Machine Learning Modeling for Cow Price Prediction

BY

A.K.M Tasnim Alam
ID: 191-15-12239

Zahid Hasan Nirob
ID: 191-15-12438

Afrin Jahan Urme
ID: 191-15-12209

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

DR. MD. TAREK HABIB
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised By

MD. SAZZADUR AHAMED
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project/internship titled **"Machine Learning Modeling For Cow Price Prediction"**, submitted by Name: A.K.M Tasnim Alam, ID No: 191-15-12239, Name: Zahid Hasan Nirob, ID No: 191-15-12438, Name: Afrin Jahan Urme, ID No: 191-15-12209 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *23.01.2023.*

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Zahid Hasan
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Fahad Faisal
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Ahmed Wasif Reza
Associate Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Md. Tarek Habib, Associate Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. Md. Tarek Habib
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

A.K.M. Tasnim Alam
ID: -191-15-12239
Department of CSE
Daffodil International University

Nirob

Zahid Hasan Nirob
ID: -191-15-12438
Department of CSE
Daffodil International University

Afrin Jahan Urme

Afrin Jahan Urme
ID: -191-15-12209
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Almighty Allah for His divine blessing that made us possible to complete the final year project successfully.

We are really grateful and wish our profound indebtedness to **Dr. Md. Tarek Habib**, **Associate Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Machine Learning to carry out this project. His endless patience scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan**, Head, Department of CSE, and **Dr. Md. Tarek Habib**, Associate Professor, Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of the CSE department of Daffodil International University.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

This study predicts the price of a cow using data provided by a cow vendor. Cows supply countless things that people utilize on a daily basis, including meat, milk, labor, leather, and many others. The Animal Department of Bangladesh estimates that during Eid al-Adha last year, 10 to 12 million cattle were sacrificed, of them, 4 to 5 million were cows. Therefore, there is a great demand for cows in our country, especially during Eid al-Adha. Most people in our country are unaware of the cost of cows. They are conned into paying extra money for it by a cow salesman. We conducted research on cow price prediction in order to reduce this fraud. We predicted the price of a cow by using machine learning. Machine learning is the finest platform that greatly impacts different corners of science & technology, including the prediction sector. We have acquired data from many sorts of cows utilizing their characteristics, such as height, weight, gender, color, horns, age, teeth, breed, location, and price. Then we applied machine learning algorithms to our processed dataset. In order to determine the accuracy of the performance, we used four well-known machine learning algorithms in our work. They are Gradient Boosting Regression, Decision Tree Regression, Random Forest Regression, and Linear Regression. Linear regression performed the best in our research across all of these machine learning algorithms, with an accuracy rate of 94.9%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 Introduction

A cow is a crucial daily supply of food, nutrition, much-needed money, and nitrogen-rich manure for regenerating soils and other uses. They also take on many different societal roles. The main reason cows are raised as livestock is for their hides, which are used to make leather, milk, and food. Therefore, it follows that it is essential for everything to understand the value of a cow. Price prediction is difficult due to the intricacy of price fluctuation in the market. It also has a significant socioeconomic value. However, in Bangladesh, where we are from, the majority of people live in rural areas. Since they live in villages, agriculture is their main source of income. Cows are the animal that is domesticated in the village's homes the most frequently as a result. However, Bangladesh, a country where Muslims make up the majority, has a high demand for cows especially, in Eid al-Adha. And also dairy cows have a big economic impact. Cows have a big influence on the economy. Those who are unfamiliar with the price of cows are regularly taken advantage of while purchasing and selling them. Knowing the price of cows before selling or buying may be advantageous as a result, and the possibility of fraud will be reduced. Experts in machine learning have previously done a substantial amount of work on price forecasts, including stock and property price predictions, but cow price predictions are quite uncommon. As a result, the goal of this work is to predict how prices using a range of important factors. The price of cows will be predicted in this study utilizing a variety of attributes as well as algorithms. A certain number of people who are unaware of the price of cows will benefit from this and can buy cows at a predicted price. People will benefit from knowing the cow’s genuine predictive price thanks to this study.

1.2 Motivation

Today, cows have a big impact on our daily life and economy. Cows are used in our daily lives to generate a wide range of products, including meat, milk, labor, and leather. The aurochs, a long-horned extinct species of wild cow, are considered to be the origin of domesticated cows. Because Bangladesh is a nation that mainly relies on agriculture, cows have a huge impact on our economy. In this agricultural nation, cows are the primary source of income for the majority of rural households. A lot of cows are needed in Bangladesh. The majority of people in Bangladesh live in rural areas. Cows are essential to farming and other agricultural jobs for people who live in rural areas. The majority of people in Bangladesh are not aware of the price of a cow. During the Eid al-Adha holiday, a large number of people go to the market to purchase cows. Because most individuals don't realize how much a cow is worth, vendors frequently mislead purchasers by offering cows at exaggerated rates. By determining the cow's forecast price, we hope to awaken the citizens of our nation. Our goal in doing this study is to safeguard customers from fraud when they go to buy cows. To forecast the price of cows, we employ machine learning.

1.3 Rationale of the Study

The results of this study will assist individuals in avoiding being duped by cow vendors at cattle markets. For this reason, the use of machine learning techniques greatly aids in the identification of predictions. If we use machine learning to identify the anticipated price of a cow, we can lower the number of customers who are taken advantage of by cow vendors in cattle markets.

1.4 Research Questions

- **(i)** How we will collect our data?
- **(ii)** How much data do we have to collect?
- **(iii)** How will our original data set look like?
- **(iv)** How we will predict the price of a cow?
- **(v)** Will our data set & machine learning will be compatible?

(vi) Which technique of machine learning we should use?

1.5 Expected Outcome

Using the results of our research, we are able to inform individuals about the actual price of a cow and assist them in determining the price that is expected for one. The majority of people in our nation don't know how much a cow costs. They visit the cattle market, where they are frequently taken advantage of by the seller of cows because they are unaware of the actual price of the animal. As there hasn't been a lot of research on cow price prediction, we've looked at features like weight, height, gender, and the colors black, red, white, brown, and grey as well as the cow's horn, age, teeth, breed, location, and price in order to predict the cow prices with a higher degree of accuracy. In turn, this allows us to demonstrate the better accuracy of the cow's price prediction.

1.6 Project Management

When we first began our research, we knew very little about machine learning and research techniques. Our honorable supervisor helps us a lot in our whole work. The data collection process was done by A.K.M Tasnim Alam, Zahid Hasan Nirob & Afrin Jahan Urme. The implementation process was done by our group leader A.K.M Tasnim Alam. In our whole report chapter 1,5 was written by Zahid Hasan Nirob, chapter 2 was written by Afrin Jahan Urme & chapter 3,4 was written by A.K.M Tasnim Alam.

1.7 Report Layout

This research paper contains the following contents as given below:

- In chapter one, we go over the research's introduction, as well as its rationale, research questions, and anticipated results.
- In chapter two, we go over related publications, a review of the research, and the size and complexity of the topic.

- In chapter three we discuss about the Data gathering procedure, Propose methodology, statistical analysis & feature implementation, and workflow of this research.
- In chapter four we discuss the result and discussion.
- In chapter five, we go over the research summary, its limitations, and possible future directions.

CHAPTER TWO

BACKGROUND STUDY

2.1 Introduction

In this section, we look at pertinent works, a study overview, the scope of the topic, and issues. In the related work section, we list many research articles, related studies, underlying methodology, classifiers, and accuracy that are pertinent to our research. For a better and simpler understanding, we compile a summary of a few chosen pertinent studies and arrange them in a table in the research summary section. The scope of the problem section describes how our organizational paradigm may exacerbate the issue. Finally, we highlight the difficulties and risks we ran into while doing this research in the Challenges section.

2.2 Related Works

This research study's literature review part will feature current related studies on cow price prediction that have been written by various researchers. We have followed and studied their work in order to better understand the techniques and strategies they used for their research.

Rahman et al. [1] researched predicting cow pricing ranges by using any cow photograph. Cow photographs were gathered from a variety of online e-commerce sites that sell cows in order to primarily anticipate the price range of cows based on the photos of the cows.

Cows are divided into four price categories: low, medium, high, and very high classes. In a machine learning-driven approach to prediction, a convolutional neural network (CNN) is employed as an image classifier and linear regression is used to forecast pricing. For this process, close to 1000 data points from various websites relating to cows are gathered. The majority of the data points were obtained from the "goruchai.com" website. 80% of the data is used for training, and 20% is used for testing.

This work has an F1 score of 65.73%, 70.05% accuracy, 68.69% precision, and 70.05% recall. The results show that the price range of a cow may be predicted with 70% accuracy.

Lawrence et al. [2] analyzed the Cattle Price Forecast Errors based on Live Cattle Futures and Seasonal Index. Using the closing futures price one week after the Cattle on Feed report was published for January, April, July, and October, he assessed the Live Cattle Futures Market projection and modified it for the prior five-year average basis. Each month's price was predicted, and the quarter's price was calculated by averaging the three months. In order to estimate a price for each of the following 12 months, the Seasonal Index was based on the monthly average price for the same month as the report (i.e., January average price following the January report), and then three months were averaged into each corresponding quarter. Then, for the first quarter of 1990 through the fourth quarter of, these predictions were contrasted with the actual average price. The projections for the four quarters, which range from -.3% to 1.1% of the actual average quarterly prices (1% of $90/cwt is $.90), in 2008 generally perform extremely well. The Index can over or under-predict events. Futures frequently miscalculated real prices by a tiny margin, ranging from.4% to 1.1%. Overall, the Seasonal Index's predictions performed somewhat better than the Futures, with an average inaccuracy of -.1% as opposed to.7%. The average forecast inaccuracy among researchers predicts that the actual price will typically be.4% higher than the basis-adjusted Futures price.

