Time Series Forecasting of Methane Emissions from Livestock using Machine Learning

MSc Research Project
Data Analytics

Shambhu Chandrakant Patole
Student ID: x19213743

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Shambhu Chandrakant Patole
Student ID:	x19213743
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Noel Cosgrave
Submission Due Date:	16/08/2021
Project Title:	Time Series Forecasting of Methane Emissions from Livestock using Machine Learning
Word Count:	4526
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	21st September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Time Series Forecasting of Methane Emissions from Livestock using Machine Learning

Shambhu Chandrakant Patole
x19213743

Abstract

Methane (CH4) is the infamous and strong greenhouse gas (GHG) in the environment. CH4 is almost 84 times stronger than carbon dioxide (CO2) and it has a global warming potential (GWP) of 25 over a 100-year time span. Methane has fewer concentrations and it stays in the atmosphere for almost 10 years. This study performed a comparative approach to forecast methane emissions from livestock (enteric fermentation), where the annual data from 1961-2019 was gathered from the Food and Agriculture of United Nations official website and it used three famous time-series models, ARIMA (Auto Regressive Integrated Moving Average), SVM (Support Vector Machine), and PROPHET with the goal to find the best model between these time series analysis algorithms. ARIMA performed the best among three models used that were evaluated using performance metrics that gave RMSE 0.03, MAPE 0.02, and MAE 0.05. Using this best model, methane emissions from livestock were also forecasted from 2020 to 2024. The results obtained in this study can help officials to concentrate more on methane emissions and bring new or change existing policies to mitigate climate change.

1 Introduction

1.1 Background

Animal agriculture is the leading cause of climate change, accounting for at least 87% of annual greenhouse gas (GHG) emissions, according to the recent findings of Global Sensitivity Analysis (GSA) done by Rao (2021). On every scale from local to global, the livestock sector appears as one of the top two or three most major contributors to the most serious environmental concerns, as stated by the Food and Agriculture Organization (FAO) of the United Nations (UN) (Steinfeld et al.; 2006). The conclusions of this paper indicate that it should be a significant policy emphasis when dealing with issues such as land degradation, climate change, and air pollution, water scarcity and pollution, and biodiversity loss. Now, the atmosphere of the Earth is made up of many gases and that contribute to keeping our planet warm. However, numerous greenhouse gases (GHGs) such as carbon dioxide (CO2), methane (CH4), nitrogen oxide (N2O), and chlorofluorocarbons have increased in recent decades as a result of diverse human activities, warming the globe and causing global warming and climate change. Being one of the most pressing issues of our times, it is crucial to address the major factors responsible for climate change. The two most important factors responsible for climate change are animal agriculture and the burning of fossil fuels. Additionally, it was found that the annual methane emissions from animal agriculture alone are responsible for incremental global warming than the combined annual carbon dioxide (CO2) emissions from all fossil fuels sources (Rao; 2021) proving it is essential to address methane (CH4) emissions from Livestock.

Billions of animals packed onto factory farms each year emit massive volumes of CH4. The gas is produced by ruminants like cows, sheep, and goats during their food digestion process known as Enteric Fermentation, and CH4 is also exhaled by the acres of cesspools topped up with the defecate of pigs, cows, and other agricultural animals. Methane is emitted through various sources including agriculture, energy, industry, and waste. However, the agriculture sector is the biggest emitter of anthropogenic methane emissions where enteric fermentation produces almost 65% of methane emissions as shown in FAO's website which is followed by emissions from rice production, other agricultural operations, and manure management (Karakurt et al.; 2012). Methane is a potent greenhouse gas, with a global warming potential (GWP) 34 times that of carbon dioxide over a 100-year period and 86 times that of carbon dioxide over a 20-year period (Stocker; 2014). Hence, forecasting methane emissions from enteric fermentation can further helpful officials monitor the highest emitter sector among livestock and make any required changes in the climate change/climate reduction policies.

1.2 Research Question

"How time series models can efficiently perform prediction of methane emissions from Animal Agriculture (Livestock)?"

1.3 Research Objectives

In terms of methane (CH4) emissions, the data from Food and Agriculture of the United Nations shows that Enteric Fermentation is the major contributor of CH4 emissions followed by rice cultivation and manure management, as shown in Figure 1. [1].

The objectives for this research are as follows,

1. To identify top emitter of methane (CH4) emissions produced due to enteric fermentation in the Livestock sector.

2. To compare the performance of the proposed time series models and find the best model for forecasting methane emissions.

3. To forecast the methane emissions for the future using the best-found time series model.

4. To publish the results that will further help changes in policies.

The research paper is structured as shown in the following format Section 2 that presents the literature review on the models, Section 3 addresses methodology approach in the field of methane emissions in livestock. Section 4 addresses with Implementation. In Section 5, Evaluation and results obtained from cleaning the data and results obtained from training all the models will be discussed. Section 6 justifies the results obtained from the trained model and insights are drawn and also the future work based on the topic is mentioned.

