Symbiosis Institute of Computer Studies & Research

Meat Consumption Prediction Using Weka

Name: Ayush Warrier

PRN: 23030124067

Guided by: Dr. Amol Vibhute

Abstract: So understanding food consumption trends is important, Why? Because it helps analyze dietary habits and also improve food production planning. In India, meat consumption has been increasing, which made me think that I look up the past trends so that I can predict how much meat consumption can happen in the future. My aim for this project, is to apply Machine Learning techniques to predict India's future meat consumption and maybe also provide insights into how demand might change in the coming years.

To help me with this study, I used a dataset that I found from Kaggle that contains historical meat consumption data from 1961 to 2021. I first started by filtering the dataset to keep only India's data because the dataset I got from Kaggle was too big since it contained meat consumption of multiple countries over multiple years. Then I created a new column called Total Meat Consumption, which adds up all the meat categories, including Poultry, Beef, Sheep & Goat, Pork, Other Meats, and Fish & Seafood. This would help me and make my task easier in predicting and allowed me to focus on overall consumption trends.

For predictions, I decided to apply Linear Regression, a widely used Machine Learning model for trend forecasting. I trained the model using 61 years of past data and then used it to predict meat consumption for the next 10 years (2022-2032). So when I evaluated the model, I found that it performed really well, and it achieved a R^2 Score of 0.9985, a Mean Absolute Error (MAE) of 0.55, and a Root Mean Squared Error (RMSE) of 0.75. These results confirmed that my model was highly accurate in forecasting future consumption trends.

The predictions from this project could be useful for food industry planners, policymakers, and researchers in making informed decisions about future food demand and supply chain management. Through this project, I was able to see how Machine Learning can be applied to real-world forecasting problems, particularly in the areas of agriculture and food consumption analysis.

Keywords:

- Machine Learning
- Python
- Meat Consumption Prediction
- Linear Regression
- Data-Analysis

1. Introduction:

Predicting meat demand is quite important for long-term resource management, coming up with agricultural policies, and planning for the food sector and environmental protection. There is a growing demand for meat so I wanted to analyze how meat consumption has changed over time so that businesses and governments can make better decisions about food productions. By predicting future meat consumption in India, companies can improve the meat production and distribution, and at the same time policymakers can plan for the food

security and trade policies.

Years ago, food consumption was analyzed using the basic trend analysis and averages. So these methods provided insights into past trends, I realized that they weren't reliable for future predictions since they didn't account for factors like population growth and economic changes. I also saw the time-series models like ARIMA (Auto-Regressive Integrated Moving Average), which have been used for food demand forecasting. However, I noticed that these models require a lot of manual adjustments and don't work well when consumption patterns fluctuate significantly [(Saudi Arabia Prediction, 2024)]. This made me question whether ARIMA would be the best approach for my project.

As I explored more advanced forecasting methods, I found that Artificial Intelligence (AI) and Machine Learning (ML) have been widely used to improve prediction accuracy. A common forecasting method is Linear Regression, which establishes a relationship between an independent variable (such as year) and a dependent variable (such as total meat consumption). I decided to go with Linear Regression because it is simple to understand and can reasonably predict meat consumption based on past data [(Türkiye Study, 2023)]. It also provides valuable insights into how population and income levels influence food demand.

Although I chose Linear Regression, I also looked into Deep Learning models and Neural Networks, which have been used for food demand forecasting in large-scale studies. These models can detect hidden patterns in data and have been successful in predicting food consumption trends [(Kazakhstan Study, 2024)]. However, I found that Neural Networks have limitations—they require large datasets and high computational power, and their decision-making process is often unclear, making them difficult to interpret. I realized that for my project, I needed a model that was easy to implement and interpret, which is why I ultimately chose Linear Regression [(Jayapal, 2022)].

I also noticed that most studies on food consumption trends focus on Western countries, and very little research has been done on India's meat consumption patterns [(Switzerland Study, 2024)]. Since India's meat consumption has been rising, I wanted to develop a model that forecasts future trends based on historical data. Many modern machine learning techniques can be computationally expensive or unnecessarily complex, which is why I felt that Linear Regression struck the right balance between simplicity and accuracy.

For this project, I decided to use Kaggle's meat consumption dataset to develop a Linear Regression model that forecasts India's meat consumption trends for the next ten years (2022–2032). So what my goal was to analyze the past meat consumption data that I got from Kaggle and evaluate the model's performance using metrices like R^2 Score, MAE, and RMSE. I also wish to visualize the findings, making it easier for policymakers and industry professionals to understand future meat demand.

