdeveloped and developing countries, the results could help the government and agriculturists produce the right quantity of food products to satisfy the country's demand.

## *Limitations*

Some of the limitations of the current work are listed below,

- The proposed solution only accounts for a single environmental factor - the emission of gases to forecast the production of crops, livestock or forestry products. Hence, the forecasted values may not be accurate in real-time since it does not account for other factors like rainfall, drought, population growth etc.
- By dropping many features that contain NaN values, it is possible that the effect of such gas emissions is not appropriately learned by the model leading to inaccuracies.

## **Future Work**

The future work of this project would focus on adding more datasets on Temperature change, Rainfall, Population, Trade flows that would add intricate factors to decide the production of

crops in the future. The solution can also be extended to cover social and economic sustainability to make the system a robust model for sustainability prediction. Technology companies like Microsoft and AWS are investing more in building sustainability solutions to reduce the carbon footprint. This research gives insights into what factors an organization should care about while designing sustainable solutions. Given the right dataset, this research can be further extended to be an in-built library to make sustainable factor predictions on social, environmental, and economic aspects.

## 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] [Report](#) of the United Nation Conference on Environment and Development, Rio de Janeiro, June 1992
- [6] 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.