[1] http://www.fao.org/faostat/en/#data/GT/visualize

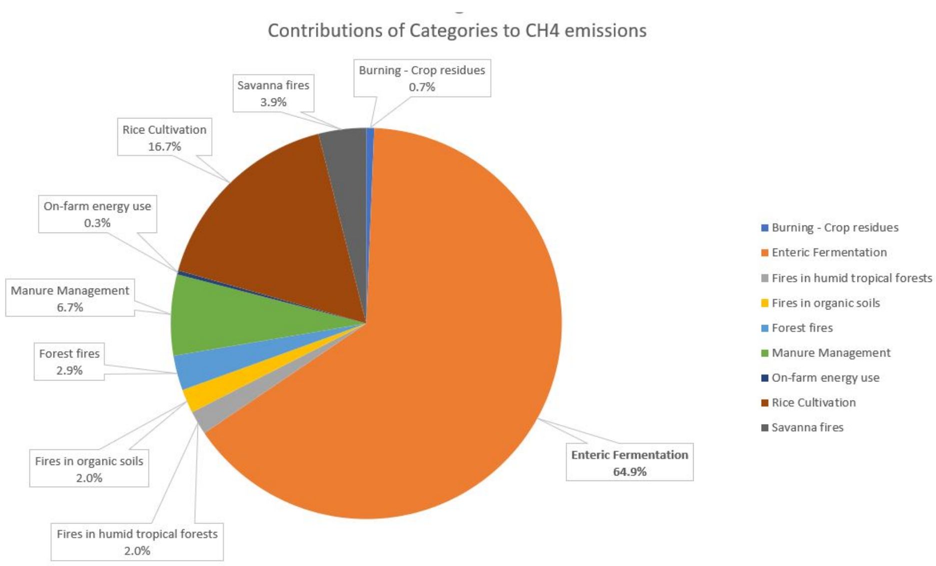


Figure 1: Contributions of categories to CH4 emissions

2 Related Work

2.1 ARIMA

The ARIMA model, in particular, has exhibited superior precision and accuracy in predicting the future lags of time series. New techniques to evaluate and forecast time series data have been developed as a result of recent advances in computer computational capacity (Siami-Namini et al.; 2018). Yusuf et al. (2014) created an ARIMA model to forecast CH4 emissions in 2014, with the purpose of calculating the amount of methane produced by livestock manure from 1980 to 2008, and then using that data to forecast future methane emissions by the sector in Malaysia from 2009 to 2020. The results of the ARIMA model, which was run through SPSS(v18) software, revealed that ARIMA(0,1,0) was the best forecasting model for cattle, buffaloes, and pigs, accounting for more than 95 percent of total methane emissions, with cattle emissions expected to rise from 67.0 Gg in 2009 to 77.0 Gg by 2020.

Hopali and Cakmak (2020) used the R Statistical Computing Software Package to anticipate energy-related CO2 emissions from manufacturing in the United States, using ARIMA, Holt Winters' Additive, and Holt-Winters' Multiplicative forecasting methods. There were recent studies done for forecasting Covid-19 cases published by Satrio et al. (2021), Talkhi et al. (2021). Former, used ARIMA and PROPHET models to forecast coronavirus disease in Indonesia, with the goal of comparing both models' ability to handle time-series data with no seasonality, random patterns, and low observations. Latter focused on comparing NNETAR, ARIMA, Hybrid, Holt-Winter, BSTS, TBATS, Prophet, MLP, and ELM network models with the main focus to find the best model among the 9 models used. ARIMA(1,0,0) and ARIMA(1,0,1) performed best in predicting the covid cases.

2.2 Prophet

Prophet Model was used to forecast air pollution based on time series Samal et al. (2019). The goal of this study was to determine the utility of analytics models in developing a system that can provide an approximate prediction of future pollution levels within a large confidence interval to predict future levels of several contaminants with a large confidence interval. Shitharth et al. (2021) used Prophet model for predicting covid-19 cases to anticipate 60 days ahead forecasts for committed cases, deaths, and recovered cases in India. The model was developed with a 95 percent prediction interval without modifying any seasonality parameters or additional regressor.

Chan (2020) used the ARIMA and PROPHET models to predict Myanmar stock prices for daily, weekly, and monthly 4 data, which was divided into two groups of training and testing, and the MAPE performance metric was used to measure the most accurate forecasting model by comparing the error analysis results, where the author emphasizes on using MAPE as it proves most useful. From 2005 to 2016, the Diurnal, Weekly, and Seasonal Cycles, as well as Annual Trends in Atmospheric CO2 and CH4, were studied at the Tower Network in Siberia by Belikov et al. (2019). This investigation proved that the Prophet model's ability to detect periodicity in environmental occurrences as well as revealing the seasonality of greenhouse gas concentrations was phenomenal.

2.3 SVM

Seasonal Autoregressive Integrated Moving Average (SARIMA) and Support Vector Machine (SVM) models were examined in the study conducted by Abdullah et al. (2021) on forecasting Rainfall in Bogor City, Indonesia. The best model with the lowest MAE, RMSE, MAPE, and r-Pearson was SVM with seasonal constraint and parameter cost = 4 and Gamma = 2. SVM was clearly shown to be a promising alternative strategy for predicting time series data in the study. Ballı (2021) conducted research on the COVID-19 pandemic between January 20, 2020, and September 18, 2020, using weekly data from the World Health Organization for the worldwide, United States, and Germany. Forecasting models included Support Vector Machines (SVM), Random Forest, Multi-layer perceptron, and Linear regression, which were evaluated using RMSE, APE, and 3 MAPE performance metrics. Following training and testing, it was discovered that the SVM model offered the lowest values for the metrics employed, as well as the best performance when compared to other models.

Bakay and Ağbulut (2021), a recently released study, focused on forecasting greenhouse gas emissions (GHG emissions) from the electricity generating sector in Turkey, with data from the Turkish Statistical Institute covering the years 1990 to 2018. Four years (2015-2018) were forecasted and assessed using five performance measures (RMSE, MBE, rRMSE, R2, and MAPE), and it was discovered that all of the algorithms individually offered satisfactory results for estimating GHG emissions in Turkey. Sun and Liu (2016) used SVM, ANN, and grey model (GM) algorithms in a similar study to estimate and analyze CO2 emissions from China's three major sectors and residential consumption. The algorithms were trained using industry data. The authors evaluated the algorithm's performance in terms of mean absolute percentage error (MAPE), root mean squared error (RMSE), maximum absolute percentage error (MaxAPE), and Median Absolute Percentage Error (MdAPE), with the SVM algorithm achieving the best results, with MAPE, MaxAPE, MdAPE, and RMSE of 0.328 percent, 0.641 percent, 0.306 percent, and 0.003 megatons, respectively.

3 Methodology

3.1 Introduction

KDD, CRISP-DM, and SEMMA are the popular processes for data mining. A comparative study done by Shafique and Qaiser (2014) shows that KDD has 9 steps, SEMMA has 5 steps whereas the CRISM-DM model has 6 steps, and also CRISP-DM is quite complete and accurate than the SEMMA model. Hence, for this research CRISP-DM is used. Figure 2 below shows the overall steps of CRISP-DM Methodology that are further explained in this section briefly.

