Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 2, November : 2023

A MACHINE LEARNING BASED APPROACH TO STUDY EFFECT OF VARIOUS PARAMETERS ON TOMOTO CROP YIELD

Raghu Kumar Bandlamudi[a] Prasad V Potluri Siddhartha Institute of Technology, Kanuru, Vijayawada
Lakshmana Rao Muppa[b] Madanapalle Institute of Technology & Science, Madanapalle
Sasidhar Muttuneni[c] **Chirnajivi Paritala**[c] Amritasai Institute of Science and Technology, Paritala, Vijayawada

ABSTRACT

In India, farming is one of the main sources of employment. The agricultural industry supports more than half of the population, Over 55% of India's population depends on agriculture for existence making it the foundation of the country's economy. Variations in weather, climate, and other environmental , economical factors are now a significant threat to the continued success of agriculture. The study aims at Analyzed the effect of variables (Temperature, Rainfall, Investment, pesticides and fertilizers) on crop yield prediction involves estimating the crop's production based on past data. Using through multi linear regression algorithm

Using this algorithm, we can predict the productivity by the given parameters like temperature, rainfall, investment, pesticides and fertilizers. This all parameters will help to predict the crop productivity and it will help the farmers to make plan and precautions according to the prediction by doing less investments. Validation has been done successfully and obtained good agreement between the prediction as well as the given data.

Keywords: Crop, Yield, Prediction, Tomato, Regression, Machine learning

INTRODUCTION

One of the most popular and consumed veggies in the world, tomatoes are astaplein many different cuisines all over the world. Due to their high nutritional content, favorable health effects, and growing popularity in the food business, tomatoes have seen an increase in output in recent years. They are utilized in a wide range of culinary recipes and are a vital source of vitamins and minerals. However, due to a variety of circumstances, including weather, disease outbreaks, and pests, the amount of tomatoes grown might fluctuate from year to year. In several nations around the world, the tomato business is a significant component to the agricultural sector. To help them to plan their operations and choose their crop wisely, tomato farmers must make a prediction of the crop's volume for the future year. To assess tomato data from previous years and produce precise forecasts about the volume of tomatoes for the following year and apply machine learning algorithms.

One of the exclusion criteria used in our analysis of the retrieved articles is whether the publication is a survey or standard review paper. These omitted works are really related works, which are covered in this section. Review research on the estimate of nitrogen status using machine learning was carried out by Culinarian and Sukkarieh (Chlingaryanetal.,2018).The report comes to the conclusion that the agriculture industry would benefit from fast advancements in ML and sensing technology. A review of papers on machine learning models for agricultural production prediction based on meteorological factors was conducted by Elavarasan etal. In order to uncover more factors that affect agricultural output, the research [5, 9] suggests searching broadly. A review paper on the use of machine learning in agriculture was released in 2018 [1]. The analysis was done utilizing papers that dealt with soil management, water management, crop management, and animal management. In order to choose the best time for harvest and estimate production, [3] conducted review research on assessing the maturity of fruits. [4, 11] Discussed the problems and approaches that are faced in the agricultural industry when using image processing and machine learning, particularly when trying to detect illnesses. The authors of paper [6] discussed several machine

learning techniques and how they applied to plant biology. A review study on the use of data mining for decision-making in the agriculture industry generally is discussed in [7, 10]. They came to the conclusion that more study is required to determine how data mining may be implemented into complicated agricultural datasets. In another review paper [8, 12] of the various data mining methods utilised for agricultural production prediction and came to the conclusion that data mining methods might be employed to resolve the problem

MULTILINEAR REGRESSION METHOD IN MACHINE LEARNING

Multi linear regression: To calculate the association between two or more independent variables and one dependent variable, utilize multiple linear regression.

The degree to which two or more independent factors and one dependent variable are correlated (for instance, how rainfall, temperature, and the amount of additional fertilizer affect crop growth).The value of the dependent variable at a particular value of the independent variables (for instance, the anticipated crop output at a specific level of rainfall, temperature, and fertilizer addition).

