



# MACHINE LEARNING MODEL TO PREDICT FARMER'S INCOME

<sup>1</sup> Anoushka Kashyap

<sup>1</sup>Data Scientist,

<sup>1</sup>Artificial Intelligence and Machine Learning,

<sup>1</sup>TuringMinds.ai, Bangalore, India

**Abstract :** India is the second largest populated country in the world. Every individual's work contributes in shaping the country. Everyone working has the equal right to earn for his/her work. Among them are the farmer's without whom human population would cease to extinct. Due to corruption and lack of knowledge, they are deprived of their basic right. This project/model will predict the fair pay and empower them with knowledge of their right.

**IndexTerms –** Decision Tree Algorithm,python,machine learning database,google colabrotory.

## INTRODUCTION

Farmer's are the backbone and pillars of our society. They are the one's who provide us all the food that we eat. The entire population of country depends upon farmer's, be it the largest country or smallest. Without them and agriculture, it would be an ordeal for the modern world and civilization. Yet they do not procure the required wage for their produce. As a result they have to deal with lot of consequences, which they don't deserve for all the hard work they have put in and majority of times it reaches to a condition where they have to give up their lives.

The objective is to make sure that the farmer receives the appropriate pay for his produce and he should gain the profit. This is achieved by applying vegetable price prediction, where the price of the crops are determined based on the type of crop, the month in which it is cultivated and weather condition for that month. The character of the model is predict the price of the crop, among which 70 % of the price will be given to the farmer and 30 % to the middleman. Middleman plays crucial role in the whole process. As he makes sure that the crop he has obtained from farmer reaches every individual through retailers.

## NEED OF THE STUDY.

Today India is progressing towards becoming a developed country. The progress of every individual farmer plays a vital role in this process. Only when farmer's progress, our country will have indeed growth. It is our duty to make sure that the pillars of our country get what they deserve. In order to achieve this, we should contribute in providing them with the most essential right and that is to get legitimate pay for their produce. Because at the end of the day if the farmers are not given the fair pay, then they will gradually start losing the interest in farming and will definitely not encourage their future generations to continue doing farming. It will automatically reduce the farmers population to continue farming and population of country is continuously increasing. This will lead to a situation where the demand and supply chain will never be balanced. Here is when the need of study begins.

The purpose is to make sure that the farmer receives the appropriate pay for his crop and he should receive the profit. This is done by applying vegetable price prediction model using machine learning. Vegetable prices are predicted by considering the crop, the month in which it is cultivated and the weather conditions for that particular month. The character of this model is to predict prices of the vegetable and from the predicted price, 70 % of the price is given to farmer and remaining 30 % is given to the middleman. Middle man is plays crucial role, as he makes sure that the vegetables reach the retail shop.

## 3.1 Population and Sample

The population in India is one forty crore. Among 140 crore, the population of farmers or the population which is dependent on agriculture is 81 crore, that is 58 % of 140 crore. More than 50 % of Indian population is dependent on agriculture. If more than half of the Indian population are not receiving their fair pay, then gradually over the next 10 years the the population will increase and simultaneously the population of farmers will decrease. The supply and demand cycle chain will not be balanced and lead to crisis. A crisis where prices will be very high and yet again lower class of society will have large impact.

The supply and demand chain has to be balanced, that is more the farmers population involved in agriculture more will be the supply only when the farmers are paid the right pay.

### 3.2 Data and Sources of Data

For this study data has been collected from the website of kaggle. Few set of vegetables are considered where the month, the season and the temperature for that particular month have been considered. Also the disaster that took place for that particular month and price per kg for that particular month have been considered.

### 3.3 Theoretical framework

Variables of the study contains dependent and independent variable. The study used pre-specified method for the selection of variables. The study used the values of the prices per kg as dependent variable. From the price of the vegetables the farmers pay are calculated.