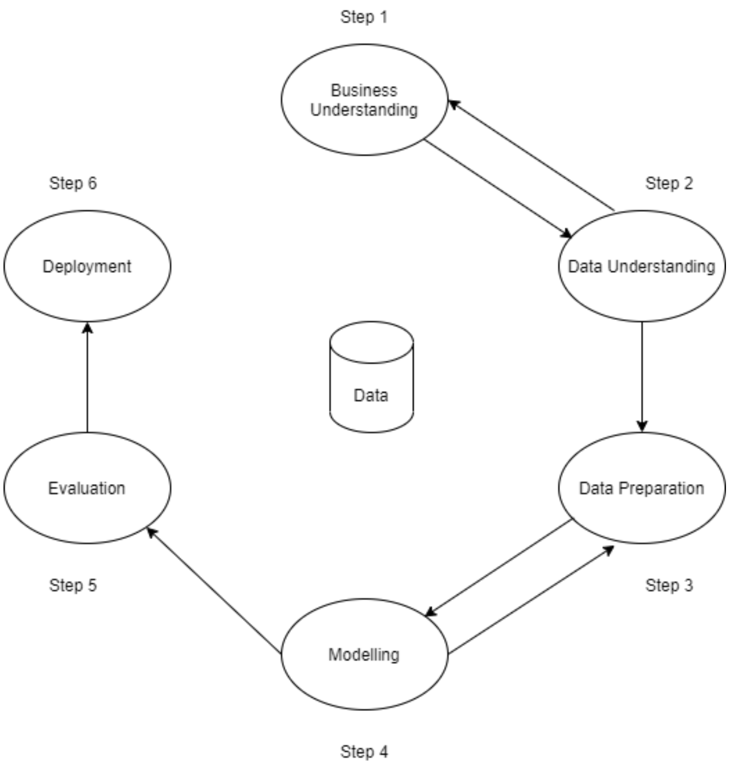


Figure 2: CRISP-DM Methodology

3.2 CRISP-DM Overview

3.2.1 Business Understanding

This is the first and most important phase of the project, in which the research objectives are established and highlighted, followed by the development of a business plan to attain those goals. Day by day, it is getting crucial to address climate change as it is severely impacting the environment and it is one of the most critical issues of our times. Apart from carbon dioxide (CO2), Methane is the notorious GHG and it is more potent as well. The data from the Food and Agriculture Organization of the United Nation shows that enteric fermentation, a digestive process in animals, contributes to almost 65% of methane emissions from agriculture followed by rice cultivation, manure management, and others. The research objective is to forecast Methane emissions from Livestock and find out the best model among the 3 machine learning models used in this research, namely ARIMA, PROPHET, and SVM.

3.2.2 Data Understanding

In this next step of CRISP-DM, the gathered data is collected and scrutinized to understand it and acquire insights related to that data. For this research, the data is downloaded from the Food and Agriculture Organization of the United Nations, which is publicly made available for research, statistics, and science. The data set contains methane emissions data from enteric fermentation by livestock species (asses, buffaloes, camels, cattle (dairy and non-dairy), goats, horses, llamas, mules, sheep, swine (breeding and market)) having yearly data from 1961 to 2019 with projections for 2030 and 2050 that contains data by country, regions and special groups, with global coverage. The data set contains 369580 rows including NA values having 14 columns (Area Code, Area, Item Code, Item, Element Code, Element, Year Code, Year, Source Code, Source, Unit, Value, Flag, Note).

3.2.3 Data Preparation

In this section, data cleaning performed during this research will be discussed. For the cleaning of the data, RStudio software was used. Original data was taken into a data frame and further data cleaning operations were performed. Out of 14 columns, only 7 columns were selected. Rows for 2030 and 2050 projections were removed. Next, Source column and Element column were filtered to include only 'FAO Tier 1' and 'Emissions (CH4)' values, respectively. Further, the Item column was filtered to include the data for 'All Animals' and Area column to include data only for 'World' values. Finally, the data frame was exported to a comma-separated values (csv) file, which was further used for modeling.

3.2.4 Modeling

In this section, the cleaned data is further given as an input to the various models that will be used in this research. The models used are as follows, ARIMA (AutoRegressive Integrated Moving Average), Support Vector Machine(SVM), and PROPHET. Before inputting the cleaned data to the models, it was divided into training and testing sets. The models were trained on data from 1961 to 2014 and the forecasting was tested on the remaining test data i.e. 2015 to 2019. Below are the details about the models used in this research,

i. ARIMA - When it comes to forecasting, AutoRegressive Integrated Moving Average (ARIMA) is one of the popular time series models. However, the requirement before running the ARIMA model is that the time series data used should be stationary. Further, based on the previous time series values from the past, this model predicts the values for the future, where mainly 3 parameters p, d, and q are inputted to this ARIMA model. p gives the number of lag observations, d gives the degree of differencing and q gives moving average size.

ii. PROPHET - This is a time-series data forecasting algorithm based on a contribution model that works well for non-linear trends data with yearly, weekly, and daily seasonality, as well as holiday impacts. Because it can handle significant seasonal and historical data, the Prophet process is one of the most accurate, rapid, and effective time series forecasting models and it has a high sensitivity to missing data, capturing trend shifts and huge outliers. [2] It is an open-source package available in both Python and R languages.

iii. SVM - Because of its simplicity and the fact that it addresses some of the flaws of ANN (Artificial Neural Network), such as overfitting, SVM (Support Vector Machine), a supervised machine learning algorithm, has become a popular algorithm for both classification and regression issues (Tang and Zhou; 2015). This algorithm uses statistical learning theory to guide its learning process.

3.2.5 Evaluation

In this fifth section of CRISP-DM methodology, evaluation of the performance of all models used is done by metrics RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error), and MAE (Mean Absolute Error). Further, the comparison of all the models is done using these metrics. Whichever model among these 3 models gives the lowest RMSE, MAPE, and MAE values will be considered as the best model and can be considered as better at forecasting than others.

3.2.6 Deployment

In this last section of CRISP-DM, the final deployment is done by preparing a presentation and a thorough research document along with a configuration manual.

4 Implementation

4.1 Introduction

This section discusses the complete flow and the implementation of the models used to meet the research project's objectives. Figure 3 shown below depicts the flow diagram of the implementation followed throughout the research.