Once we are ready with collected the data, we can use a multi-linear regression model to predict the future tomato crop yield and quantity based on the farmer's data. The multi linear regression model is used on the farmer's data as input variables and the process parameters of tomato crop yield in quantity as the output variable.

To train the multi-linear regression model, you would need to use awe used a data set that includes both the farmer's data and the process parameter i,e; tomato crop yield in quantity. The model would then be optimized to minimize the difference between the predicted tomato crop yield in quantity and the actual tomato crop yield in quantity.

The model will take the farmer's data as input variables and use the coefficients to calculate the predicted tomato crop yield in quantity.

Multiple Linear Regression

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\beta_3X_{i3}+\beta_4X_{i4}+ \ldots\ldots\ldots\ldots +\beta_pX_{ip}+\epsilon$$

Where:

- Y_i=dependent or predicted variable
- β_0=y-intercept
- β_1 and β_2 are the regression coefficients representing the change in y relative to a one-unit change in x_{i1} and x_{i2}, respectively.
- ϵ = random error (residual)term.

Multi linear Regression Algorithm steps

Step 1: Import the necessary packages

The necessary packages such as pandas, NumPy, sklearn etc are imported**.**

Step 2: Import the CSV file:

The CSV file is imported using pd. read_csv ()method. To access the CSV file, click here. The 'No 'column is dropped as an index is already present.df. head () method is used to retrieve the first five rows of the data frame.df. columns attribute returns the name of the columns. The column names starting with 'X' are the independent features in our dataset. The column 'Y' tomato crop yield in quantity is the dependent variable column. As the number of independent or exploratory variables is more than one, it is a Multi linear regression.

Step 3: Create a scatter plot visualize the data:

A scatter plot is created to visualize the relation between the 'X4 number of convenience stores' independent variable and the 'Y 'tomato crop yield in quantity a dependent feature.

Step4: Create feature variables:

To model the data, we need to create feature variables, X variable contains independent variables and y variable contains a dependent variable.

Step 5: Split data into train and test sets:
Here, train_test_split () method is used to create train and test sets, the feature variables are passed in the method. test size is given as 0.3,which means30% of the data goes into test sets, and train set data contains 70% data. The random state is given for data reproducibility.
Step 6: Create a linear regression model
A simple linear regression model is created. Linear Regression () class is used to create a simple regression model, the class is imported from sklearn.linear model package**.**
Step 7: Fit the model with training data.
After creating the model, it fits with the training data. The model gains knowledge about the statistics of the training model. fit () method is used to fit the data.
Step 8: Make predictions on the test data set.
In this model. Predict () method is used to make predictions on the Test data, as test data is unseen data and the model has no knowledge about the statistics of the test set.
Step 9: Evaluate the model with metrics.
The multi-linear regression model is evaluated with mean_squared_error and mean_absolute_error metric. when compared with the mean of the target variable, we'll under stand how well our model isp redicting.mean_squared_error is the mean of the sum of residuals. mean_absolute_erroris the mean of the absolute errors of the model. The less the error, the better the model performance.

Methodology
Figure 1 shows the methodology for predicting using multi linear regression. The methodology for predicting, using machine learning would involve the following steps:
Data Collection: Collect the data on tomato quantity from the farmer's data per one-acre land. This can be obtained from Madanapalle and beside villages.
Suitable Method: Trains a model on known input and output data to predict future outputs, and machine learning, which uses hidden patterns or internal structures in the input data.
Multi linear regression: After analyzing the data, we used a statistical technique like multi linear regression to predict future crop yield. This involves creating a mathematical model that relates the independent variables to crop yield. The model is then used to make predictions based on new data.
Multi linear regression algorithm: Which is a part of AI, uses an assortment of accurate, probabilistic, and upgraded techniques that empower computers to pick up from the past point of reference and perceive hard-to-perceive patterns from massive, noisy, or complex datasets.
Solution: To manually predict future crop yield, one need to follow the above steps and perform the calculations and analysis by hand.
Optimum Solution: Solution is one whose measure of quality is close to the best that could theoretically be obtained. Typically, agents do not need optimal solutions to problems; they only must get close enough.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 2, November : 2023