Bozic et al. [3] work on Rating Livestock Margin Insurance for Dairy Cattle Using Parametric Bootstrap Tests for Futures Price and Implied Volatility Biases. For testing the presence of bias in futures prices and implied volatility in deferred contracts with overlapping time-to-maturity horizons, the authors of this paper have developed a parametric bootstrap method. They use their methodology to test the hypotheses that futures prices are accurate and impartial predictors of terminal prices and that implied volatility squared times remaining time to expiration is an accurate predictor of terminal log-price variance. They apply the test to data for Class III milk futures and options from 2000 to 2011 as well as maize, soybean meal, and soybean meal.

They employ the LGM-Dairy rating method, which is predicated on the unbiasedness of implied volatilities derived from at-the-money options and futures prices. Contrarily, biases in Class III milk implied volatilities may imply that LGM-Dairy premiums are undervalued. They also performed Monte Carlo Experiments, which suggest that depending on the risk management technique adopted, correcting for biases in implied volatility coefficients could increase the cost of LGM-Dairy insurance by between 3% and 21%.

Eldridge et al. [4] have used included two distinct yet closely related components of the decision tool for grazing management and profitability predictions and the pricing analysis of feeder cattle. In this analysis, the premium paid to CPH cattle for cattle weighing 400 to 499 pounds, 500 to 599 pounds, and 600 to 699 pounds, respectively, was over $6/cwt, $3.50/cwt, and $1.50/cwt. Since this study only included lots of 20 or more heads, one might speculate that a sizable percentage of this premium is for the health program that CPH calves receive rather than just their lot size. The state of Kentucky's four main auction markets were also included in this model.

For the majority of the eight weight categories, three out of the four showed no discernible price difference. However, in each of the six models in which it was included, the Kentucky Tennessee Livestock Market displayed considerable discounts ranging from over $2/cwt to over $5/cwt.

Kinnischtzke et al. [5] analyzed identifying key, easily accessible features of the beef supply that influence changes in feeder and slaughter cattle prices. Data were derived from the USDA's quarterly and monthly cattle-On-Feed data. Theoretical predictions of specific monthly feeder and slaughter cattle price patterns were made using logistic regression. With monthly cattle-on-feed, placements, and marketing serving as the independent variables and upward or downward trends in cattle prices as the dependent variable, multivariate models are used. The relevant variables that coincided with changes in price trends were identified using multi-period analysis of monthly, quarterly, and yearly price trends and related data.

Overall monthly equation results showed a projective accurate price trend of 81.4, 83.7, and 78.2 percent on average. When the number of steers weighing more than 1100 lb. is rising, a downward price trend is about to begin. Numbers rising in the steer weight category under 500 1b and the heifer weight category between 500 and 699 1b. indicate rising price indices.

Muwanga et al. [6] conducted relying on a few U.S. cattle marketplaces, this study examined the spatial linkages, price distributions, and price forecasts for cattle. In this research, the characteristics of distinct cattle price series, the application of a rational expectations model to cattle pricing for various marketing locations and cattle classes, and the spatial price correlations for certain markets were all investigated.

The factors that influence cattle prices, the theoretical and practical ramifications of price dispersion features for market structure and behavior, and a comparison of forecasting systems' accuracy.

Marsh et al. [7] predicting Quarterly Live Cattle Prices Using a Rational Distributed Lag Model. In order to reduce issues with specification mistakes in the disturbance structure, quarterly U.S. feeder cattle, and fed cattle prices were calculated inside a logically distributed lag framework. Each reduced form equation's price variation was explained by the behavior of certain input and output variables in the cattle market. The findings showed dynamic stability, with feeder cattle prices having a polynomial rational lag structure and fed cattle prices having a geometric rational lag structure. It was discovered that these dynamics outperformed both static-serial correlation and wholly autoregressive requirements.

Franzmann et al. [8] devised Trend Models of Feeder, Slaughter, and Wholesale Beef Cattle Prices. In order to give a low-cost alternative method of projecting prices at three market levels in the beef industry and to provide information on the temporal interrelationships among these market levels, harmonic regressions were fitted to monthly data. The findings confirm preexisting notions about time linkages and have little predictive power.

Coffey et al. [9] examined the impacts of alterations in price momentum and market factors on hedging live cattle. The effect of shifting fundamental economic variables and price trends on the predictability of the basis for live cattle is quantified in this study. In order to understand how changes in the live cattle market and current market conditions affect basis predictability, basis prediction errors are modeled as functions of pertinent economic variables.

Price momentum, which is utilized to identify trends in price movement, is one of the explanatory factors. The procedure is performed for each of the five main Livestock Mandatory Price Reporting (LMR) zones (Colorado, Iowa/Minnesota, Kansas) for live cattle prices. Results show that contemporaneous factors, we were also unfamiliar with Anaconda, Jupyter Notebook, and several new machine-learning algorithms. At first, it took us some time to comprehend and learn about it, but with the help of our supervisor and further practice, we were able to pick it up quickly. Then, we keep working diligently and enthusiastically. r, the weight of cattle marketed, and the thinness of the negotiated cash market statistically affect basis prediction errors.

Moser et al. [10] measured the precision of direct genetic values in Holstein bulls and cows was assessed using certain subsets of SNP markers. This study's goal was to assess how well low-density assays predicted direct genomic value (DGV) for two profit index characteristics, a survival index, an overall conformation trait, and five milk production traits (APR, ASI). Dense SNP genotypes for 2,114 Holstein bulls and 510 cows were obtained for 42,576 SNP. To fit models with various sets of pre-selected SNP, a subset of 1,847 bulls born between 1955 and 2004 served as the training set. To create prediction equations and order SNP depending on the magnitude of the regression coefficients, ridge regression (RR) and partial least squares regression (PLSR) were utilized. A single subset of evenly spaced SNP or subsets of the high-ranking SNP for each particular trait were both used to choose subsets of SNP. SNPwae has been chosen based on their rank for ASI, APR, or minor allele frequency within intervals of roughly similar length. For low-density assays, PLSR performed better than RR, but RR performed better for higher-density SNP sets. RR and PLSR performed quite similarly to predict DGV.

DGV predictions for variables with a greater heritability, such as productivity, were more accurate (0.52-0.64) when utilizing all SNPs than for qualities with a lower heritability, such as survival (0.19-0.20). When at least 3,000 SNP were used, there was only a slight (5–6%) improvement in accuracy when utilizing subsets that comprised the highest-ranking SNP for each attribute as opposed to a common set of uniformly distributed SNP. For cows, subsets with 3,000 SNP offered more than 90% of the accuracy possible, and for young bulls, 80% of the accuracy was possible with a high-density test.

Blakely et al. [11] assessed a quarterly cat farm price projection model with a view to developing a futures market approach. According to the structural model, the inventory of heifers, steers, and bulls weighing less than 500 pounds, the price of maize, the price of fed cattle, and the price of fed cattle with a two-period lag are some of the more significant variables in determining the price of feeder cattle. Analytically reduced forms have low usefulness in predicting the price of feeder cattle; instead, it would appear preferable to create an estimate of the price of feeder cattle using a straightforward OLS model. The transfer function forecasting model can then employ an OLS estimator as an input series. The transfer function model will work best if a leading indicator series can be established that can be utilized as an input series. Although the futures market would seem to be the ideal source for such a series, the database is currently too small to include futures prices in a function that forecasts the price of feeder cattle. A few straightforward tactics that combine futures markets with price projections show promise for both lowering the revenue variance that would be observed under a cash-only business and increasing the average income obtained by feeder cattle producers. It appears encouraging to conduct more research in the field of developing strategies. The output series model's MSE was 29.34, the transfer function model's MSE was 5.80, and the futures market's MSE was

74.74. The exclusive cash approach produced a $4.03 average per head with a $43.33 standard error.

2.3 Comparative Analysis and Summary

The machine learning system has already been used to forecast and detect changes in cattle prices. With the use of tobacco user identification, disease detection, and alcohol user prediction, machine learning technology is now more widely used. This section has demonstrated the comparison between these connected works. Here, Table 2.1 provides a comparison of several research projects with respect to their subject, technique, and conclusion.

TABLE 2.1: SUMMARY OF RELATED RESEARCH WORK

Authors	**Animal (deal with)**	**Descriptions**	**Size of Dataset**	**Algorithm/Method**	**Accuracy**
Rahman et al. [1]	Cow	Prediction of cow prices using machine learning	1000	CNN and Image Processing	70%
Lawrence et al. [2]	Cattle	Live cattle futures and the seasonal index both predict cattle prices correctly	NM	The Index forecast method	NM
Bozic et al. [3]	Cattle	Rating Livestock Margin Insurance for Dairy Cattle Using Parametric Bootstrap Tests for Futures Price and Implied Volatility Biases	500	The LGM-Dairy rating method	21%
Eldridge et. al. [4]	Cattle	Cattle price predictions and grazing management	508	Linear regression model, Feeder Cattle Price Model	90%

Kinnischtzke et al. [5]	Cattle	Trends in Cattle Price Prediction	1100	Logistic regression, Binary Choice Models, Linear Probability model	83.7%

Muwanga et al. [6]	Cattle	Spatial linkages, price distributions, and price forecasts for cattle	800	VAR Model, ECG Model, Granger-Causality Error Correction Model, SGC Model, Hypotheses and Empirical Price Dispersion Models	90%
Marsh et al. [7]	Cattle	Prices for Quarterly Live Cattle Based on a Rational Distributed Lag Model	NM	Distributed Lag Model	NM
Franzmann et al. [8]	Cattle	Feeder, Slaughter, and Wholesale Beef Cattle Price Trend Models	NM	The linear trend model	NM
Coffey et al. [9]	Cattle	Changes in Market Fundamentals and Price Momentum and the Effects on Hedging Live Cattle	205	LMR	NM

Moser et al. [10]	Bulls and cows	Accuracy of direct genomic data in Holstein bulls and cows utilizing subsets of SNP markers	2654	Ridge regression (RR), Partial least squares regression (PLSR)	90%
Blakely et al. [11]	Cattle	A model for forecasting quarterly cattle prices with application to the development of a futures market strategy	NM	The forecasting model, The Moving Average Model, Transfer Function Models	81%

Recent advancements in machine learning, AI, and deep learning are being studied for use in any type of prediction model. Several machine learning techniques have recently been used to diagnose, predict, and find content. For any prediction model, it is usual to employ techniques like CNN, SVM, the forecast model, logistic regression, and linear regression. In this work, a number of methods, including linear regression, random forest regressor, K-Neighbors classifier, random forest classifier, gradient tree regressor, decision tree classifier, and more, were used to forecast the price of cows in Bangladesh.