### 3.4 Methodology

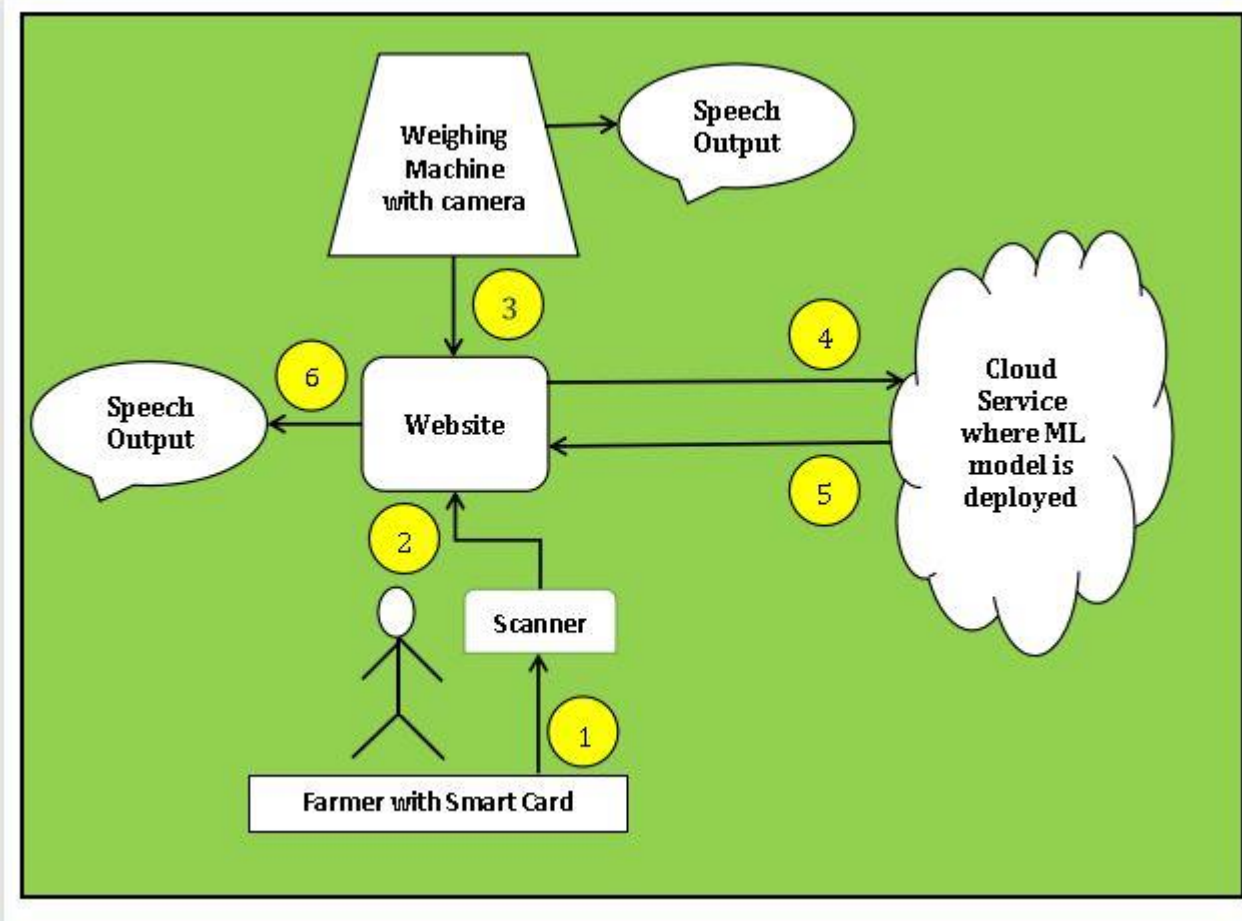


Fig 1 : schematic diagram of the prototype

This prototype will be implemented through a website where the model will be deployed in a free cloud service. The model has to be trained continuously with the details of season in which the crop is grown, temperature, disaster happened in last 3 months, whole sale index for each crop.