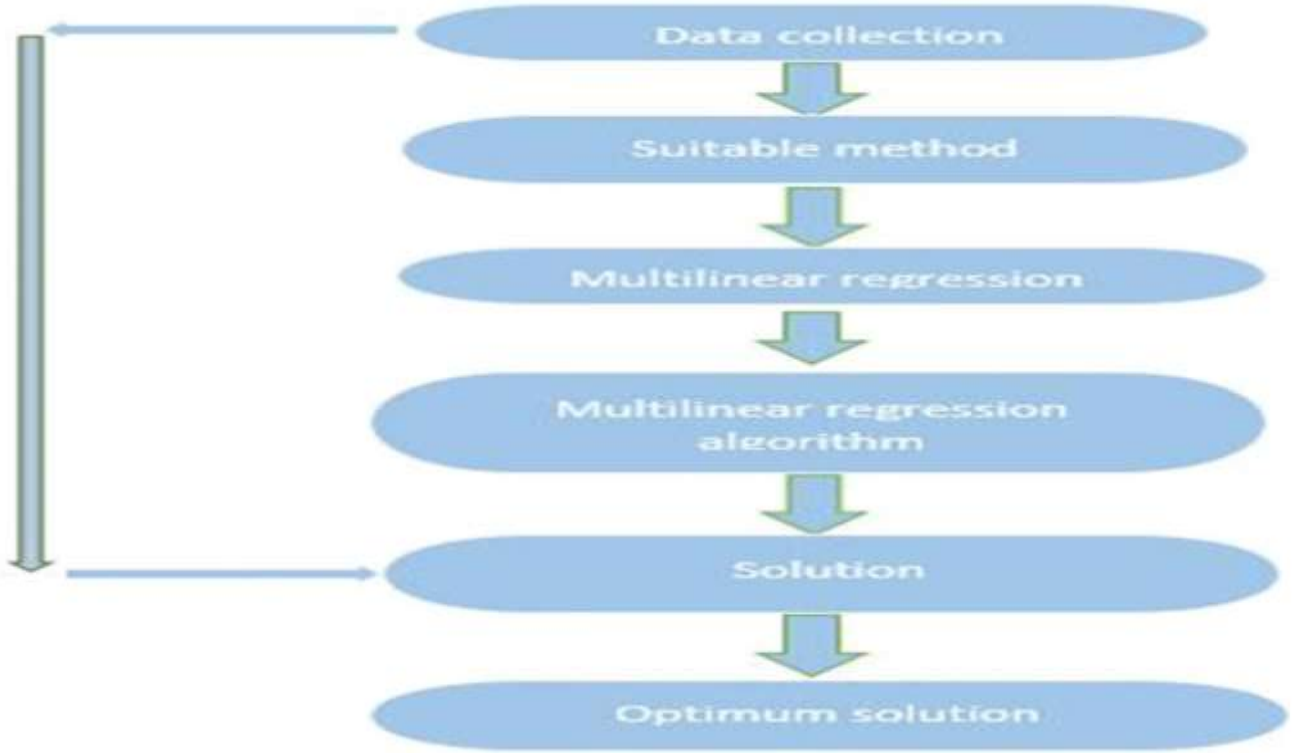

Figure 1 The methodology

DATA COLLECTION: We collected the following data from formers

Table : Data collection from Formers

S.NO	Production (Tons)	Temperature (degree)	Rainfall (cms)	Investment (Rs)	Fertilizers (mm)	Pesticides (L)
1	5	30	20	50000	65	50
2	4.5	20	50	55000	135	45
3	5.5	32	45	20000	85	55
4	7.5	20	80	80000	165	75
5	6.5	38	70	70000	155	65
6	7	39	15	45000	115	70
7	6.5	38	70	70000	130	65
8	3.5	25	40	45000	125	35
9	4.5	28	9	55000	135	45
10	5.5	12	55	65000	39	55
11	4	27	48	50000	120	40
12	5	30	60	20000	140	50
13	5.5	32	55	65000	160	55
14	6	16	55	45000	120	60
15	3.5	25	40	15000	120	35
16	3	22	35	35000	120	23
17	5.5	32	21	65000	145	55
18	2	20	20	10000	90	20
19	8	45	42	90000	170	20
20	8.5	50	90	100000	120	85
21	4.5	28	45	20000	135	39
22	4	27	53	50000	130	40
23	6.5	38	70	15000	140	25
24	4.5	28	45	55000	135	45