2.4 Scope of the Problem

Our research focuses on developing models through data analysis and machine learning techniques. The price of cows can be predicted using our suggested approach. The economy and society will be significantly impacted by this prediction. Because it is crucial that everyone comprehends the worth of a cow. Because price signals and databases are so intricate, price prediction is challenging. However, there is a significant demand for cows in Bangladesh, a country where Muslims are the majority. On the other side, some people frequently take advantage of those who are unaware of the value of cows when buying and selling them.

It may be useful to know the price of cows before selling or buying, and the likelihood of fraud will be decreased. In order to forecast the price of cows, this approach would be helpful for regular people and ethical individuals. Cattle price predictions are rather unusual, despite the fact that machine learning experts have previously done a significant amount of work on price forecasts, including stock and property price predictions. Therefore, the objective of this research is to forecast cow's price using a variety of significant criteria. And in this area, our findings will undoubtedly be helpful. The outcomes are pretty acceptable when machine learning and artificial intelligence are employed recently for a variety of object predictions. Therefore, we made the decision to develop a model for predicting cow prices using machine learning.

2.5 Challenges

We encountered various issues when conducting our research. We had some difficulty gathering data because cowherds did not respond to our requests in a timely manner. Because of this, we were forced to gather information through in-person interviews with cattlemen, which proved challenging given the pandemic situation. Additionally, it is difficult to distinguish between price projections for cows based on their weight, height, and color.

We visited numerous cow markets, read numerous publications, and spoke with numerous cattlemen, but the majority of them refused to provide any information about their cows. Since we knew few professionals who could support our target class, it was extremely difficult to gather accurate information. We were then able to gather our data from several cowherds and cattlemen. We looked for other farms and markets, but they refused to provide us with any information due to privacy concerns. Finally, with much difficulty because of their extreme commercial activity throughout the Eid holiday, we were able to speak with a number of professionals. We were also unfamiliar with Anaconda, Jupyter Notebook, and several new machine-learning algorithms. At first, it took us some time to comprehend and learn about it, but with the help of our supervisor and further practice, we were able to pick it up quickly. Then, we keep working diligently and enthusiastically

CHAPTER THREE
RESEARCH DATA AND METHODOLOGY

3.1 Introduction

This study's primary goal is to create a model for estimating the price of a cow. The number of factors determines how much a cow will cost. We utilized some machine-learning algorithms to forecast the price. In this research, we used Decision Tree Regression, Random Tree Regression, Linear Regression & Gradient Boosting Regression. These algorithms' primary function in the approach was a regression. There are sixteen important attributes in the dataset that is necessary to forecast the cost of a cow. We meticulously processed and examined our dataset prior to implementation. To choose the best method for the model, we estimated and computed each algorithm's accuracy.

3.2 Data Gathering

The collection of proper data was our top priority because this area was extremely challenging for us and served as the focal point of our research. Cow salesmen occasionally avoid interacting with us because we weren't legitimate buyers and also they get easily distracted by others customers. We used to gather information from buyers and sellers at the cattle market door. Sometimes, we separate and interview people who have previously bought a cow to gather information from them. Online data was never an option because we intended to get data directly from the field. We gathered our data using the following criteria.

1. What is the weight?
2. What is the height?
3. What is gender?
4. Which kind of female?
5. What is the color?

6. Is the cow single color or multicolor?
7. What is horn status?
8. What is the age?
9. Number of teeth?
10. What is bread?
11. From which place does the cow come?
12. What is the price?

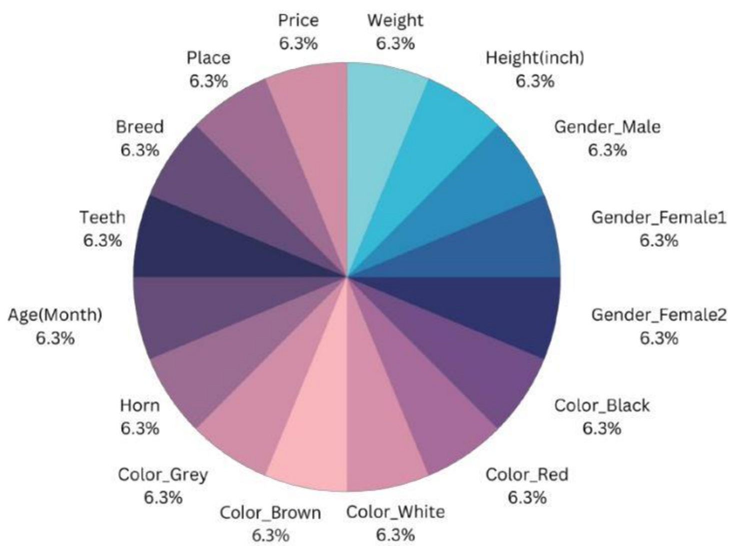

Figure 3.2: Diagram of Dataset

To forecast the price of a cow, we had to take into account every one of these criteria. By speaking with various experts and cattle traders, and using several websites and papers, we learn about all of these characteristics.

3.3 Research Instrumentation

People can use these predictive models to create price predictions for cows. For any type of prediction and detection, data mining, machine learning techniques, and deep learning are particularly popular. To determine which algorithm will work best for our model, we have applied a variety of algorithms to the data we have gathered. Numerous machine-learning techniques, including linear regression, Decision Tree, random forests, and Gradient Boosting have been applied. In our research, we selected 'Microsoft Excel' for our dataset, 'Python' as our programming language, and 'Anaconda Navigator', 'Jupyter Notebook', and other tools for data mining [4].

3.4 Proposed Methodology

We first gathered 1000 data from the field and after that, we analyzed our dataset. We utilized seven feature selection techniques after preprocessing our dataset, including taking all characteristics. We have used some machine-learning algorithms like Random Forest (RF), Decision tree (DT), Linear regression (LR), and Gradient Boosting regression (GBR). We assessed how well each technique performed. We chose the most effective approach based on their performance. Then we evaluate the regression models with metrics like Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