- The first step is the farmer will be given the smart card which will contain the unique identification number and aadhar number. The website will have scanning option, where he will scan his card. Once he scans the card his photo, the details of the crop which he cultivates and unique identification number given to him will be displayed. This process will make sure to validate the farmers details and will eliminate the chances of the farmer getting exploited.
- Second step is where the farmer will weigh his cultivated crop using weighing machine and the weight of the crop will be declared to the farmer through voice output. This weighing machine will have camera which will recognize the crop. The image of the crop recognized and the weight of the crop measured will be sent to the model.
- The model will then process the data fed to it and calculate the price of the crop which the farmer should receive with respect to the weight of the crop. The amount that the farmer should be paid will be verified with speech output in the native language. This speech output will minimize the chances of farmers getting exploited.

- Algorithm : Decision Tree Algorithm
- Framework : scikit-learn, Flask
- Software : Python
- Cloud Services : Google Cloud AutoML
- Database: Machine Learning Database (MLDB)
- Web IDE : Google Colaboratory

### 3.5 Statistical tools and econometric models

Decision Tree Classification Algorithm is used to predict the price of the crop based on various attributes. From the price that is predicted, farmers pay will be predicted.

#### 3.5.1 Decision Tree Classification Algorithm

It is a supervised learning technique that can be used for both classification and regression. It is a tree structured classifier, where internal nodes represent the features of dataset, branches represent the decision rules and each leaf node represents the outcome. In a decision tree there are two nodes. Decision nodes are used to make any decisions and have multiple branches. Leaf nodes are the output of those decisions and do not contain any further branches. The decisions are performed on the features of dataset.

##### 3.5.1.1 Classification and Regression Tree (CART) Algorithm

In order to build tree we use CART algorithm. It is a predictive model, it explains how target values can be predicted based on other attributes. In a decision tree root node is split into sub nodes on the basis of threshold value of an attribute. The root node is taken as training set.

The CART uses Gini index. It is a metric for classification tasks. It splits the dataset into decision tree. It does that by searching for the best homogeneity for the sub nodes with the help of Gini index criteria. It works on categorical variables, provides outcomes either successful or failure and hence conducts binary fitting only. It stores the sum of squared probabilities of each class. It computes the degree of probability of a specific variable that is wrongly being classified when chosen randomly and a variation of Gini coefficient. Mathematically we can write Gini index as

$$\text{Gini} = 1 - \sum_{i=1}^n (p_i)^2$$

Where  $p_i$  is the probability of an object being classified to a particular class,  $n$  is the class.

## IV. RESULTS AND DISCUSSION

### 4.1 Visualization of Dataset



	Vegetable	Season	Month	Temp	Deasaster Happen in last 3month	Vegetable condition	Price per kg
0	potato	winter	jan	15	no	fresh	20
1	tomato	winter	jan	15	no	fresh	50
2	peas	winter	jan	15	no	fresh	70
3	pumkin	winter	jan	15	no	fresh	25
4	cucumber	winter	jan	15	no	fresh	20
...	...	...	...	...	...	...	...
116	brinjal	winter	jan	15	yes	fresh	33
117	ginger	winter	jan	15	no	fresh	88
118	potato	summer	apr	32	no	fresh	24
119	peas	summer	apr	33	no	fresh	33
120	radish	summer	apr	32	no	fresh	9

121 rows × 7 columns

Fig 2 : Dataset  
Type equation here.

	Vegetables	Actual Price	Predicted Price	Farmers Pay	Retailers Pay
2	peas	70.0	38.500000	26.950000	11.55
10	garlic	150.0	121.500000	85.050000	36.45
17	ginger	20.0	82.666667	57.866667	24.80
31	califlower	20.0	13.000000	9.100000	3.90
32	chilly	200.0	75.000000	52.500000	22.50
33	okra	15.0	30.000000	21.000000	9.00
35	brinjal	40.0	20.000000	14.000000	6.00
42	peas	40.0	38.500000	26.950000	11.55
44	cucumber	12.0	15.000000	10.500000	4.50
46	Raddish	15.0	19.000000	13.300000	5.70
48	onion	28.0	28.000000	19.600000	8.40
53	chilly	80.0	77.500000	54.250000	23.25
54	okra	40.0	35.000000	24.500000	10.50