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 2, November : 2023

MULTI LINEAR REGRESSION CODE

```
import matplotlib.pyplot as plt
from scipy import stats
plt.scatter(df["Temperature?"],df["Productionin(Kg)"])
```

```
<matplotlib.collections.PathCollection at 0x218990aa640>
```

```
x=df[["Temperature?","Rainfall(mm)","Investement(Rs)","Fertilizers(kg)","Pesticides(L)"]]
y=df["Productionin(Kg)"]
#print(x,y)
from sklearn import linear_model as ml
re=ml.LinearRegression()
re.fit(x,y)
p=re.predict([[30,20,50000,65,50]])
print(p)
```

```
[4.86271354]
```

RESULTS AND DISCUSSION

The expression is:

Production = $8.8 \times 10^{-2} + 7.12 \times 10^{-2}$ * Temperature + 1.61×10^{-2} * Rainfall + 1.83×10^{-5} * Investment + 7.24×10^{-4} *Fertilizer + 2.82×10^{-2} * Pesticides

And corresponding R^2 value = 0.73 (May be the expression is true as R^2 is close to 1)

And F static is 9.734 > critical value (0.9407 for 18 dof)

From above we can say that the expression is may be consider

Crop production analysis has been done through the multi linear regression analysis, Effect of temperature, rainfall, pesticides, fertilizer and investment on production analyzed through graphs. Effective variables temperature, rainfall, pesticides, and fertilizer increases at certain level crop yield

decreases. So this analysis to be helpful for the farmers obtaining optimum variables through machine learning analysis