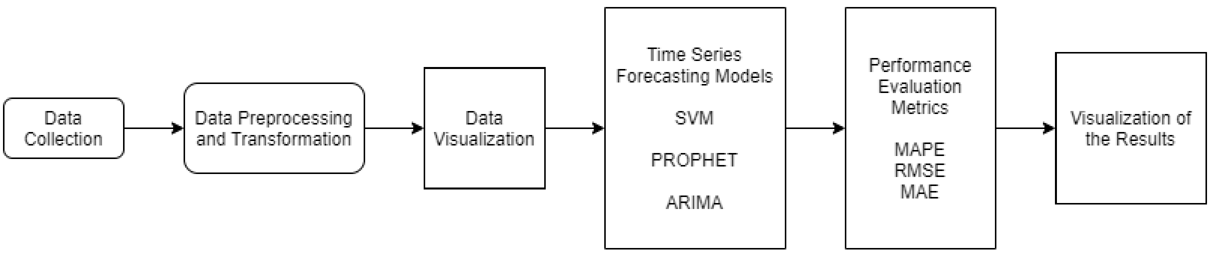


Figure 3: Flow diagram of the Implementation

[2] https://facebook.github.io/prophet/

4.2 Data Collection

The data set used for this research is downloaded from the Food and Agriculture Organization of the United Nations official website. The collected data is thoroughly scrutinized to understand it and find insights related to it. For the purpose of this research, the data was downloaded from the Food and Agriculture Organization of the United Nations website [3], that is publicly available for science, statistics, and research purposes. The data consists of comma-separated values (CSV) that contain annual methane (CH4) emissions from enteric fermentation by livestock species (asses, buffaloes, camels, cattle (dairy and non-dairy), goats, horses, llamas, mules, sheep, swine (breeding and market)) for period 1961-2019 with projections for 2030 and 2050 having data by country, regions, special groups and global coverage. The data set contains 369580 rows including NA values having 14 columns (Area Code, Area, Item Code, Item, Element Code, Element, Year Code, Year, Source Code, Source, Unit, Value, Flag, Note). The sample rows of the data set can be seen in Figure 4 shown below.

Area Code	Area	Item Code	Item	Element Code	Element	Year Code	Year	Source Code	Source	Unit	Value	Flag	Note
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1961	1961	3050	FAO TIER 1	kilotonnes	240.6831	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1962	1962	3050	FAO TIER 1	kilotonnes	245.3106	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1963	1963	3050	FAO TIER 1	kilotonnes	255.8285	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1964	1964	3050	FAO TIER 1	kilotonnes	259.065	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1965	1965	3050	FAO TIER 1	kilotonnes	265.598	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1966	1966	3050	FAO TIER 1	kilotonnes	276.994	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1967	1967	3050	FAO TIER 1	kilotonnes	280.094	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1968	1968	3050	FAO TIER 1	kilotonnes	288.821	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1969	1969	3050	FAO TIER 1	kilotonnes	286.382	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1970	1970	3050	FAO TIER 1	kilotonnes	290.26	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1971	1971	3050	FAO TIER 1	kilotonnes	287.79	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1972	1972	3050	FAO TIER 1	kilotonnes	231.527	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1973	1973	3050	FAO TIER 1	kilotonnes	244.979	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1974	1974	3050	FAO TIER 1	kilotonnes	262.836	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1975	1975	3050	FAO TIER 1	kilotonnes	282.074	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1976	1976	3050	FAO TIER 1	kilotonnes	288.225	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1977	1977	3050	FAO TIER 1	kilotonnes	280.876	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1978	1978	3050	FAO TIER 1	kilotonnes	280.285	A	
2	Afghanistan	1755	All Anima	7225	Emissions (CH4)	1979	1979	3050	FAO TIER 1	kilotonnes	274.235	A	

Figure 4: Raw Data gathered from UN's FAO Official Website

4.3 Data Pre-processing

Before inputting the actual data to the model and for it to run efficiently it is important to clean the raw data to remove unwanted information like missing values (NA) or any special characters like '@' or '$', etc. To do so, RStudio software was used to perform the data cleaning required for this research. Below are the detailed steps followed for pre-processing the data,

1. Initially, out of the 14 columns in the data set, only 7 relevant columns were selected. Rest other columns were excluded.

2. Rows for 2030 and 2050 projections were removed.

3. Source column was filtered to include only 'FAO Tier 1' data and 'UNFCCC' was excluded as it did not contain data for all years and had some missing values.

[3] http://www.fao.org/faostat/en/#data/GE

4. Element column was filtered to include only 'Emissions (CH4)' values excluding the rows for 'Stocks' that was not required for this research.

5. Item column was filtered to include the data for 'All Animals' and Area column to include data only for 'World' values.

6. Finally, the data frame was exported to a comma-separated values (csv) file and this cleaned data was further used for modeling.

Similar changes were also performed separately using Microsoft Excel without making any changes to the original dataset. To gain more insights, data was analyzed in excel for finding out the top emitter of methane emissions in Livestock. The results of the analysis shown in Figure 5 clearly show that cattle is the highest contributor to methane emissions from Livestock. The first objective of this research paper is achieved here.