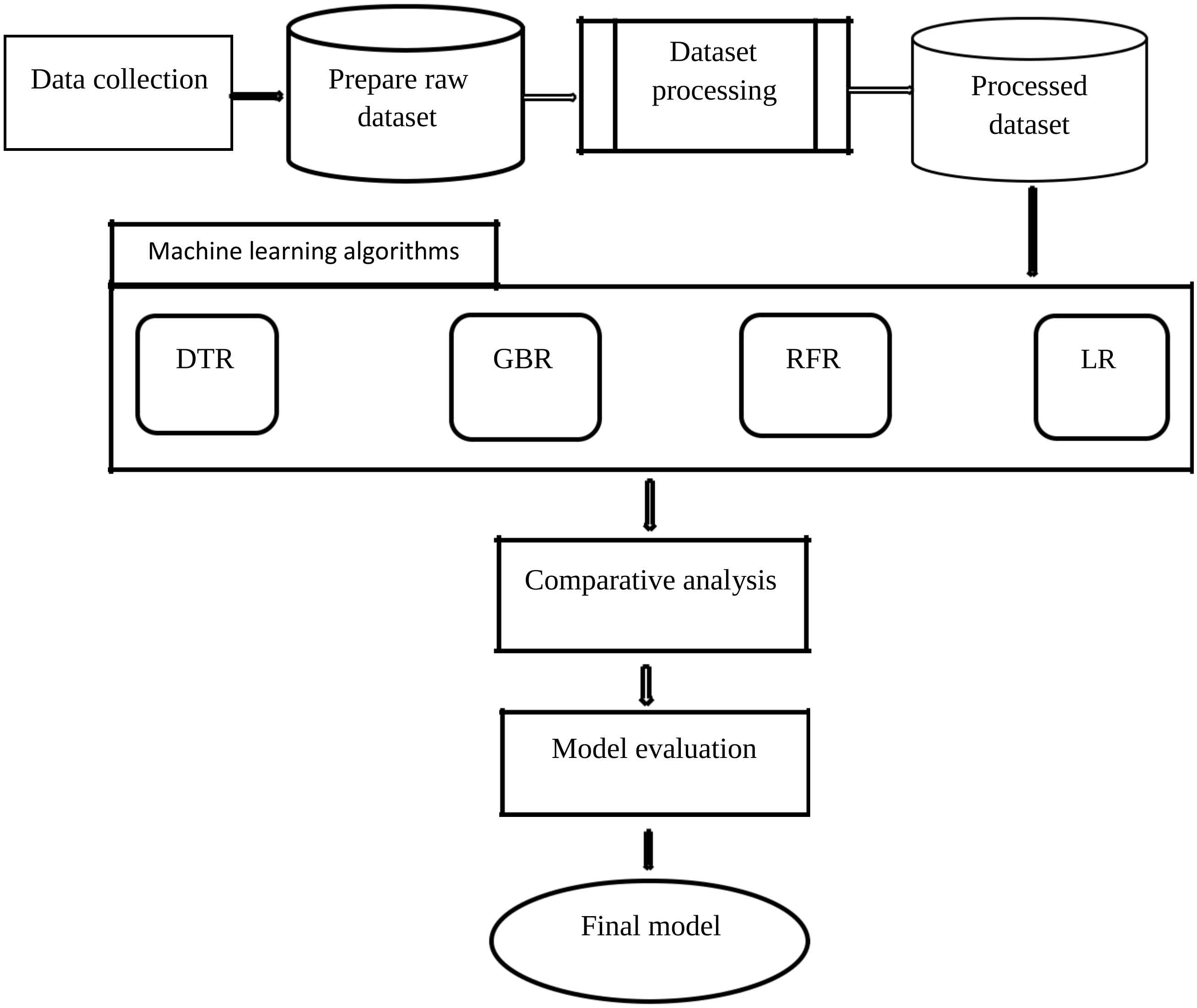

Figure 3.4 Proposed methodology of the research

3.5 Data Processing

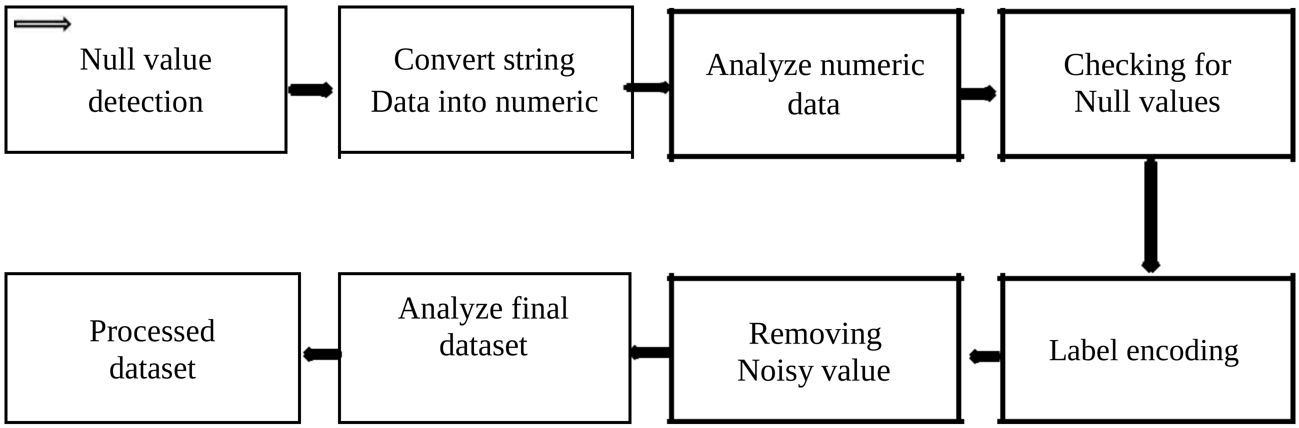

Figure 3.5: Data preprocessing steps

The first and most crucial step in creating a machine learning model is data preprocessing, which entails getting the raw data ready and making it suitable for the model because real-world data is in an improper format or has missing values or noise, it can't be used straight in machine learning models. In order to clean data and make it usable for a machine learning model, data preparation is an essential task. This increases the model's effectiveness and accuracy [1]. First, from the collected data, we obtained basic categorical data, numerical data, and text data. Then we made the decision to apply data processing to transform this data into an algorithmic format. Start the data cleansing process first. We moved on to the following stage if there is a null value in the data set.

It has been encoded at the level where text data is converted to numerical data. The correlation heat map was then examined as a means of data integration. This map displays the ratio of each data connected to the other. We used outlier quantile detection to filter out noisy values. Data reduction and data visualization in feature engineering was made easier for us by using separate histograms for each feature. Throughout the entire data processing procedure, the "Jupyter Notebook" and "Anaconda navigator" were used. In Figure 3.6, the correlation heat map is displayed.

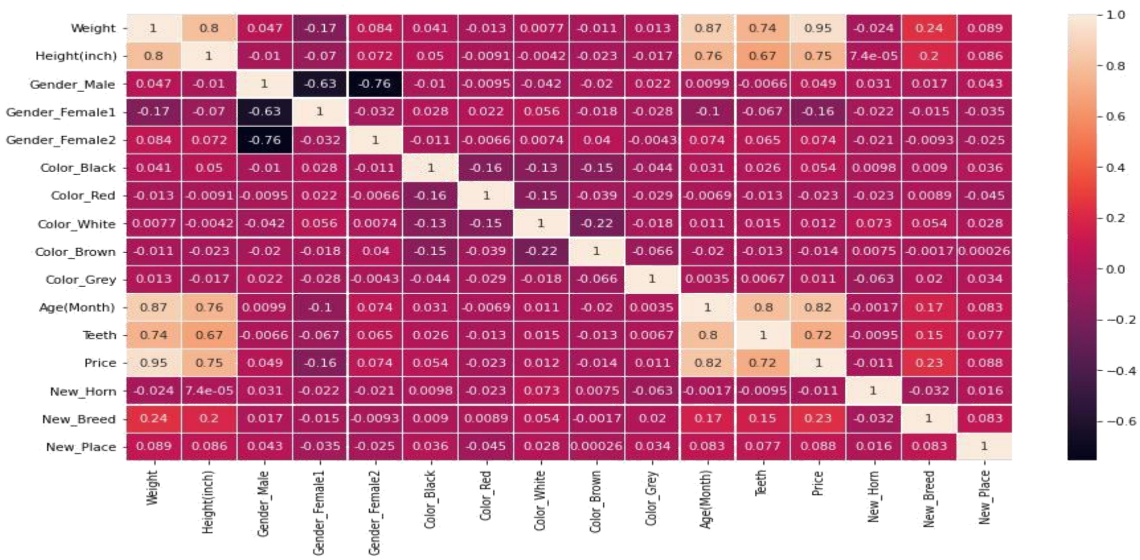

Figure 3.6: Heat-map

3.7 Statistical Analysis

We created sixteen features after talking with the professionals and cow dealers. Feature names are:

TABLE 3.7: LIST OF FEATURES NAME

SL	Questions	Feature Name
1	What is the weight?	Weight
2	What is the height?	Height(inch)
3	Is the gender male?	Gender_Male
4	Is the cow female but never a have a baby?	Gender_Female1
5	Is the cow female and was pregnant before?	Gender_Female2
6	Is the cow's color black?	Color_Black

7	Is the cow’s color red?	Color_Red
8	Is the cow’s color white?	Color_White
9	Is the cow’s color brown?	Color_Brown
10	Is the cow’s color grey?	Color_Grey
11	How is the horn?	Horn
12	What is the age?	Age (Month)
13	How many teeth does he have?	Teeth
14	What is the breed?	Breed
15	Where is he come from?	Place
16	What is the price?	Price

3.8 Conversion of Text Feature Value to Numeric Value

TABLE 3.8: DESCRIPTION OF TEXT FEATURE VALUE INTO NUMERIC VALUE

SL	**Feature Name**	**Value Name**	**Description**	**Figure Number**
1	Horn	1	Straight	1
		2	Curly	
		3	Defected	
2		1	Deshi	14
		2	Shahiwal	
		3	Red Shindhi	

	Breed	4	Friesian	
		5	Greycattle	
		6	Mirkadim	
		7	Hasba	
		8	Haryana	
		9	Australian	
		10	Potash	
		11	NBG	

3	Place	1	Chapainawabganj	15
		2	Sirajganj	
		3	Pabna	
		4	Bogura	
		5	Punjab	
		6	Jaypurhat	
		7	Lalmonirhat	
		8	Manikganj	
		9	Kustia	
		10	Natore	
		11	Rajshahi	
		12	Hydrabad	
		13	Bikaner	
		14	Haryana	

		15	Rajasthan	
		16	Naogaon	
		17	North holland	
		18	Friesland	
		19	Schleswig-Holstein	
		20	Hungary	
		21	Bihar	
		22	Uttar pradesh	
		23	Australia	
		24	Noakhali	
		25	Dhaka	
		26	Laxmipur	
		27	Feni	
		28	Cumilla	
		29	Chittagong	

3.9 Statistical analysis of Weight Range

The price of a cow primarily depends on its weight. Caws can gain up to 500–600 kg depending on their age and breed. In our dataset minimum weight is 47kg and the maximum weight is 460kg. 50% of our data are around 170kg. The histogram, of weight, is shown in figure 3.9.

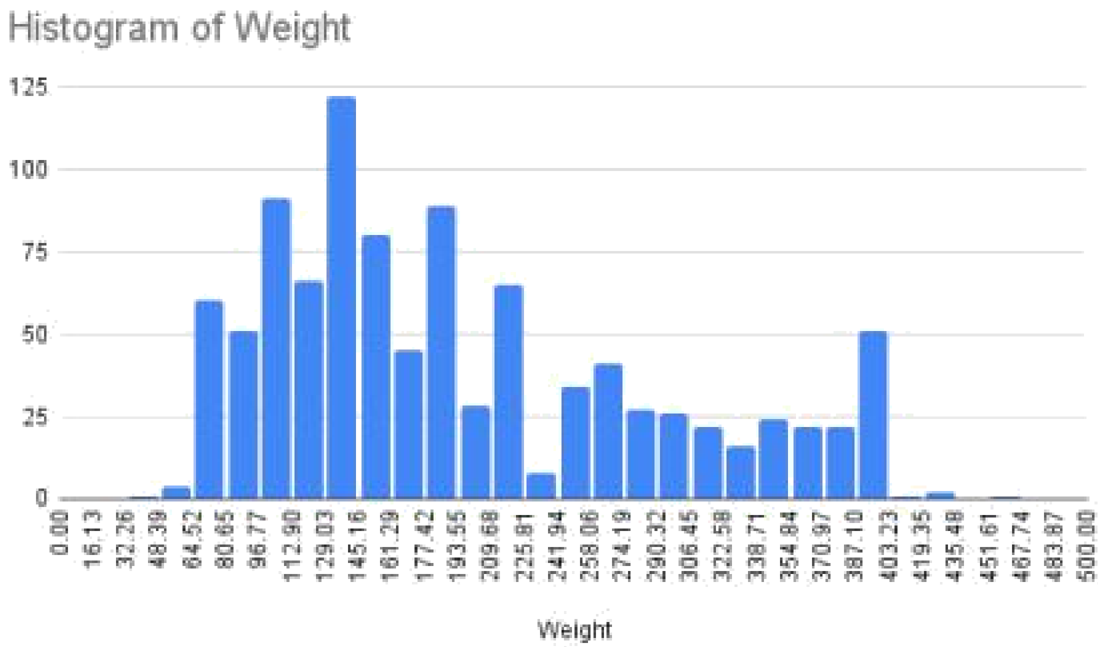

Figure 3.9: Histogram of Weight Range

3.10. Statistical Analysis of Height

Height is the main consideration when determining a cow's shape. In addition to weight, a cow's price is also influenced by its height. The minimum height in our dataset is 31inch and the max is 75 inches. The statistics of height are shown in 3.4.3.