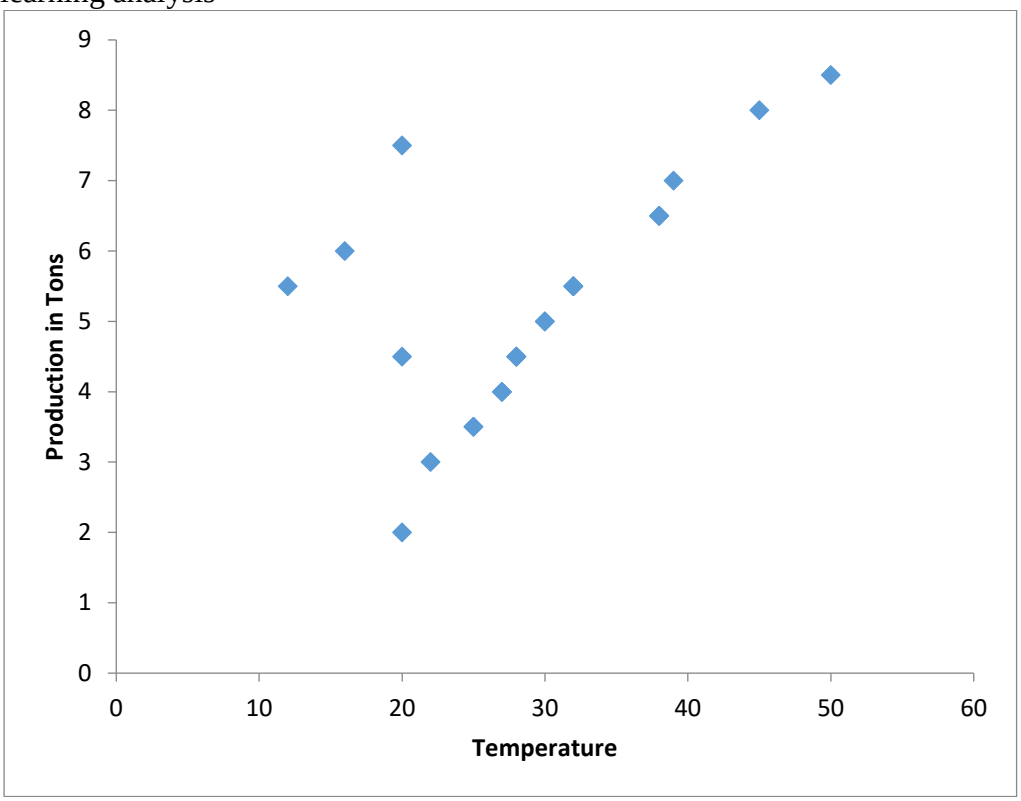

Fig 2 Production vs Temperature

Figure 2 shows relationship between production and temperature. X-axis represents temperature(c) and Y-axis represents production in Kg. Here production is increasing with increasing temperature upto certain value, after that clearly we can see a drop in production

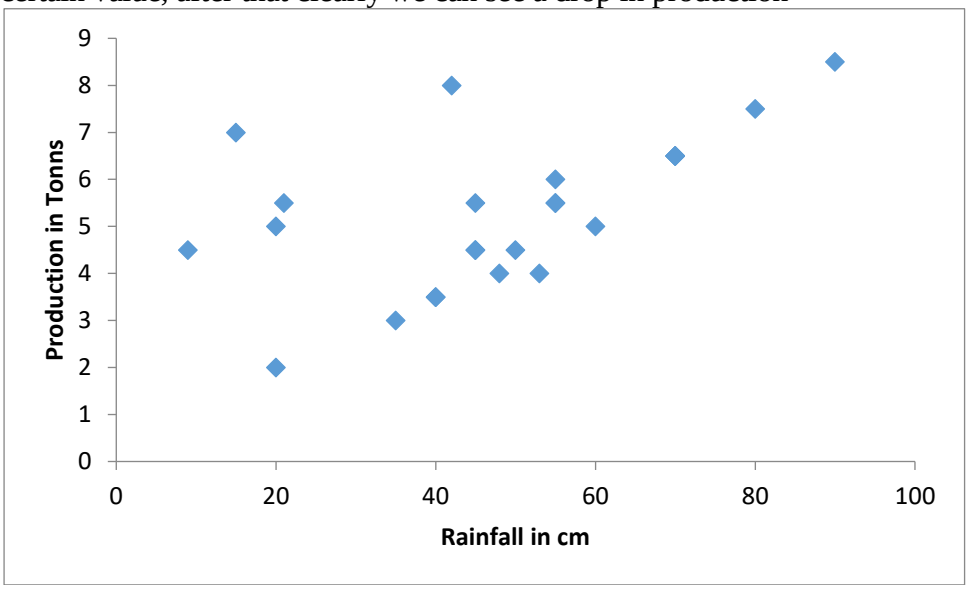

Figure 3 Production vs Rain fall

Figure 3 shows the relationship between production vs rainfall. X-axis represents Rainfall (mm) and Y-axis represents the Production. Here production increases with rainfall upto certain value, and beyond that production is decreasing