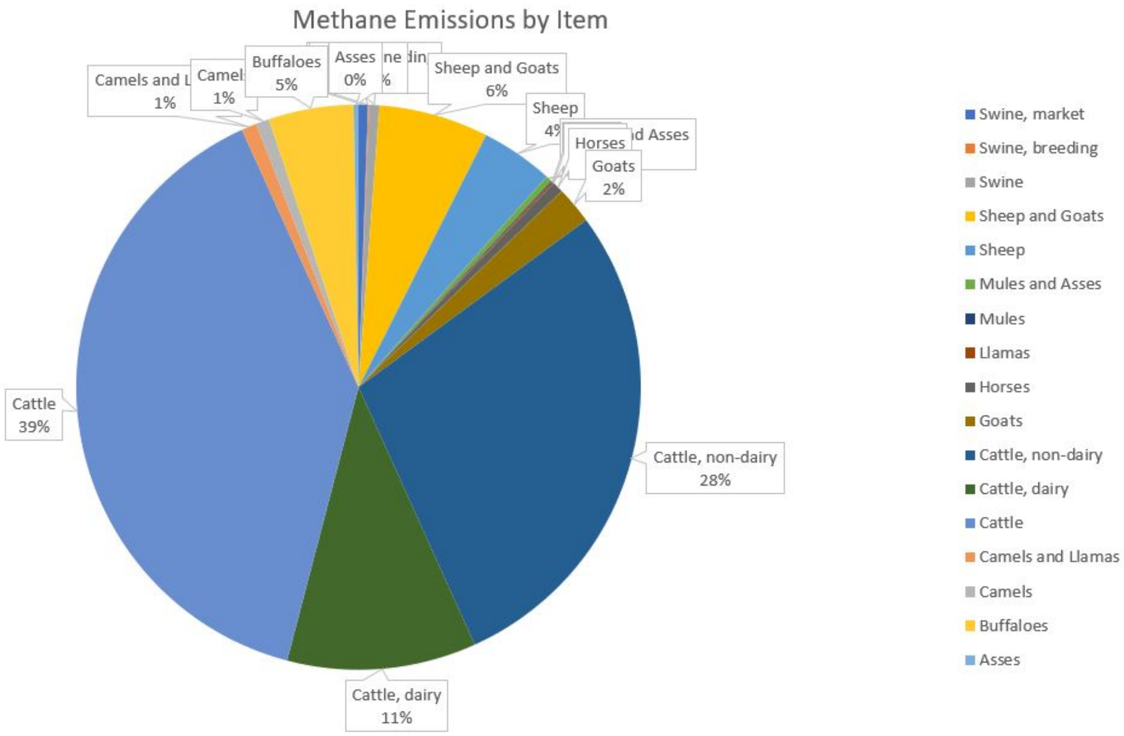
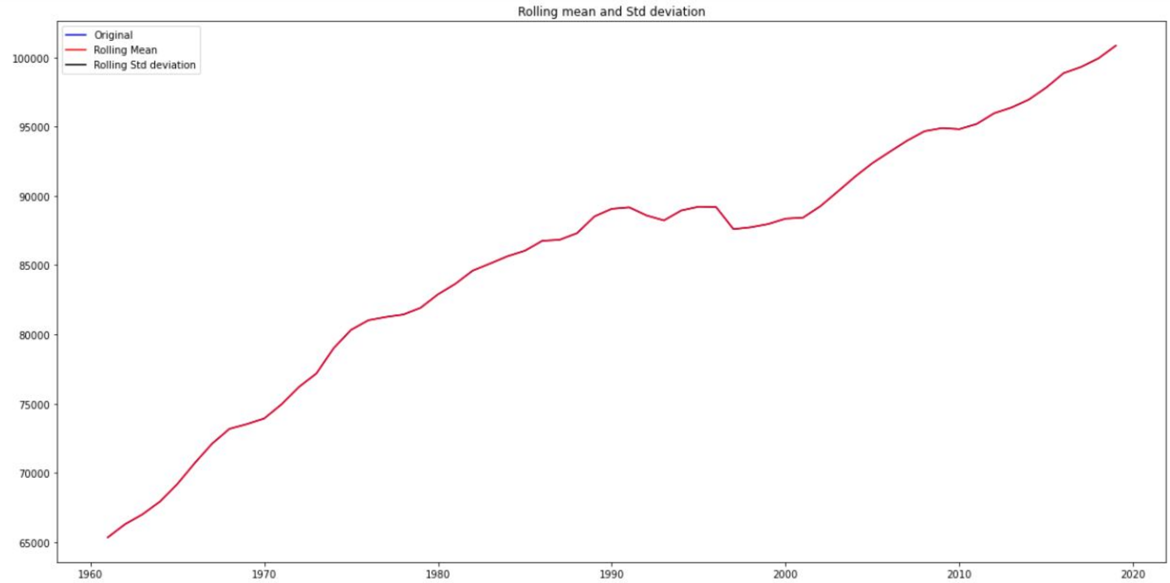


Figure 5: Methane emissions by Item

4.4 Model Construction

4.4.1 ARIMA

Initially, the required libraries are loaded into Jupyter Notebook and the cleaned data is loaded into a data frame by parsing the year column to convert it into a DateTime format. Columns Area, Item, Element, Source, and Unit are removed and only Year and Values columns are further used for input to the model. Major libraries used for constructing/building ARIMA model are pandas, NumPy, matplotlib, sklearn, statsmodel.tsa. At the start, it is checked if the time series data is stationary or not. Figure 6 shows the results after performing the Dickey-Fuller test that shows there is a trend in time series which means the time series data is not stationary.




```
Dickey fuller has been performed and the results are:
Test-Statistic          -1.531466
p-value                  0.517892
Lags                     1.000000
Observations Used       57.000000
```

Figure 6: Original data having trend

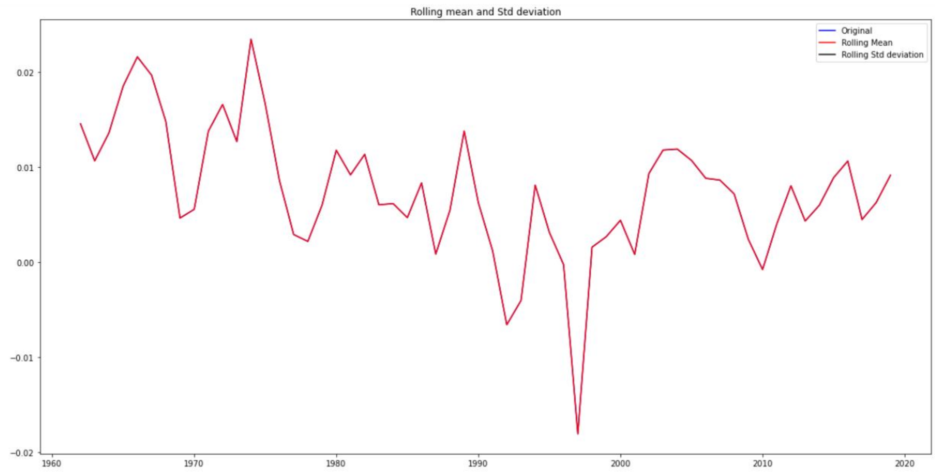
A separate test stationarity function is created that determines whether the given data set is stationary or not. Rolling mean and Standard deviation is plotted and then the Dickey-Fuller test is performed using 'adfuller' function on the time series data. A series of output is generated having Test statistics, p-value, lags, and observations used. As the time series data used was not stationary, the p-value obtained was higher than 0.05. Hence, 1-order differencing (value of d) was performed and again the Dickey-Fuller test was performed that gave a p-value less than 0.05 and the time-series was stationary that can ben seen from the output shown in Figure 7.

Figure 8 and Figure 9 shows the ACF and PACF plot respectively, that further help choose the parameter that needs to be passed to ARIMA model. Figure 10 shows the graph after the model is trained.fitting vs original for the 1 order differencing

Further, it was passed to check Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to get the values of p and q respectively. Now, the data is divided into train and test, and using the values p,d, and q, the ARIMA model is trained against train data and .fit() function is used for fitting the model. Once the model is trained, function .forecast() is used to predict the values for the test data. Both values from the dataset and predictions are checked using the evaluation metrics RMSE, MAPE, and MAE.