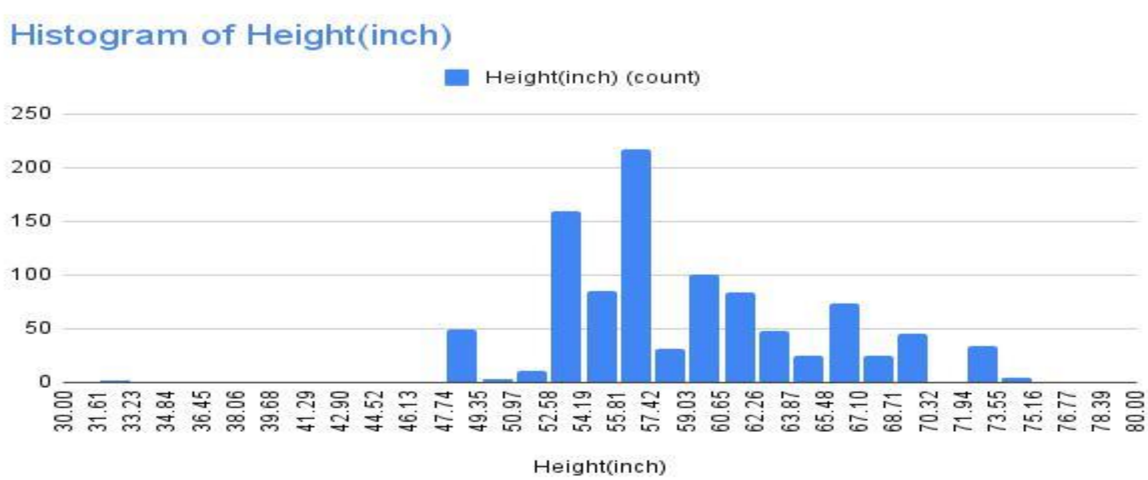

Figure 3.10: Histogram of Height

3.11 Statistical Analysis of Gender

The price of a cow might vary significantly depending on the cow's gender. About 95% of our data is male and other around 5% of our data is female. The statistics of gender are shown in Figure 3.11.

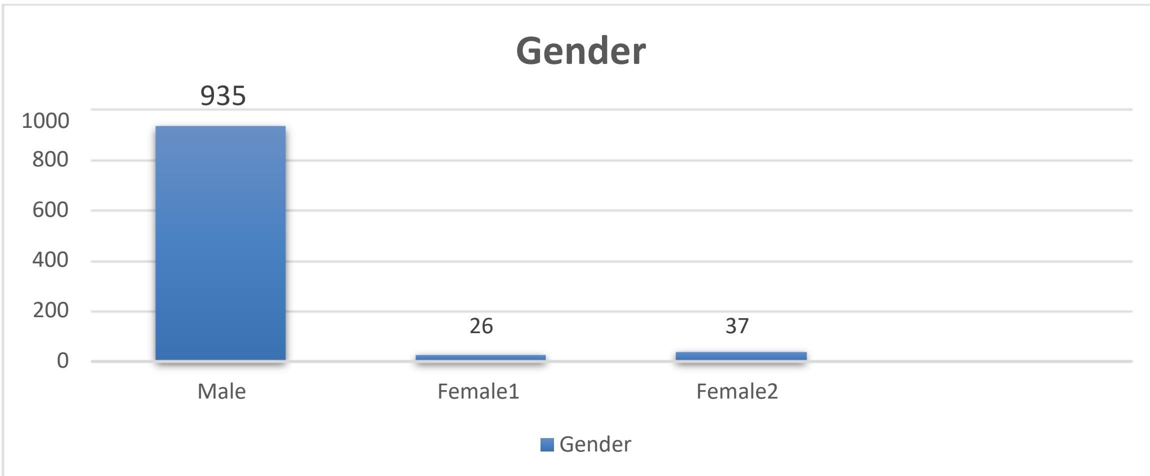

Figure 3.11: Histogram of Gender

3.12 Statistical Analysis of Color

Color is what makes a cow beautiful. A cow might have one color or several. Around 70% of cows in our dataset are in multi-color. The statistics of color are shown in figure 3.12.

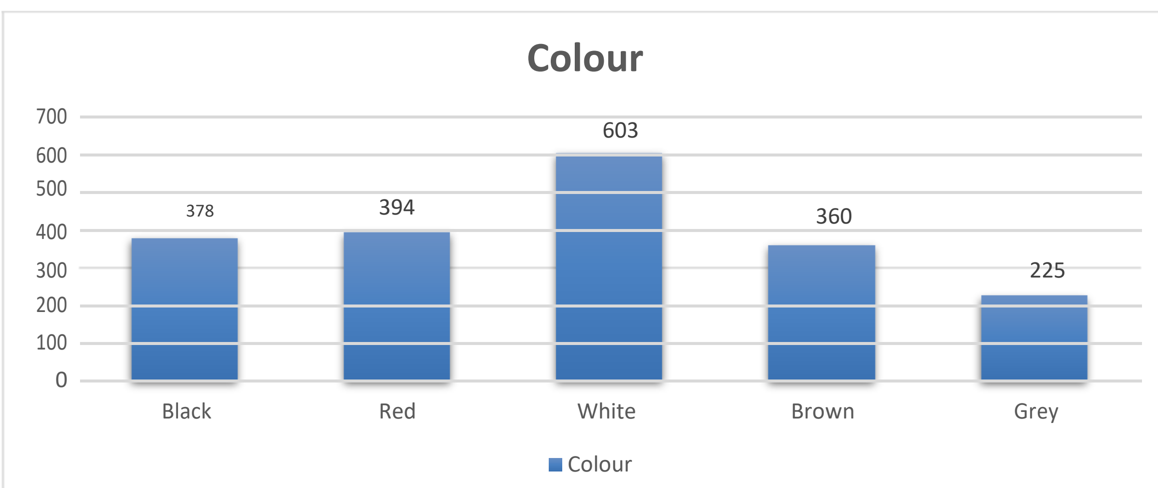

Figure 3.12: Histogram of Color

3.13 Statistical Analysis Horn

Horn is also a beautiful part of a cow. There are three types of horns in our dataset, Straight, Defected and Curly. The statistics are shown in figure 3.13.

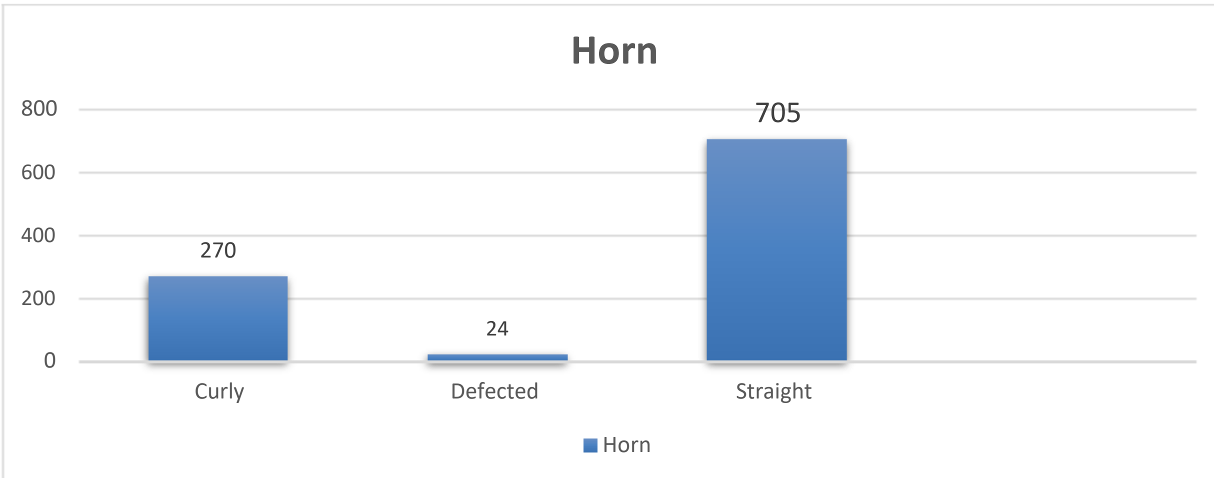

Figure 3.13: Histogram of Horn

3.14 Statistical Analysis of Age

The primary determinant of a cow's weight is age. A cow is permitted for Eid-ul Adha after a specific age. The statistics of age are shown in figure 3.14.