This report covers the dataset used, the literature review, methods and techniques applied, results, and conclusions to present a complete analysis of India's future meat consumption trends.

2. Literature Review:

I explored a variety of forecasting techniques, and then I realized that predicting food consumption has always been an important part for agricultural planning, food economics, and supply chain management. Over the years, researchers have tried various methods, from traditional statistical models to more advanced machine learning algorithms, to make food demand predictions more accurate.

One of the earliest and most commonly used methods is ARIMA (Auto-Regressive Integrated Moving Average), which relies on historical data to make predictions. While ARIMA has worked well for short-term forecasting, I found that it struggles when economic conditions change or consumer preferences shift. A study on red meat consumption in Saudi Arabia confirmed this issue, as ARIMA failed to capture economic changes, leading to inaccurate long-term forecasts [(Saudi Arabia Prediction, 2024)]. This made me realize that I needed a model that could adapt better to changing food demand trends without requiring constant manual adjustments.

Since Linear Regression is a simpler and widely used forecasting method, I looked into how well it performs in food demand predictions. A study conducted in Türkiye applied Linear Regression to predict meat consumption trends, and while the results were easy to interpret, the model struggled with non-linear patterns and external factors like climate change and government policies [(Türkiye Study, 2023)]. Despite its limitations, I found that Linear Regression is still a great choice when the trend is steady and increasing over time, which fits the pattern seen in India's meat consumption trends.

Another issue I came across was that most food consumption studies focus on Western countries, leaving very little research on India's meat consumption trends. A study on meat consumption in Poland found that eating habits vary based on age and economic conditions, which highlights the importance of regional studies [(Switzerland Study, 2024)]. Since India's dietary preferences and meat consumption trends are changing, I wanted to work on filling this research gap by focusing specifically on India's meat consumption trends.

I also looked into other machine learning techniques, like Neural Networks and Deep Learning Models, which have been used in food consumption forecasting. A study in Kazakhstan found that Neural Networks performed significantly better than traditional methods [(Kazakhstan Study, 2024)]. However, I noticed that these models have several downsides—they require large datasets, high computational power, and act like "black-box" systems, making it difficult to understand how they make predictions. Since I wanted a simple, interpretable model that doesn't require high-end computational resources, I decided to go ahead with Linear Regression [(Jayapal, 2022)].

My aim by choosing Linear Regression is to provide an effective solution for predicting India's meat consumption in the future. Why? Because this method is accurate, simple and straightforward, making it useful for people who need easy-to-interpret forecasts. Through this project, I hope to contribute to India-specific research on food consumption trends, helping businesses and policymakers make informed decisions for the future.

3. Data Used in the Project:

The dataset I used in this project was taken from Kaggle and contains information on meat consumption per capita for multiple countries for years 1961 to 2021. Since my report focuses only on India's meat consumption, I selected only India's data and made a different file for it. This makes sure that the predictions I get in this project are specific to India's food consumption patterns.