4.4.2 SVM

Support Vector Regressions (SVR) All the required libraries are loaded including the SVR library from sklearn package. At the start, the cleaned data is loaded into a data




```
Dickey fuller has been performed and the results are:
Test-Statistic          -3.704697
p-value                  0.004048
Lags                     0.000000
```

Figure 7: One order of differencing

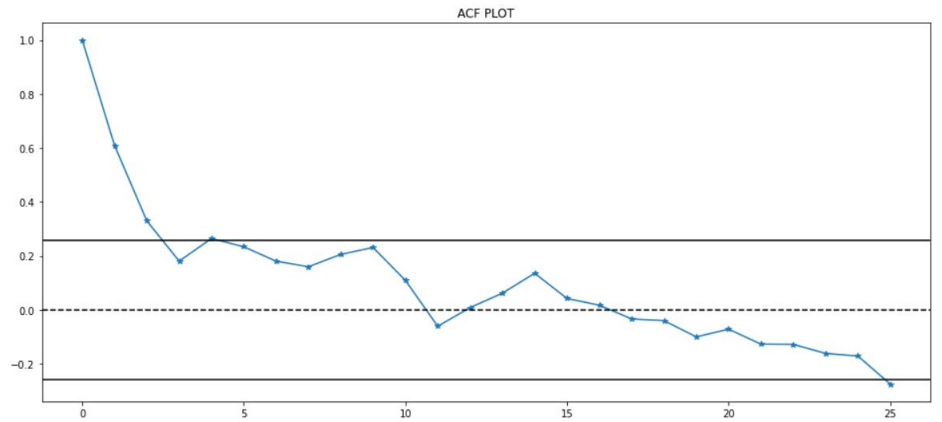


Figure 8: ACF PLOT

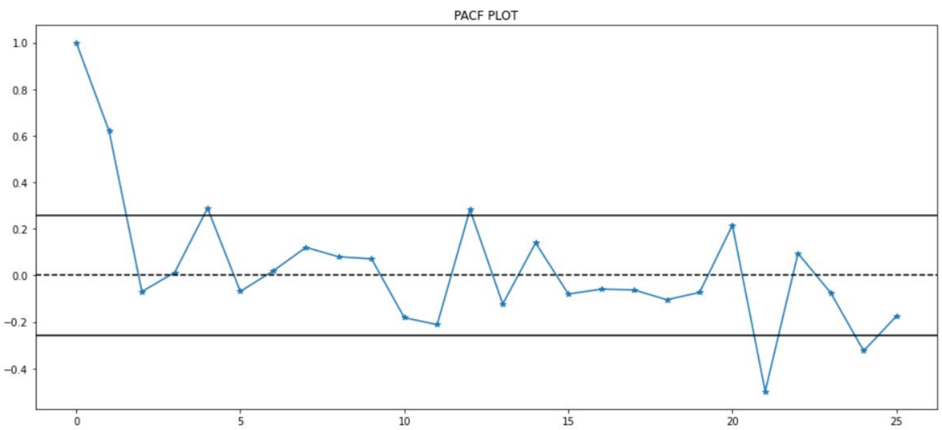


Figure 9: PACF PLOT

Figure 10: Original vs Fitting

frame in Google Colab where the year column is parsed using DateTime functions to convert it into a DateTime format. Columns Area, Item, Element, Source, and Unit are deleted, keeping only Year and Values columns that are further given as an input to the SVR model. Year column (used as an index) and the Values column are split into two different sets of train and test (the former contains the index value and the latter contains the methane emissions value in kilotonnes). One important input to SVR is the type of kernel to be used. For this research, 'rbf' is used. Initially, best kernel is chosen for the data set being used (among the kernels, 'linear','poly','rbf','sigmoid') and then the parameter tuning is performed to get better accuracy. The model is then trained using 'rbf' kernel and fitted using the function .fit(). After the training, the .predict() function is applied and the SVR prediction values are obtained for the defined test index. Finally, the original and predicted values are tested using the RMSE, MAPE, and MAE evaluation metrics.

4.4.3 PROPHET

The third model chosen for this research is PROPHET developed by Facebook which has proven efficient in forecasting time series data. Google Colab is used for this model construction. In the beginning, the cleaned data is loaded into a data frame. Year column is parsed using DateTime function to convert it into a DateTime format. Columns Area, Item, Element, Source, and Unit are excluded, taking only Year and Values columns that are further renamed as 'ds' and 'y', respectively, as it is the requirement of the prophet model. Data is divided into training and testing. As the data used for this research is yearly data, the frequency is given as 'YS' that means for yearly data. For occupying future dates, a new data frame is created using the .make_future_dataframe function. The prophet model is fitted using .fit() function for training data and once it is trained the prediction values are forecasted. Test values and prediction values are tested using RMSE, MAPE, and MAE metrics and finally original and predicted values are plotted.

4.5 Conclusion

To conclude, the Implementation stage provides a detailed brief about the complete flow of the implementation including cleaning of data, diving data into train and test, building the models using libraries defined in python for forecasting the time series data. For model construction purposes, various articles from online were referred. For ARIMA, [4] For prophet, [5] and for SVM (Namboori; 2020) were referred.

5 Evaluation and Results

5.1 Introduction

In this segment, the detailed evaluation of all the models implemented will be checked. RMSE. MAPE and MAE are the performance metrics that will be used to check the performance of all models in terms of forecasting the methane emissions from livestock.

RMSE:

Root Mean Squared Error (RMSE) is a performance metric that can help determine the standard deviation of errors. The lower the RMSE value, the better the model's performance.

MAE:

Mean Absolute Error (MAE) is a performance metric that gives the absolute average by summing the difference between original and forecast values and dividing it by all the number of rows.

MAPE:

Mean Absolute Percentage Error (MAPE) is a performance metric that helps in determining the error in the predicted values that shows the accuracy of the model after forecasting. The lower the MAPE value, the more accurate the model is, which demonstrates the model has a superior prediction/forecasting ability.

5.2 Experiment 1 - Evaluation of ARIMA Model

In this experiment, pre-processed data was given as an input to the ARIMA model to train it with order (3,1, 2) and the later model was fitted. Once the ARIMA model was trained and fitted, a forecast was done on test data that gave below metrics values for RMSE, MAE, and MAPE as shown in Table 4.