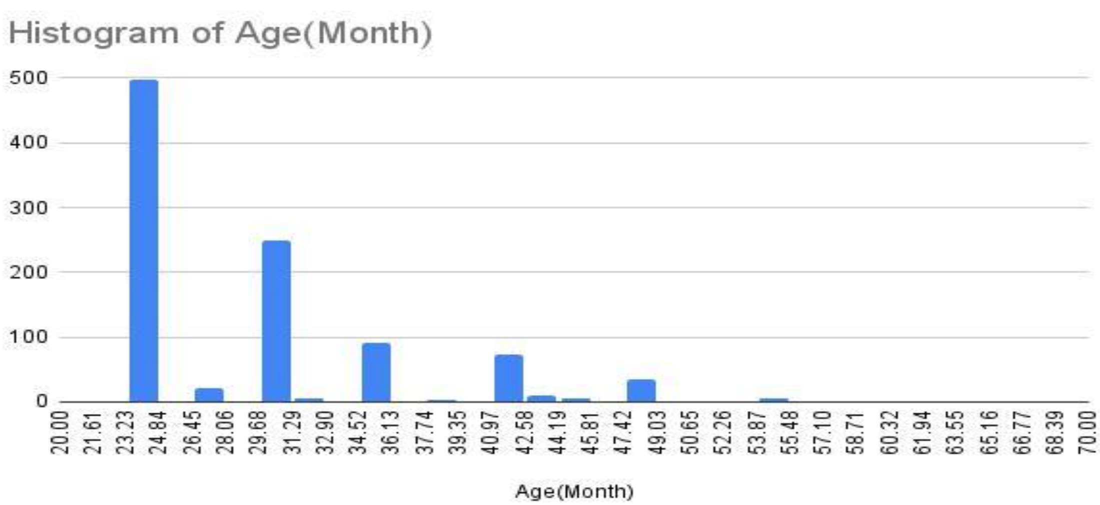

Figure 3.14: Histogram of Age (Month).

3.15 Statistical Analysis of Teeth

A cow must also have a certain number of teeth in order to receive permission for sacrifice for Eid al-Adha. The statistics of teeth are shown in figure 3.15.

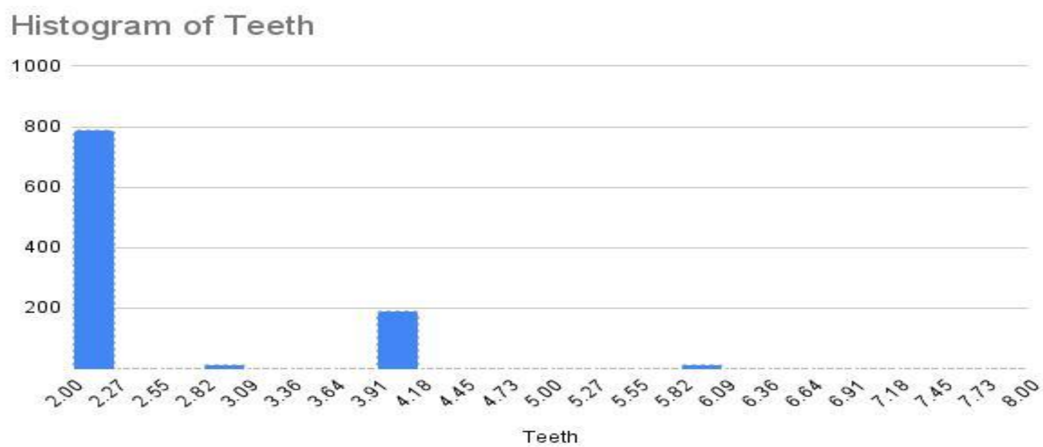

Figure 3.15: Histogram of Teeth.

3.16 Statistical Analysis of Breed

The breed of a cow can also affect its price. In our dataset, there are eleven breeds of cows. They are Australian, Deshi, Friesian, Grey cattle, Haryana, Hasba, Mirkadim, NBG, Potash, Red Sindh, and Sahiwal. The statistics of Breed are shown in figure 3.16.

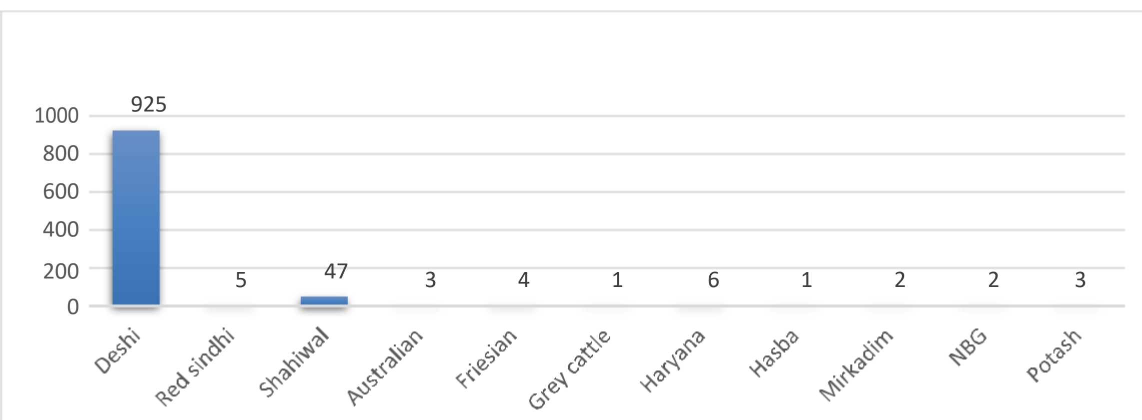

Figure 3.16: Histogram of Breed.

3.17 Statistical Analysis of Place

Our dataset includes primarily deshi cows. Bangladesh is where the cow was born the majority of the time. Depending on where the cow comes from, the price may differ. The statistics of Place are shown in figure 3.17.

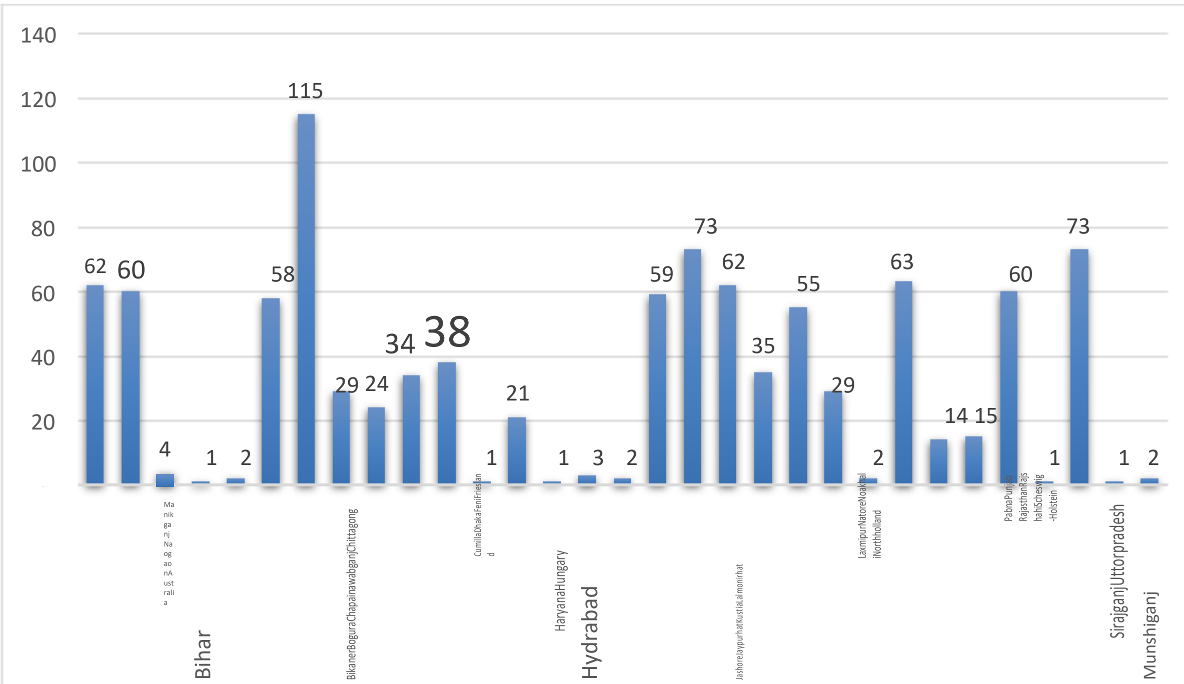

Figure 3.17: Histogram of Place

3.18 Statistical Analysis of Price

The main motive of this research was predicting cow's price. Price can differ based on previous attributes. The statistics of price are shown in figure 3.18.

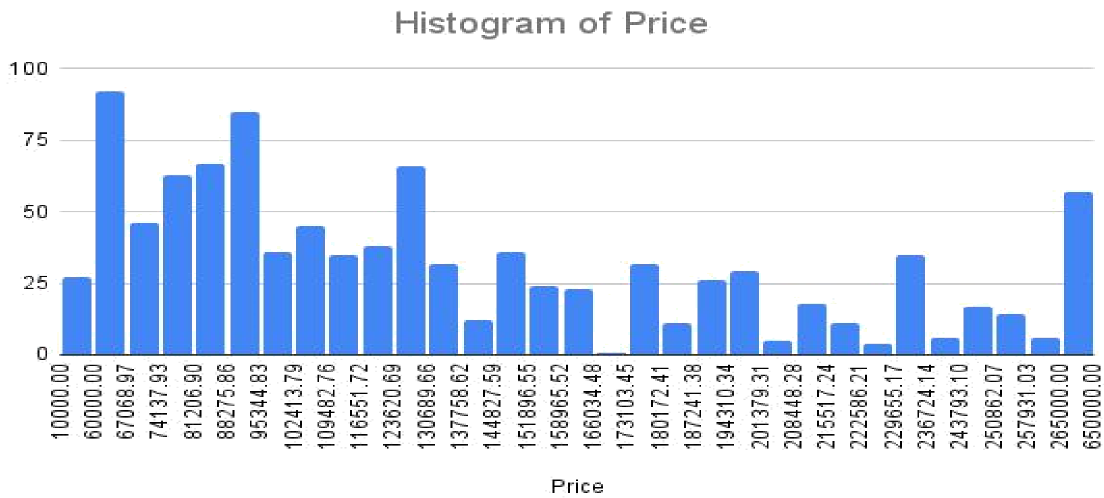

Figure 3.18: Histogram of Price

3.19 Implementation Requirements

To carry out our research, we require tools for data mining, data processing, and data storage. We collect our data from the field and from the cow dealers. We used Microsoft Excel to produce data sets. The software "Anaconda Navigator and Jupyter Notebook" was utilized for data pretreatment and algorithm implementation.