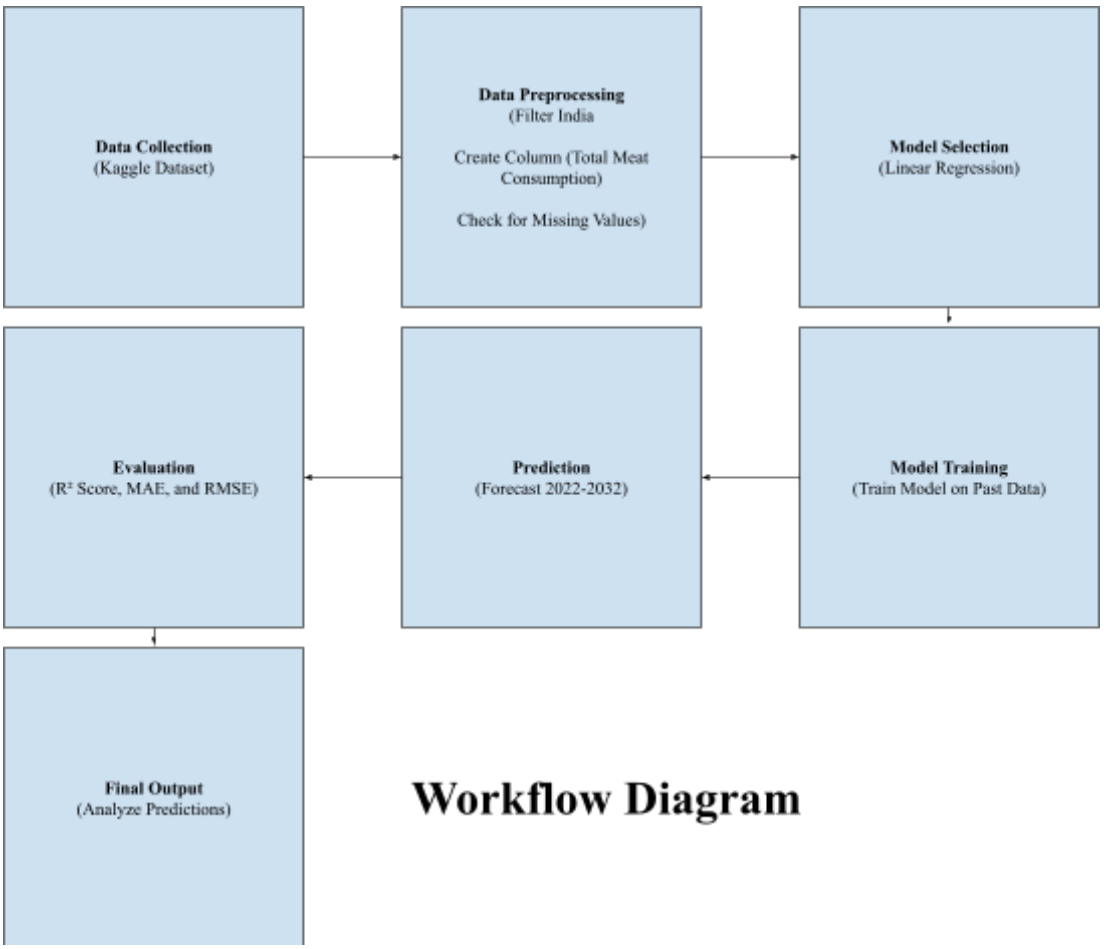
The columns in the dataset are:

- Year
- Poultry
- Beef
- Sheep and Goat
- Pork
- Other Meats
- Fish and Seafood
- Total Meat Consumption – I added a new column by adding all meat categories for each year.

In the end, I just converted the dataset into a CSV file to make it easier for me to use it in Python. The reason why I added the Total Meat Consumption column was to help analyze trends and make predictions. I checked the dataset and found that there were no missing values, so no extra steps were needed for data cleaning.

I purposely took this dataset because this dataset is small, which makes it perfect for using Linear Regression. Since the data covers from 1961 to 2021, it provides a good historical reference for forecasting future meat consumption trends in India.

4. Methods & Techniques Used:



Workflow Diagram

4.1 Data Preprocessing:

Before I start training the model, the dataset needed to be cleaned and prepared to make sure that I get accurate predictions. So for that I applied the following preprocessing steps:

i) Selecting India's Data: So, the original dataset had records of multiple countries but my focus was only on India, and so I removed all other country records from the dataset.

ii) Creating the Total Meat Consumption Column: Since the dataset had separate columns for the different types of meat, I added a new column called Total Meat Consumption and summed up the values of all meat categories of a single year. This would help me in simplifying trend analysis and forecasting.

iii) Checking for Missing Values: The dataset that I used luckily didn't have any missing values, but this is a mandatory step to make sure the data has all the records filled.

iv) Converting Data for Model Training: The Year column was set as the independent variable (let's take X for it), and Total Meat Consumption was set as the dependent variable (for this we will take Y). This will help the model to understand the relation between Year and Meat Consumption.

4.2 Feature Engineering:

Process of selecting important data attributes to improve model accuracy. In my project like I mentioned in the Data Processing, Year was used as the independent variable (X) since time affects food consumption trends, while Total Meat Consumption was the dependent variable (y) as it represents overall meat intake. Other columns like Poultry, Beef, Pork, and Fish & Seafood were used only to calculate Total Meat Consumption, this way I avoided making it complex for myself. Since the dataset had few features, Principal Component Analysis (PCA) was not needed, and manual feature selection kept the model simple and easy to understand (Turkiye Study, 2023).

4.3 Machine Learning Model Used:
I selected Linear Regression as the predictive model because it is simple, effective, and widely used for time-series trend analysis. It assumes a linear relationship between the independent variable (Year) and the dependent variable (Total Meat Consumption).