Further, matplotlib library was used to compare the original and the forecast values as shown in Figure 11,

[4] https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/
[5] https://machinelearningmastery.com/time-series-forecasting-with-prophet-in-python/

Table 1: Performance Metrics

Metrics	ARIMA
RMSE	0.03
MAPE	0.02
MAE	0.05

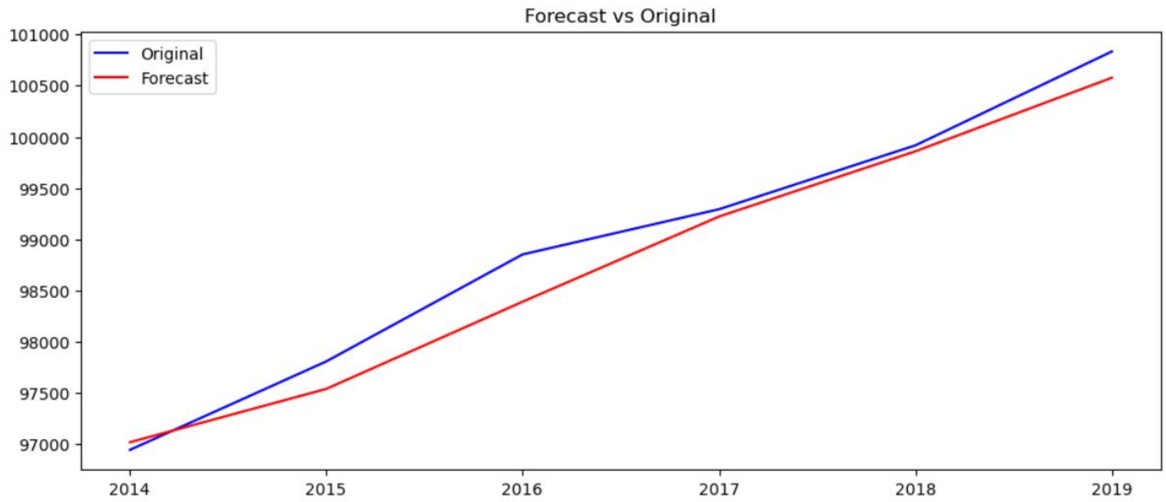


Figure 11: ARIMA Orignal vs Forecast

5.3 Evaluation of PROPHET Model

In this experiment, pre-processed data was given as an input to the PROPHET model to train it and fitted. Once the model was done with training, a new data frame was created with yearly frequency and for periods the same as the length of the test data set. The forecast was done on test data that gave below metrics values for RMSE, MAE and MAPE as shown in Table 2.

Table 2: Performance Metrics

Metrics	PROPHET
RMSE	13.43
MAPE	0.12
MAE	121.45

Figure 12 shows the original values against the forecast values.

5.4 Evaluation of SVR Model

Before training the SVR model, first, the best kernel was chosen for the data set used in this research. As 'rbf' In this experiment, pre-processed data was given as an input to the SVR model to train it and it was fitted. Once the model was done with training, a new data frame was created with yearly frequency and for periods the same as the length

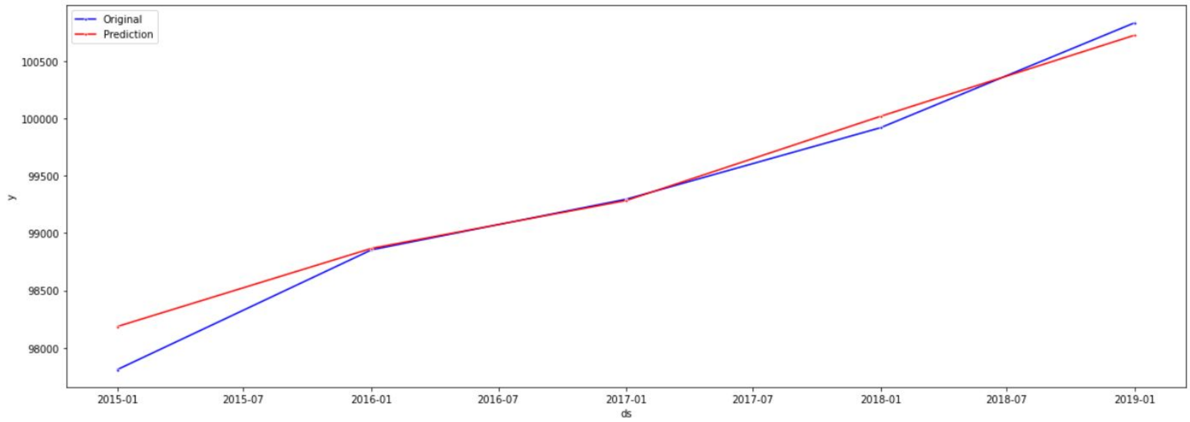


Figure 12: Prophet Original vs Forecast

of the test data set. The forecast was done on test data that gave below metrics values for RMSE, MAE, and MAPE as shown in Table 3.

Table 3: Performance Metrics

Metrics	**SVR**
RMSE	0.19
MAPE	1.72
MAE	0.44

5.5 Discussion

After evaluating all the three models used above, it was found that the ARIMA model worked best than the other models with better accuracy giving RMSE, MAE, and MAPE metrics. SVM gave a good RMSE value compared to PROPHET but couldn't surpass the results from the ARIMA model. PROPHET model showed poor performance compared to the other two models in terms of performance metrics used.

Table 4: Performance Metrics

Metrics	**ARIMA**	**SVM**	**PROPHET**
RMSE	0.03	0.19	13.43
MAPE	0.02	1.72	0.12
MAE	0.05	0.44	121.45

5.6 Results

The results of all the models used were evaluated using the performance metrics RMSE, MAE and MAPE performance metrics shown in Figure **??**. The results obtained indicate that ARIMA performed best in forecasting methane emissions from livestock, all the values for the evalutaion metrics were lowest for ARIMA as compared to other models.

6 Conclusion and Future Work

The purpose of this research was to find the best forecasting model for methane emissions from Livestock after which the published results can help make new or change existing policies related to mitigating climate change. To work on this research, the data was taken from Food, Agriculture and Organization of United Nations website that has data for methane emissions from enteric fermentation for the period 1961-2019. Exploratory data analysis was performed that revealed Cattle are the highest emitter of methane emissions from livestock. Further, the forecasting was done for world data, filtering the data for all the animals combined using the 3 machine learning models ARIMA, SVM, and PROPHET. To evaluate the performance of the models, RMSE, MAE and MAPE performance metrics were used. After training and testing each model, it was found that the ARIMA model worked best among them. ARIMA model was used to predict methane emissions in kilotonnes from the year 2020 to 2024 as seen in Figure 13. Finding the best model and predicting future methane emissions objective is achieved here.