An example of a graphical user interface for the desktop is Anaconda Navigator, which enables users to start applications and conda packages, environments, and channels without the need for command-line commands. Anaconda also contains finished and open-source data science packages [11].

3.20 Methods and Algorithms

Mostly, Machine Learning algorithms are applied to predict prices. In this research, we used Decision Tree Regression, Random Tree Regression, Linear Regression, and Gradient Boosting Regression.

3.21 Decision Tree

A decision tree is a tree-based model. It uses splitting criteria to divide the features into smaller sections with comparable response values. The tree diagram is constructed using the divide-and-conquer strategy. The categorical features can be controlled by decision trees with little to no preprocessing.

3.22 Random Forest

Among the supervised learning techniques, Random Forest is a well-known machine learning algorithm; It is based on the idea of ensemble learning and can be applied to both classification and regression issues in machine learning. It combines several classifiers to solve a challenging problem and enhance the model's performance.

3.23 Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. It executes a regression operation. Regression uses independent variables to model a goal prediction value. It is mostly used to determine how variables and forecasting relate to one another. Regression models vary according to the number of independent variables they use and the type of relationship they take into account between the dependent and independent variables. The dependent variable in regression has many different names. It can be referred to as a regress and, endogenous variable, criteria variable, or outcome variable. The independent variables may also be referred to as predictor variables, regressors, or exogenous variables [1] [9].

3.24 Gradient Boosting Regression.

A machine learning method called gradient boosting is used, among other things, for classification and regression tasks. It provides a prediction model in the form of an ensemble of decision trees like weak prediction models. The resulting technique, known as gradient-boosted trees, typically beats random forest when a decision tree is a weak learner.

The construction of a gradient-boosted trees model follows the same stage-wise process as previous boosting techniques, but it generalizes other techniques by enabling the optimization of any differentiable loss function.

3.25 Working Procedure

To finish the working procedure, we adhere to specific steps. The actions are:

i. Data Gathering.
ii. Data Analysis.
iii. Data Processing.
iv. Data Partitioning.
v. Apply Machine Learning Algorithms.
vi. Accuracy Assessment.

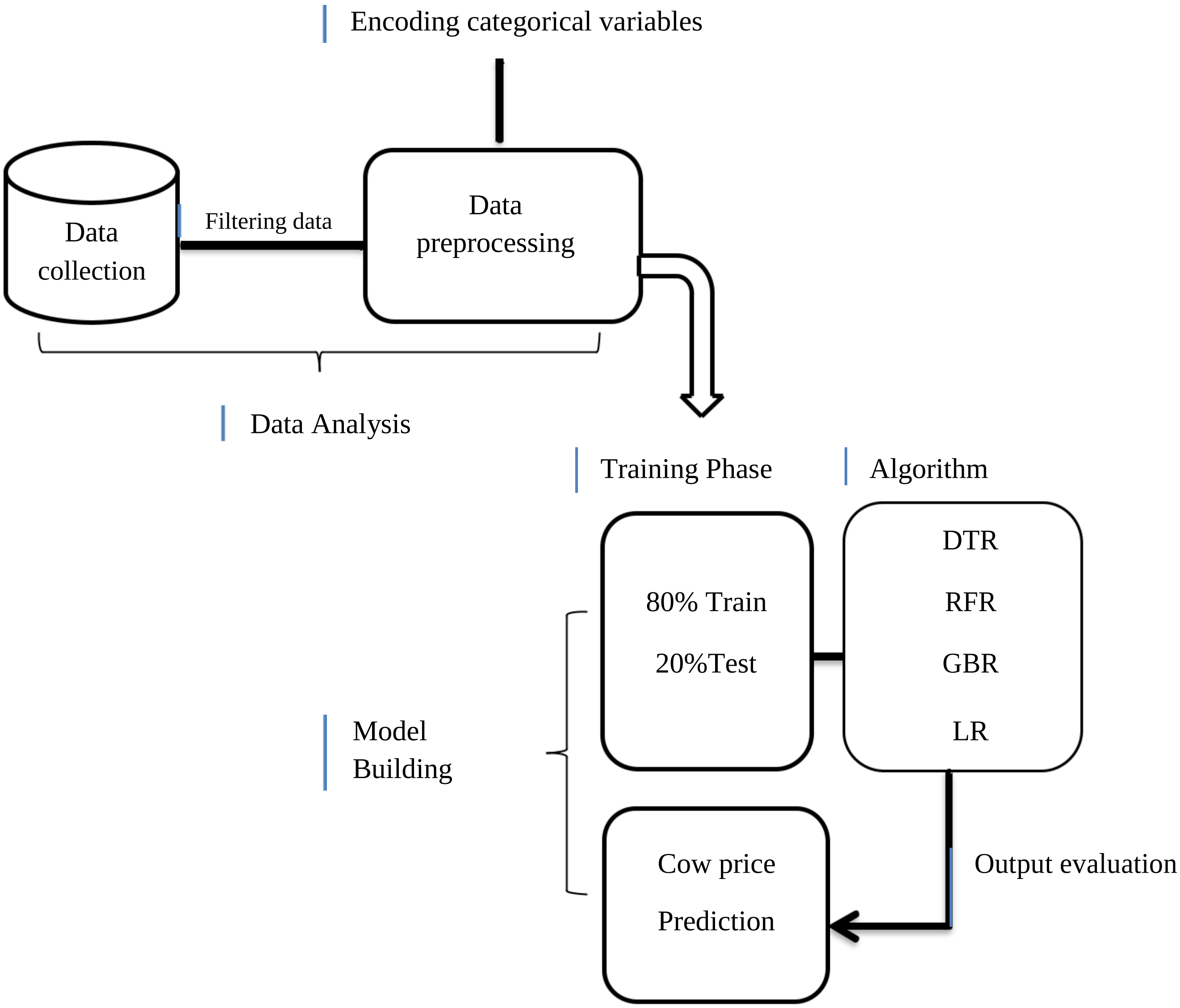

Figure 3.25: Working Process of the proposed model

3.26 Data Gathering and Data Analysis

We gather unprocessed data from the field and cow traders. We begin by gathering data on paper, which we then upload into Microsoft Excel. The phase that required data collection from the field was the most challenging

3.27 Data Pre-processing

Since we have collected our data from the field, we need to preprocess the dataset so that the algorithms can provide better outputs.

The first and most crucial step in creating a machine learning model is data preprocessing, which entails getting the raw data ready and making it suitable for the model because real-world data is in an improper format or has missing values or noise, it can't be used straight in machine learning models. In order to clean data and make it usable for a machine-learning model, data preparation is an essential task [5].

3.28 Data Partitioning

A data set is divided into two, sometimes three, smaller data sets using partitioning. These three are referred to as training, validation, and testing. Using this method to build a predictive model is the best practice, but it can only be done with enough data.

3.29 Apply Machine Learning Algorithms.

The accuracy of a dataset depends on which algorithms are we using. So, we choose some appropriate algorithms for this price prediction model. They are Decision Tree Regression, Random Tree Regression, Linear Regression, and Gradient Boosting Regression [1] [4].

3.30 Accuracy Assessment

Accuracy is the main concern of a project. Accuracy depends on the algorithms and the dataset. We tried to maximize our accuracy by processing our dataset.

CHAPTER FOUR

RESULT AND DISCUSSION

4.1 Introduction

This study covers the price prediction model of cows. There were 1000 total pieces of data collected. The primary sources of data were cow markets or cow traders. We used four machine-learning algorithms to find out the best approaches to predict better price accuracy. Finding the most effective method for this task required an in-depth investigation of many methods which was a critical component of this study.

4.2 Result and Analysis

We used Gradient Boosting Regressor, Decision Tree Regressor, Random Forest Regression, and Linear Regression in a dataset. Results are shown in figure 4.2.

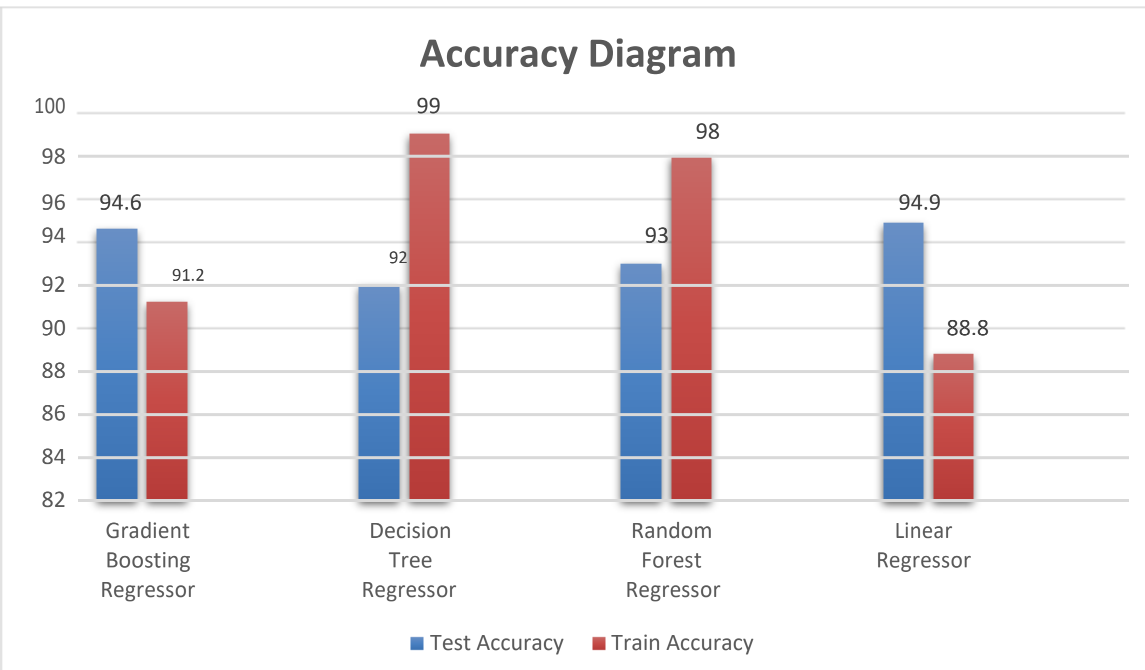

Figure 4.2: Accuracy Diagram

Figure 4.2 highlights that Linear Regression, outperforms Decision Tree Regressor, Gradient Boosting Regression, and Random Forest Regressor by a small margin.