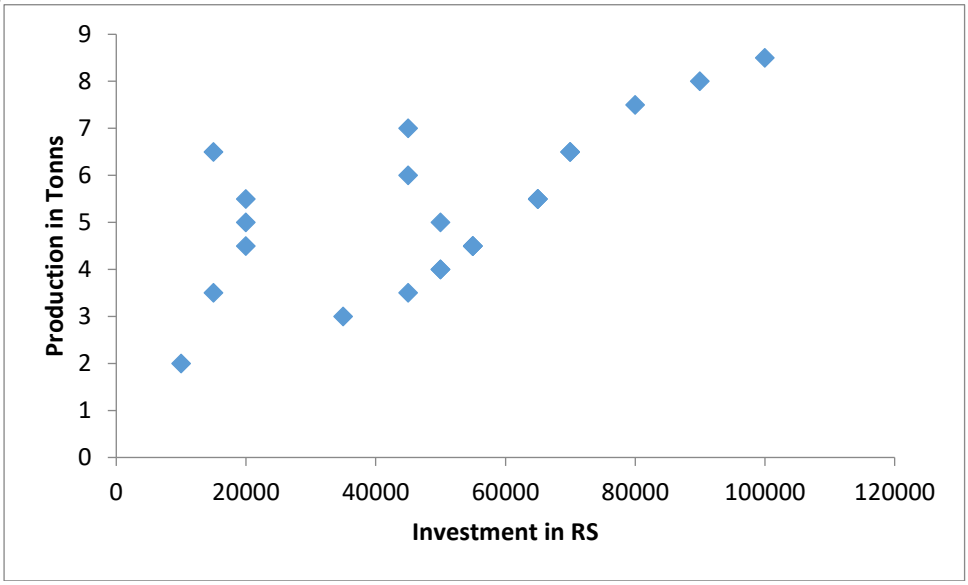

Figure 4 Production vs Investment

Fig. 4 shows the relationship between production and investment. X-axis represents investment and Y-axis represents production. Here production increases with investment

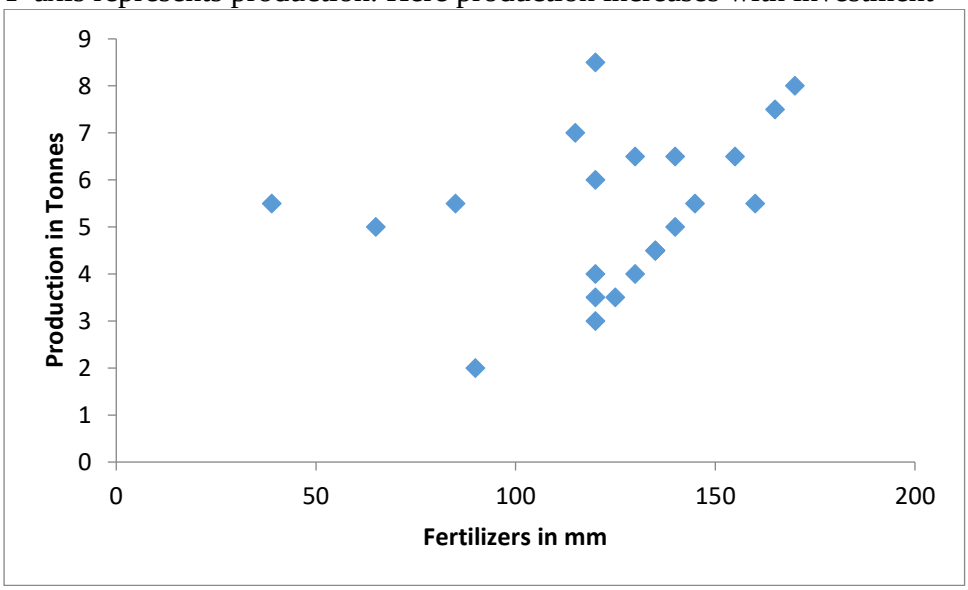

Figure 5 Production vs Fertilizers

Figure 5 shows the relationship between production and fertilizer, X-axis represents fertilizer, Y-axis represents the Production. Here fertilizer production increases with fertilizers