Fig 3 : result

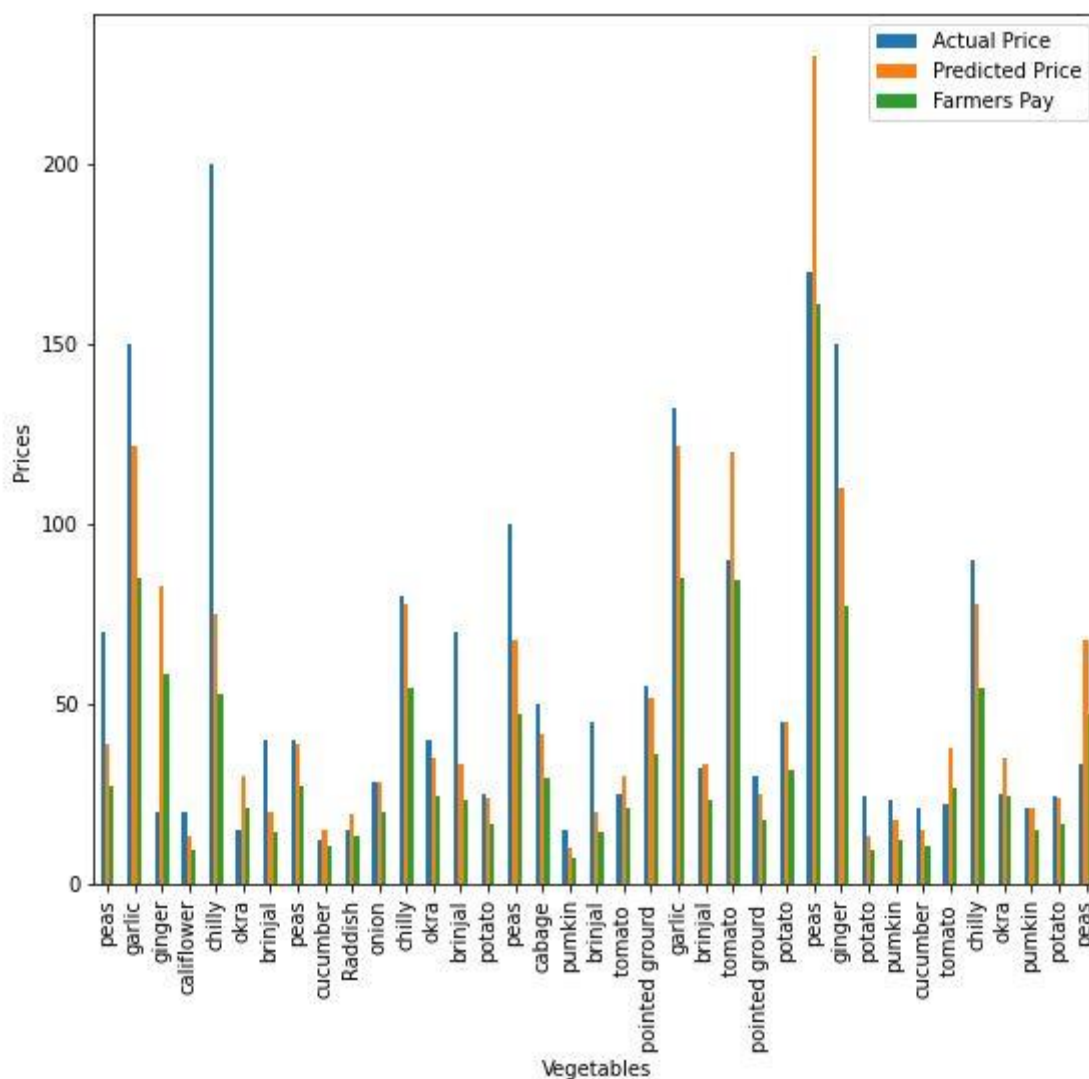


Fig 4 : Actual Pay, Predicted Pay, Farmers Pay



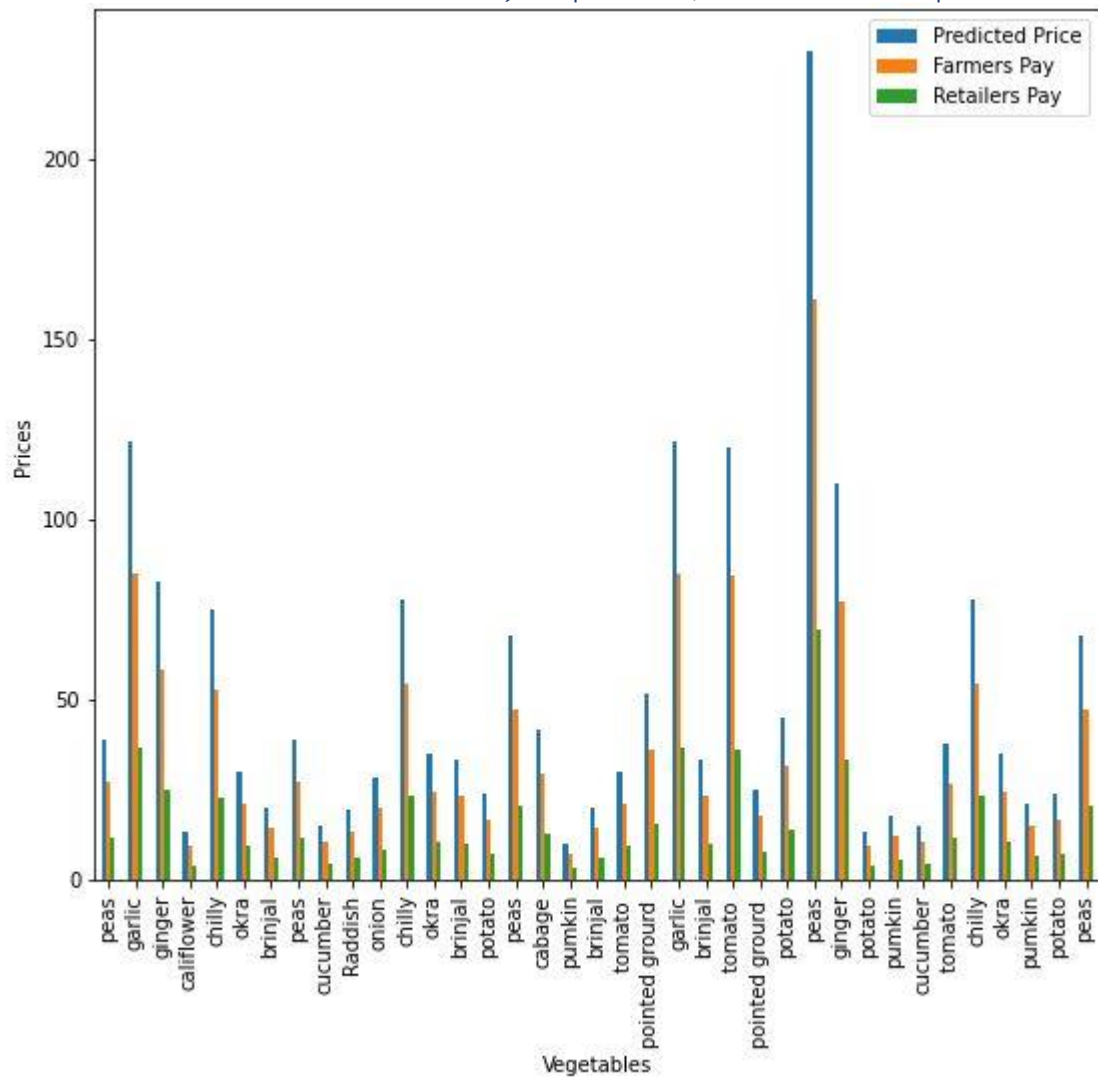


Fig 5 : Predicted Pay, Farmers Pay, Retailers Pay