Gradient Boosting regressor with 94.6% accuracy, Decision Tree with 92%, Random Forest Regressor with 93% accuracy, and Linear Regression with the highest accuracy of 94.9%.

TABLE 4.2: ERROR SCORE

Algorithms	**MAE**	**MSE**	**RMSE**
Gradient Boosting Regression	6848.926	238129307.615	15431.438
Decision Tree Regression	11530.0	2107847500.0	45911.300
Random Forest Regression	7936.416	484447745.222	22010.173
Linear Regression	6822.152	223943251.320	14964.733

The performance of the algorithms was assessed using the three most popular accuracy measures for regression models: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Relative Mean Square Error (RMSE). By measuring the, the RMSE highlights the occurrence of outliers. variations between estimates and actuals. We used RMSE as a critical, essential measure to assess the performance of machine learning models. The MSE calculates a prediction approach's degree of accuracy. It evaluates the average square root of the discrepancy between the actual and expected values. The MSE is equal to zero when a model is error-free. It can be contrasted with models whose mistakes are computed in different units. Although MAE calculates the absolute mean difference between the actual and expected data, it does not permit appreciable error rates. The squared mean difference between the estimated and real data is calculated via MSE.

4.3 Comparison with Other Works

TABLE 4.3: PROPOSED MODEL COMPARISON WITH OTHER WORKS

Authors	Animal(deal with)	Descriptions	Size of data sets	Algorithm/Method	Accuracy
This work	Cow	Cow price Prediction	1000	LR	94.9%
Rahman et al. [1]	Cow	Prediction of cow prices using machine learning	1000	CNN and Image Processing	70%
Lawrence et al. [2]	Cattle	Live cattle futures and the seasonal index both predict cattle prices correctly	NM	The Index forecast method	NM
Bozic et al. [3]	Cattle	Rating Livestock Margin Insurance for Dairy Cattle Using Parametric Bootstrap Tests for Futures Price and Implied Volatility Biases	500	The LGM-Dairy rating method	21%

Eldridge et al. [4]	Cattle	Cattle price predictions and grazing management	508	Linear regression model, Feeder Cattle Price Model	90%
Kinnischtzke et al. [5]	Cattle	Trends in Cattle Price Prediction	1100	Logistic regression, Binary Choice Models, Linear Probability model	83.7%
Muwanga et al. [6]	Cattle	Spatial linkages, price distributions, and price forecasts for cattle	800	VAR Model, ECG Model, Granger-Causality Error Correction Model, SGC Model, Hypotheses, and Empirical Price Dispersion Models	90%
Marsh et al. [7]	Cattle	Prices for Quarterly Live Cattle Based on a Rational Distributed Lag Model	NM	Distributed Lag Model	NM
Franzmann et al. [8]	Cattle	Feeder, Slaughter, and Wholesale Beef Cattle Price Trend Models	NM	The linear trend model	NM

Coffey et al. [9]	Cattle	Changes in Market Fundamentals and Price Momentum and the Effects on Hedging Live Cattle	205	LMR	NM
Moser et al. [10]	Bulls and Cows	Accuracy of direct genomic data in Holstein bulls and cows utilizing subsets of SNP markers	2654	Ridge regression (RR), Partial least squares regression (PLSR)	90%
Blakely et al. [11]	Cattle	A model for forecasting quarterly cattle prices with application to the development of a futures market strategy	NM	The forecasting model, The Moving Average Model, Transfer Function Models	81%

4.4 Discussion

For analyzing the price prediction of cows we have gathered sixteen features of 1000 cows. In this research, we have divided our dataset into two parts. Training & testing. 80% data we have used in training & 20% data we have used in testing. We have used four different regressors in this research. Applying all of these four models we get 94.6% accuracy in the gradient boosting regressor, 92% accuracy in the decision tree regressor.93% accuracy in the random forest regressor, 94.9% accuracy in the linear regressor. Among all of these four models, the linear regressor has given the best accuracy. Then we evaluate regression models using Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

CHAPTER 5

SUMMARY, CONCLUSION & FUTURE RESEARCH

5.1 Summary of the Study

Cows are a very important part of our daily lives. We must buy cows in the livestock market. To do that, we need to be aware of the cows' anticipated prices. Our study is based on the estimation of cow prices. 1000 data have been gathered. Following that, the dataset is processed for system input. After cleaning the data, we enter it into the system to be trained, and the result is a price prediction for cows. We use four classic machine learning algorithms gradient boosting regressor, decision tree regressor, random forest regressor, and linear regression & were employed them in the model-building procedure. Data gathering, Data preprocessing, methodology implementation, statistical analysis, data analysis & many others are the sections of our work. We gathered all of the essential information.

5.2 Conclusion

Our Bachelor's degree curriculum includes this research as one of its components. We didn't have a lot of knowledge about machine learning and how it is used in prediction sites when we first started our project. We discovered four machine-learning concepts while conducting this research, and we continue to learn more about them. The price of the cow is predicted using a variety of cow varieties in our article. To forecast cow prices, we employ four traditional machine-learning algorithms. Out of everyone gradient boosting regressor & linear regressor was the model that performed best.

5.3 Limitations

For the purpose of predicting the price of the cow, we have used machine learning. We are constrained in our work and model. Every time we go to collect data, we run into issues. Although a larger dataset would have been more practical, we only employed a small dataset for our research. Using a larger dataset will give a more accurate accuracy to predict the price. There are a lot of male data compared to female data in our dataset. More variety in the data set can provide more accurate data. So, there is a scope for gathering more female data. Predicting cow prices with image processing would be a great option.

5.4 Future Research

In our modern technology machine learning makes our life very faster, easier & more comfortable in every sector of human life. In the future, we will try to extend our dataset. Machine learning works extremely very well with large amounts of data, a large amount of dataset should be collected. Other machine-learning techniques can be utilized for creating a model that is more efficient & intelligent. We hope our research will help people to know the price of a cow & it will help people to purchase a cow easily.

Appendix

Abbreviation

GBR=Gradient Boosting Regressor.

DTR=Decision Tree Regressor.

RFR=Random Forest Regressor.

LR=Linear Regression.

MSE= with Mean Square Error

RMSE = Root Mean Square Error.

MAE= Mean Absolute Error.

References

[1] Rahman and A. Md, "A Machine Learning-Based Price Prediction for Cows," *A Machine Learning-Based Price Prediction,* vol. 5, no. 1, pp. 64-69, May 2021.

[2] J. Lawrence and C. Hoffman, "Cattle Price Forecast Errors: Live Cattle Futures & Seasonal Index," *Live Cattle Futures & Seasonal Index,* vol. 1, no. 1, pp. 2000-2008, 1999.

[3] M. Bozic and J. Newton, "Parametric Bootstrap Tests for Futures Price and Implied Volatility Biases with Application to Rating Livestock Margin Insurance for Dairy Cattle," *Livestock Margin Insurance,* vol. 1, no. 1, p. 29, 2012.

[4] Eldridge and W. R. , "Kentucky feeder cattle price analysis: Models for price predictions and grazing management," *Price predictions and grazing management,* vol. 1, no. 1, p. 103, 2005.

[5] Kinnischtzke and N. J., "Forecasting cattle price trends," *Price trends,* vol. 1, no. 1, p. 24, 1988.

[6] Muwanga and S. G., "Cattle price dispersions, spatial relationships, and price predictions: A case study of selected United States cattle markets," *Doctoral dissertation,* vol. 1, no. 1, p. 24, 1998.

[7] Marsh and M. J., "A Rational Distributed Lag Model of Quarterly Live Cattle Prices," *American Journal of Agricultural Economics,* vol. 63, no. 3, pp. 539-547, 1983.

[8] Franzmann and R. J., "Trend Models of Feeder, Slaughter, and Wholesale Beef Cattle Prices," *American Journal of Agricultural Economics,* vol. 54, no. 3, pp. 507-512, 1972.

[9] Coffey and K. B. , "Impacts of Changes in Market Fundamentals and Price Momentum on Hedging Live Cattle," *Agricultural and Resource Economics,* vol. 1, no. 1, pp. 18-33, 2018.

[10] Moser and G., "Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers," *Genetics Selection Evolution,* vol. 42, no. 1, pp. 1-15, 2010.

[11] Blakely and K. Paul , "A quarterly feeder cattle price forecasting model with application towards the development of a futures market strategy," *Doctoral dissertation,* vol. 1, no. 1, p. 134, 1976.

Plagiarism Report

ORIGINALITY REPORT

17%	15%	9%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ageconsearch.umn.edu Internet Source	2%
2	www.extension.iastate.edu Internet Source	2%
3	www.mdpi.com Internet Source	2%
4	gsejournal.biomedcentral.com Internet Source	1%
5	vtechworks.lib.vt.edu Internet Source	1%
6	www.uky.edu Internet Source	1%
7	medium.com Internet Source	1%
8	academic.oup.com Internet Source	1%
9	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%