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 2, November : 2023

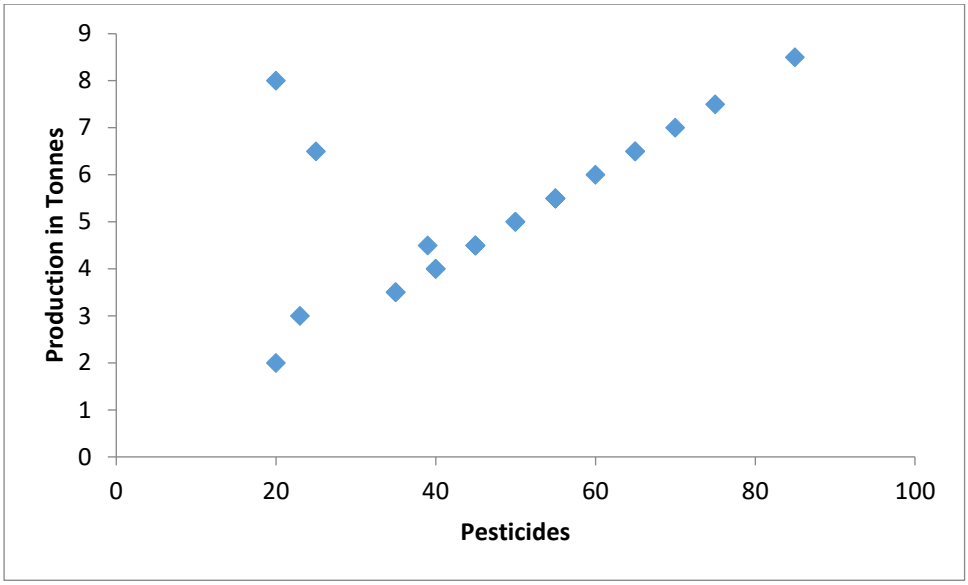

Figure 6 Production vs Pesticides

Figure 6 show the relationship between production and pesticides. X-axis represents pesticides and Y-axis represents Production. Here production increases with pesticides up to certain level,

CONCLUSIONS

Multi linear regression algorithm applied successfully for analyzing the variables on crop production. This analysis is helpful to the farmers for finding suitable options. The tomoto crop increase with temperature, rainfall, fertilizers, pesticides upto certain respective limits and increases with investment.

REFERENCES

1. (Shidnal S, 2021 Oct 13) Shidnal S, Latte MV, Kapoor A. Crop yield prediction:two-tiered machine earning model approach. International Journal of Information Technology.2021Oct; 13:1983-91.
2. (Afonso M, 2020 Nov 19) Afonso M, Fonteijn H, Fiorentin FS, LensinkD, Mooij M, Faber N, Polder G, Wehrens R. Tomato fruit detection and counting in green houses using deeplearning. Frontiers in plant science.2020Nov19;11:571299
3. (DarwinB,2021Mar27) DarwinB, Dharmaraj P,PrinceS, PopescuDE, HemanthDJ. Recognition of bloom/ yield in crop images using deeplearning models for smart agriculture: a review. Agronomy.2021Mar27;11(4):646
4. (Alibabaei K, 2021 May2) Alibabaei K, Gaspar P D, Lima T M. Crop yield estimation using deep learning based on climate big data and irrigation scheduling. Energies.2021May22;14(11):3004.
5. (Elavarasan D, 2018 Dec1) Elavarasan D,Vincent DR, Sharma V, Zomaya AY, Srinivasan K. Forecasting yield by integrating agrarian factors and machine learning models : A survey. Computers an delectronics in agriculture. 2018 Dec 1; 155:257-82

6 (Gong L, 2021 Jul 1). Gong L, Yu M, Jiang S, Cutsuridis V, Pearson S.DeeplearningbasedpredictionongreenhousecropyieldcombinedTCNandRNN.Sensors.2021Jul 1;21(13):4537

7. (Malik P, 2021 Jan 28) Malik P, Sengupta S, Jadon JS. Comparative analysis of soil properties to predict fertility and crop yield using machine learning algorithms. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2021 Jan 28
8. (Mu Y, 2020 May 25;) Mu Y, Chen TS, Ninomiya S, Guo W. Intact detection of highly occlude dimmature tomatoes on plants using deep learning techniques. Sensors. 2020 May 25;20(10): 2984
9. (Reddy DJ, 2021 May 6) Reddy DJ, Kumar MR. Crop yield prediction using machine learning algorithm. In2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) 2021 May 6 (pp.1466-1470).IEEE.
10. (MF, n.d.) Suescún MF. Machine learning approaches for tomato crop yield prediction in precision agriculture (Doctoral dissertation, Universidade NOVA de Lisboa(Portugal))
11. (Tatsumi K, 2021 Dec;17)Tatsumi K, Igarashi N, Mengxue X. Prediction of plant-level tomato biomass and yield using machine learning with unmanned aerial vehicle imagery. Plant Methods.2021Dec;17(1):1-7.
12. (VermaMS,2021Mar25) Verma MS, Gawade SD. A machine learning approach for prediction system and analysis of nutrients uptake for better crop growth in the Hydroponics system. In2021 international conference on artificial intelligence and smart systems(ICAIS)2021Mar25 (pp.)