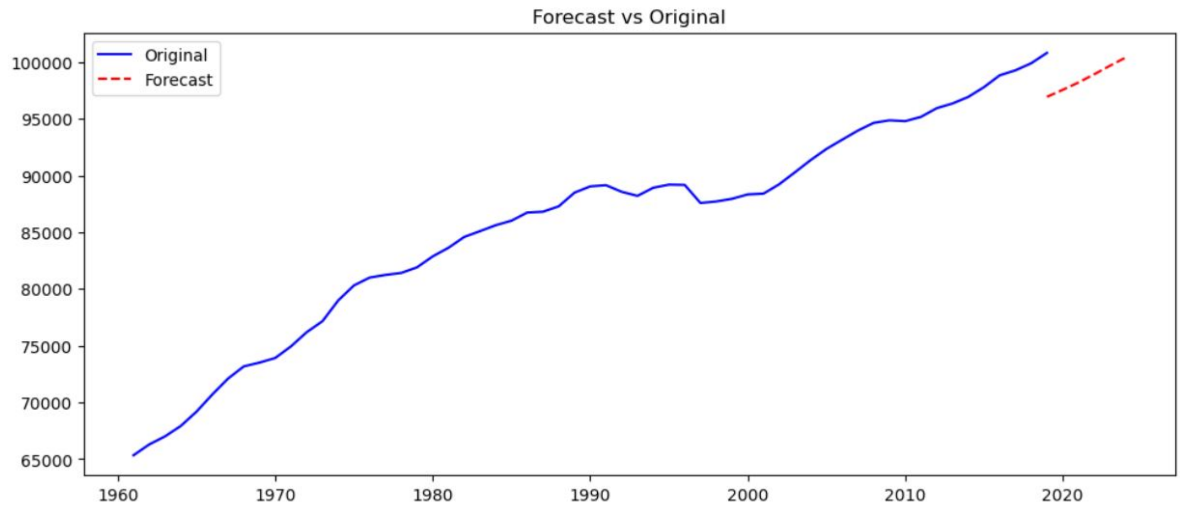


Figure 13: ARIMA Model forecasting from 2020 to 2024

For future work, different other models like LSTM can be used to compare the performance. However more historical and especially weekly or monthly data would be required.

7 Acknowledgement

It gives me immense pleasure to thank my research supervisor Noel Cosgrave for all his efforts and valuable inputs and feedback. I would also like to thank my family and friends for their online support.

References

Abdullah, A., Ruchjana, B., Jaya, I. et al. (2021). Comparison of sarima and svm model for rainfall forecasting in bogor city, indonesia, *Journal of Physics: Conference Series*, Vol. 1722, IOP Publishing, p. 012061.

Bakay, M. S. and Ağbulut, Ü. (2021). Electricity production based forecasting of greenhouse gas emissions in turkey with deep learning, support vector machine and artificial neural network algorithms, *Journal of Cleaner Production* **285**: 125324.

Ballı, S. (2021). Data analysis of covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods, *Chaos, Solitons & Fractals* **142**: 110512.

Belikov, D., Arshinov, M., Belan, B., Davydov, D., Fofonov, A., Sasakawa, M. and Machida, T. (2019). Analysis of the diurnal, weekly, and seasonal cycles and annual trends in atmospheric co2 and ch4 at tower network in siberia from 2005 to 2016, *Atmosphere* **10**(11): 689.

Chan, W. N. (2020). Time series data mining: Comparative study of arima and prophet methods for forecasting closing prices of myanmar stock exchange, *J. Comput. Appl. Res.* **1**: 75–80.

Hopali, E. and Cakmak, A. (2020). Prediction of daily co2 emissions of a factory using arima and holt-winters seasonal methods, *Business and Management* **12**(3).

Karakurt, I., Aydin, G. and Aydiner, K. (2012). Sources and mitigation of methane emissions by sectors: A critical review, *Renewable energy* **39**(1): 40–48.

Namboori, S. (2020). *Forecasting Carbon Dioxide Emissions in the United States using Machine Learning*, PhD thesis, Dublin, National College of Ireland.

Rao, S. (2021). Animal agriculture is the leading cause of climate change.
URL: *https://climatehealers.org/the-science/animal-agriculture-position-paper/*

Samal, K. K. R., Babu, K. S., Das, S. K. and Acharaya, A. (2019). Time series based air pollution forecasting using sarima and prophet model, *proceedings of the 2019 international conference on information technology and computer communications*, pp. 80–85.

Satrio, C. B. A., Darmawan, W., Nadia, B. U. and Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in indonesia using arima model and prophet, *Procedia Computer Science* **179**: 524–532.

Shafique, U. and Qaiser, H. (2014). A comparative study of data mining process models (kdd, crisp-dm and semma), *International Journal of Innovation and Scientific Research* **12**(1): 217–222.

Shitharth, S., Mohammad, G. B., Ramana, K. and Bhaskar, V. (2021). Prediction of covid-19 wide spread in india using time series forecasting techniques.

Siami-Namini, S., Tavakoli, N. and Namin, A. S. (2018). A comparison of arima and lstm in forecasting time series, *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, pp. 1394–1401.

Steinfeld, H., Gerber, P., Wassenaar, T. D., Castel, V., Rosales, M., Rosales, M. and de Haan, C. (2006). *Livestock's long shadow: environmental issues and options*, Food & Agriculture Org.

Stocker, T. (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*, Cambridge university press.

Sun, W. and Liu, M. (2016). Prediction and analysis of the three major industries and residential consumption co2 emissions based on least squares support vector machine in china, *Journal of Cleaner Production* **122**: 144–153.
URL: *https://www.sciencedirect.com/science/article/pii/S0959652616002237*

Talkhi, N., Fatemi, N. A., Ataei, Z. and Nooghabi, M. J. (2021). Modeling and forecasting number of confirmed and death caused covid-19 in iran: A comparison of time series forecasting methods, *Biomedical Signal Processing and Control* **66**: 102494.

Tang, Y. and Zhou, J. (2015). The performance of pso-svm in inflation forecasting, *2015 12th International Conference on Service Systems and Service Management (ICSSSM)*, IEEE, pp. 1–4.

Yusuf, R. O., Noor, Z. Z., Abba, A. H., Hassan, M. A. A., Majid, M. R. and Medugu, N. I. (2014). Predicting methane emissions from livestock in malaysia using the arima model, *Management of Environmental Quality: An International Journal* .