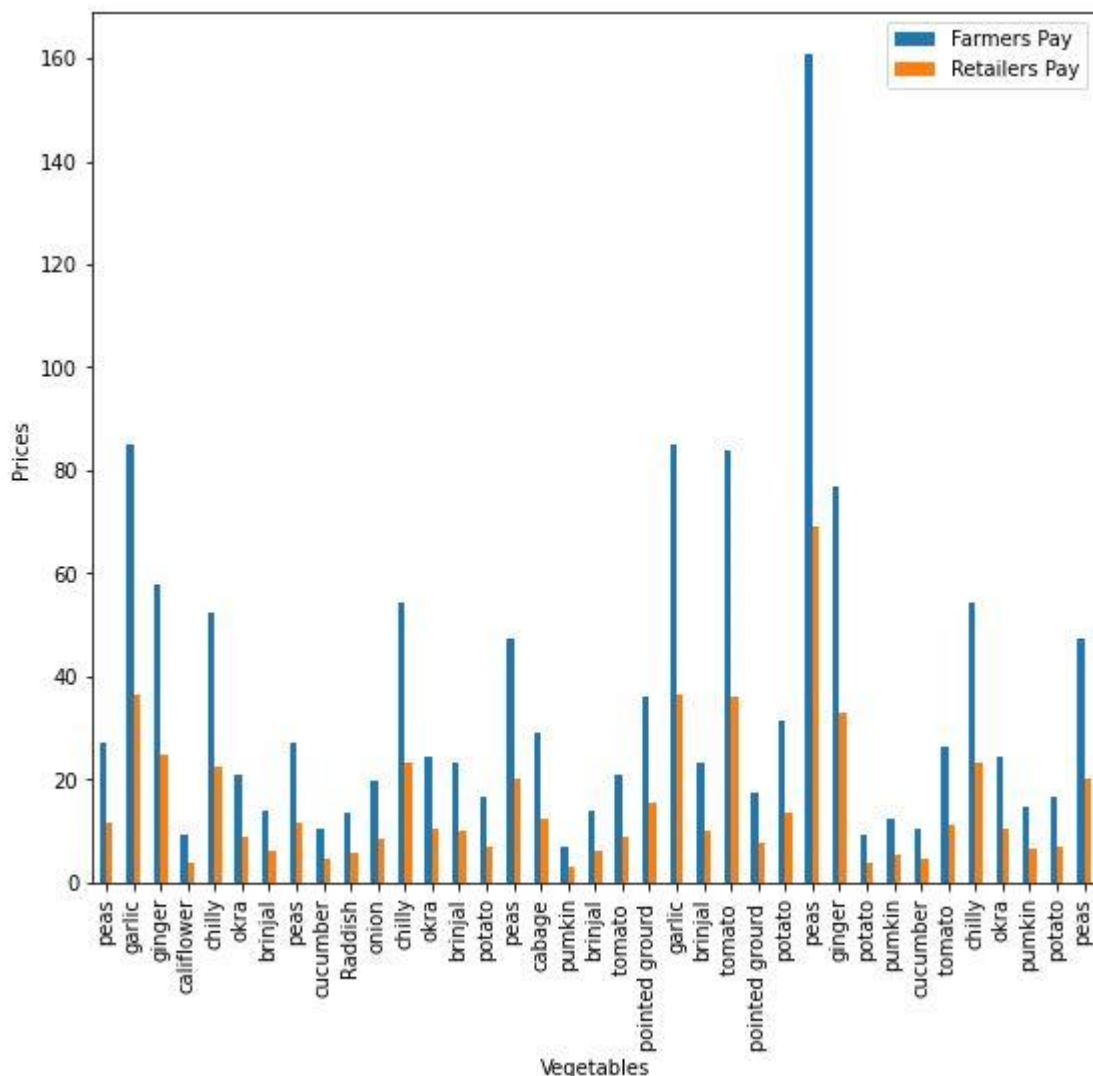


Fig 6 : Farmers Pay,Retailers Pay

**REFERENCES**

- [1] AIGGPA:Crop price prediction using machine learning in Madhya Pradesh: A pilot study,Atal Bihari Vajpayee Institute of Good Governance and Policy Analysis, March 2020.
- [2] Ranjani Dhanapal et al 2021 J Phys.: *Conf.Ser.***1916** 012042.
- [3] CROP PRICE PREDICTION USING MACHINE LEARNING by GS Kakaraparthi.

