Commodity Stock Recommendations and Price Prediction using Prescriptive Analytics Techniques

Dr.Kishor Honwadkar, Vishal S. Wankhede, Harish D. Hatmode, Pragati S. Bankar and Onkar G. Kulkarni.

Smt. Kashibai Navale College of Engineering,

Pune,411041, India

Email: knhonwadkar@sinhgad.edu ;vishalwankhede44@gmail.com; hdhatmode@gmail.com; pragati.bankar1@gmail.com; onksgk@gmail.com

***Abstract***[[1]](#footnote-1) —Prescriptive Analytics is a method which helps businesses to make better decisions by analyzing raw data. Historic raw data can be analyzed using data mining algorithms to get knowledge from that data. Using this knowledge, we can make commodity stock and price forecasting. But, fluctuations in these predictions affects the global economic activity. They have major impact on overall performance. So, we have proposed an efficient method for commodity stock and price predictions to improve the product sales. In order to forecast commodity stock and price, we have used different data mining algorithms like decision tree, K-Nearest Neighbor, gradient boost, linear regression and random forest. Further, we have compared all these algorithms to get best performing algorithm and provide predictions according to that algorithm. This provides more accurate results of stock and price forecasting.

***Index Terms***—Data Mining Algorithms, Decision Tree, KNN, Gradient Boost, Linear Regression, Random Forest, Stock Forecasting, Price Forecasting

##### Introduction

India is the land of agriculture and as per 2010–11 census report, about 70% of its population depends on agriculture for their livelihood. Presently, due to lack of proper facilities and market knowledge, farmers sell their agricultural produce to intermediators at very low cost. These intermediators take advantage of farmer’s ignorance and helplessness. Due to this, economical growth of Indian farmers is deteriorating with growing disinterest in farm and farm investments. Hence, business intelligence (BI) will help farmers to make wise decisions and improve their financial condition.

Business Intelligence (BI) is nothing but a process that includes collecting agricultural data, preprocess it, extract useful insights from it. Prescriptive Analytics is a common function of BI technologies. Using prescriptive analytics techniques, farmers can make decisions like which crop should be selected, whether to sell farm produce now or later, what will be the price of particular crop in future, how much stock should be produced. Prescriptive analytics helps farmers by predicting price of crops and recommending stock of crop to be produce.

The proposed system uses various prescriptive analytics algorithms like decision tree, K-Nearest Neighbor (KNN), gradient boost, linear regression and random forest to predict commodity price and recommend stock. This system compares all these algorithms to get best output.

1. Related Work

Commodity price prediction and stock recommendations are very important for farmers. They should be aware of these predictions so that they can make wise decisions to improve their financial condition [1]. The agriculture related data and weather information are time series in nature. Time series data is a sequence of well-defined data points measured at certain interval of time. The deep learning techniques are suitable for prediction based on these time series data. These techniques discover patterns in time series data and extrapolate them into future values of time series. There are various studies that gives application of deep learning [2] and forecasting [3] of agricultural produce price.

In data mining techniques, information can be transformed into useful format using supervised and unsupervised learning [4]. Effective business decision making can be done using appropriate sales prediction technique [5]. We can use data mining techniques to produce models that are comprehensive and reliable [6].

1. Research Methodology

The main purpose of this research is to get best results for price prediction and stock recommendation by evaluating and analyzing various data mining techniques.

## Data Collection and Preparation

The data about various agricultural commodities is available on Open Government Data platform [7]. The weather information is provided by Agricultural Meteorology Division portal [8].

The inputs to this system are daily market price and weather datasets. The daily market price dataset with data from 2012 to 2017 is taken from Agmarknet. It contains daily market price data of oilseed and pulses. The dataset for Agricultural Commodities contains agricultural commodity name, state, district, market, date of arrival, minimum price of the day, maximum price of the day, average price of the day and total quantity that arrived on the day, as attributes. The weather data has local time, cloudiness, precipitation, pressure, temperature, humidity, wind (speed, gusts, direction), sunrise/sunset, moon rise/moonset and moon phase as attributes.

## Data Preprocessing

Before using raw dataset, which is discussed in previous sub section, data preprocessing techniques are applied on it. To get accurate and quality results, data cleaning is required to be performed on datasets. Hence, we applied data cleaning techniques to remove outliers and for filling of missing values. We integrated the information of market price dataset with temperature and rainfall attributes of weather data. Data transformation techniques were used to normalize the values of dataset. After transformation, we get cleaned and useful data.

## Data Repository

After data preprocessing, the extracted, cleaned and integrated data is loaded into the Data Repository. Data is stored in a data model which will be used for analytics and visualization of the system.

## Prediction

Prediction deals with future events. Data Mining can be used to solve classification and clustering problems. It uses statistics to solve these problems. In this paper, we discussed about various Data Mining algorithms which can be applied to predict commodity price. These algorithms are Decision Tree, K-Nearest Neighbor (KNN), Gradient Boost, Linear Regression and Random Forest. All these algorithms are prescriptive analytics algorithms.

This system takes agriculture price data along with weather data as input. The preprocessed dataset and all the data mining algorithms are provided to model builder to train the model. After training, the trained data helps to calculate the accuracy of all the algorithms. System compares these accuracies to give efficient algorithm as output. Using this efficient algorithm, the predictions about commodity prices are provided to the user.

This is for particular raw product. For another raw product, the process of model building, training dataset and giving efficient algorithm with highest accuracy is repeated. Fig. 1 shows the system architecture for particular raw product.