The mathematical representation of Simple Linear Regression is:

$$y=mx+c$$

Where:

y = Total Meat Consumption
x = Year
m = How much meat consumption changer each year
c = The starting value of meat consumption when the year is 0

I used Scikit-Learn's Linear Regression model, because the algorithm calculates m and c by minimizing the sum of squared errors. This ensures that the model finds the best-fit line that captures the trend in meat consumption over time (Jayapal, 2022).

4.4 Model Evaluation:
So, to measure the performance of the model, I used three evaluation metrics:

i) R^2 Score (Coefficient of Determination)
The R^2 Score showed how well the model fits the data. It is calculated using the formula:

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

A score close to 1 indicates that the model fits well.

ii) Mean Absolute Error (MAE)

MAE measures the average absolute difference between actual and predicted values. The formula for MAE is:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i |$$

iii) Root Mean Squared Error (RMSE)
RMSE gives a higher penalty to large errors, making it useful for checking prediction reliability. It is calculated using the formula:

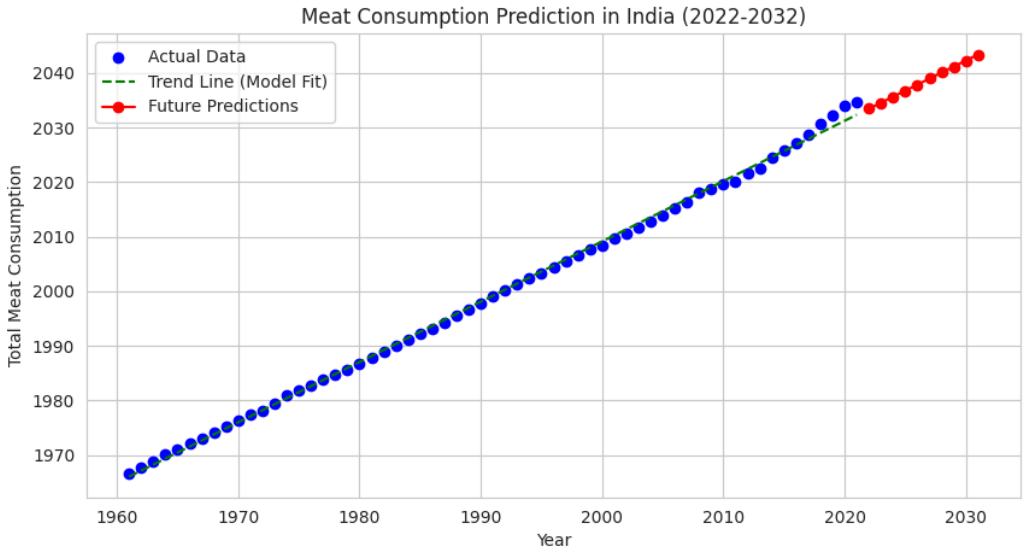
$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

My results after performing these metrics were as follows:

Metric	**Value**
R^2 Score	0.9985
MAE	0.55
RMSE	0.75

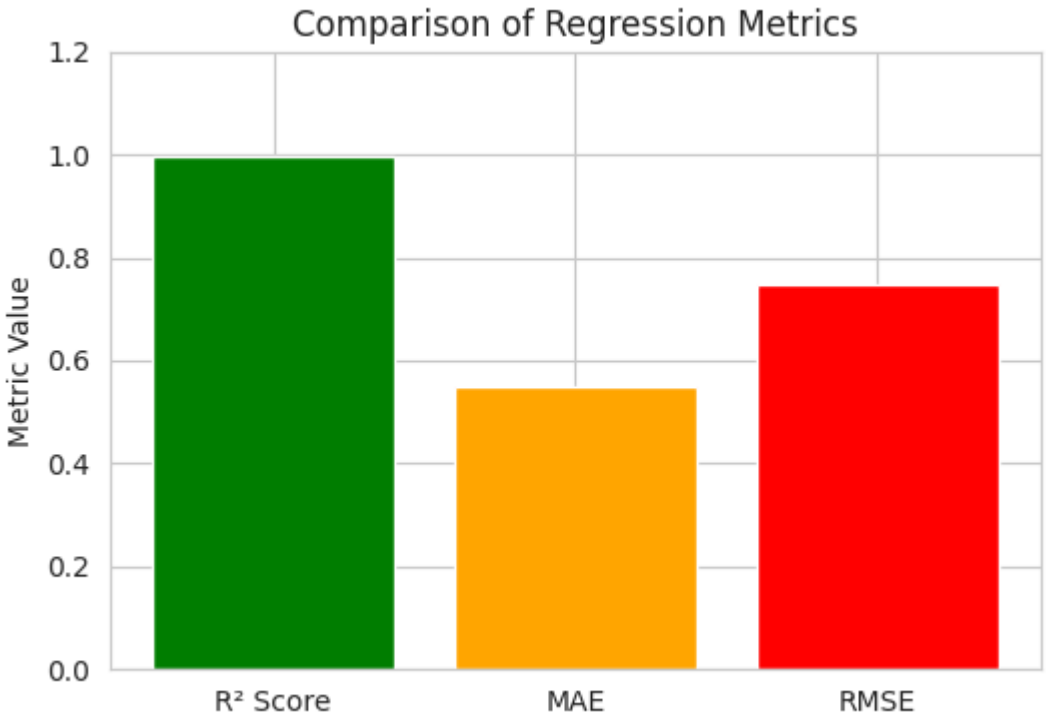
5. Results & Discussion:

After training my Linear Regression Model, I analyzed the predictions for India's Meat Consumption from 2022 to 2032. To ensure the model performed well, I visualized the results, calculated evaluation metrics, and compared the predictions with past trends. In this section, I go through my observations and key findings.



- Blue dots represent actual historical data from 1961 to 2021.
- Green dashed line shows the trend line fitted by the Linear Regression Model.
- Red dots represent the predicted values for 2022 to 2032.

Since my dataset follows a steady increasing trend, Linear Regression was the best choice because it is simple, fast and highly interpretable. While other advanced models like Neural Networks could have been tested, they require large datasets and high computational power, which wasn't necessary for this project.



I found that Linear Regression performed exceptionally well because meat consumption in India has followed a predictable pattern for decades. The model's high R^2 Score (0.9985) confirms that the predictions are reliable (since we have to try to get as close as possible to 1). To make the evaluation clearer, I created the above bar chart comparing my model's R^2 Score, MAE, and RMSE.

6. Conclusion:

In this project, I successfully applied Linear Regression to analyze and predict India's future meat consumption trends. By using historical data from 1961 to 2021, I trained the model to forecast meat consumption up to 2032. The results showed that India's total meat consumption per capita is expected to continue increasing over the next decade.

The evaluation metrics confirmed that the model performed exceptionally well. The R^2 Score of 0.9985 indicated that the model explains almost all variations in the dataset, while low MAE (0.55) and RMSE (0.75) proved that the predictions were highly accurate. The visualizations, including the scatter plot and regression trend line, further confirmed that Linear Regression was a great fit for this dataset.

Linear Regression worked well, but I recognize that other factors such as nation's economic condition, other government policies, and dietary shifts could also impact meat consumption. In the future, I could improve the model by incorporating more external variables or testing advanced machine learning models like Neural Networks.

In the end this project demonstrated that Linear Regression is not only simple but also a powerful tool for forecasting food consumption. The things I found during this project could be useful for policymakers, food industry professionals, and researchers to make informed decisions about India's future food demand.

Acknowledgments:

Thank You, Dr. Amol Vibhute for helping me with Weka and Machine Learning tools which helped me completed this project.

Additionally, I got from Kaggle the Meat Consumption Per Capita dataset, which helped me. I also appreciate the contributions of researchers whose work helped me understand different forecasting techniques and evaluate my model effectively.

References:

Dataset:
https://www.kaggle.com/datasets/scibearia/meat-consumption-per-capita

Other Research Papers:

https://academicjournals.org/journal/AJAR/article-full-text-pdf/998ECB272214
https://jfhs.scientificwebjournals.com/en/download/article-file/2785129
https://www.mdpi.com/2227-7099/12/1/11
https://norma.ncirl.ie/6594/1/sasikumarjayapal.pdf
https://pmc.ncbi.nlm.nih.gov/articles/PMC5404422/

Linear Regression and Metrics Basics:
https://farshadabdulazeez.medium.com/essential-regression-evaluation-metrics-mse-rmse-mae-r%C2%B2-and-adjusted-r%C2%B2-0600daa1c03a
https://www.geeksforgeeks.org/linear-regression-python-implementation/

Python Libraries Documentation:
https://numpy.org/doc/2.2/
https://pandas.pydata.org/docs/
https://matplotlib.org/stable/index.html
https://seaborn.pydata.org/tutorial/introduction.html
https://scikit-learn.org/0.21/documentation.html