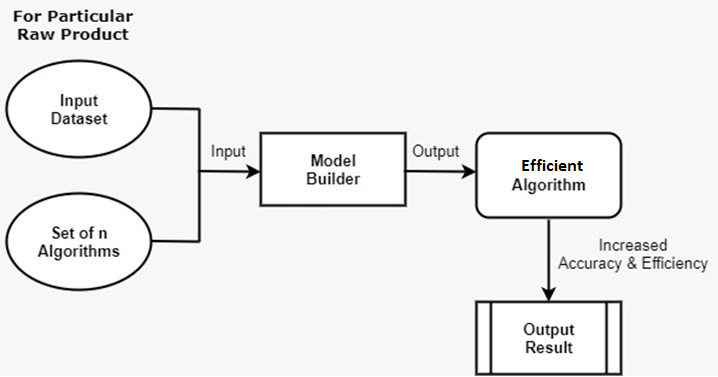


Fig. 1. System Architecture.

##### Algorithms

Various data mining algorithms are used to implement this system. These algorithms are Decision Tree, K-Nearest Neighbor (KNN), Gradient Boost, Linear Regression and Random Forest.

## Decision Tree

Decision Tree is a supervised learning technique that can be used for solving classification problems. In a decision tree, data is spilt based on certain decision kept at root node.

## K-Nearest Neighbor (KNN)

KNN is supervised learning technique that assumes the similarity between new data and available data. It put new data into the category that is most similar to it.

## Gradient Boost

This is a prediction model composed of group of decision trees which are all weak learners. It uses boosted machine learning technique. It builds one tree at a time. Each time newly built tree helps to correct errors made by previously trained tree.

## Linear Regression

Linear regression tries to find out linear relationship between variables. There are two types of linear regression: Simple Linear Regression and Multiple Linear Regression. Simple Linear Regression finds relationship between one dependent and one independent variable while Multiple Linear Regression finds relationship between two or more independent variables and one dependent variable.

## Random Forest

Random Forest is the combination of decision trees where each tree depends on the values of a random vector sampled independently with the same distribution for all trees in the forest. The principal is that single decision tree can be a weak learner, while together they can comprise a strong learner.

##### Evaluation Of Algorithms

All the algorithms discussed in previous section are evaluated using metrics such as accuracy, precision, recall and F1 score. These metrics are measured using confusion matrix. Confusion matrix is a tool for analyzing how well our classifier can recognize tuples of different classes.

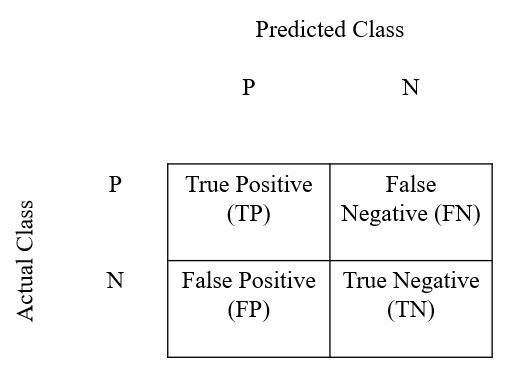


Fig. 2. Confusion Matrix.

## Accuracy

Accuracy of any classification algorithm is the percentage of test set tuples that are correctly classified by the model.

|  |  |  |
| --- | --- | --- |
|  | *Accuracy* = . | (1) |

TP = True Positive, TN = True Negative

FP = False Positive, FN = False Negative

## Precision

Precision means exactness of algorithm. It is the percentage of tuples which are correctly classified as positive are actual positive.

|  |  |  |
| --- | --- | --- |
|  | *Precision* = . | (2) |

## Recall

Recall is the measure of completeness. It is the percentage of positive tuples which the classifier labelled as positive.

|  |  |  |
| --- | --- | --- |
|  | *Recall* = . | (3) |

## F1 Score

F1 Score gives the balance between Precision and Recall. It is the harmonic mean of precision and recall.

|  |  |  |
| --- | --- | --- |
|  | *F1* = . | (4) |

Depending upon these metrics, we decide that which algorithm is best for prediction of prices of given commodity.

##### Conclusions

The process of comparison between data mining algorithms based on metrics such as accuracy, precision, recall and F1 score gives us efficient algorithm for each commodity. This efficient algorithm may differ from commodity to commodity depending upon their measure of accuracy for that commodity. Due to this, we get more accurate and quality results. Accurate predictions of commodity prices and stock will help farmers in decision making process.

##### Conflict of Interest

Vishal S. Wankhede, Harish D. Hatmode, Pragati S. Bankar and Onkar G. Kulkarni declare that they have no conflict of interest

##### Author Contributions

Onkar and Pragati conducted the research, Vishal and Harish Created and analyzed the data. Different Algorithms were studied by all members. Pragati wrote the paper. all authors had approved the final version.

##### References

1. S. Shrivastava, S. N. Pal, and R. Walia, “Market Intelligence for Agricultural Commodities Using Forecasting and Deep Learning Techniques,” *Springer Nature Switzerland*, 2019.
2. Kumar U. D., “Business Analytics: The Science of Data-Driven Decision Making,” *Wiley, New York*, 2018.
3. Larose D. T., Larose C. D., “Data Mining and Predictive Analytics,” *Wiley, New York,* 2017.
4. Mann A. K., and Kaur N., “Review paper on clustering techniques,” *Global Journal of Computer Science and Technolog*y, 2013.
5. Shah N., Solanki M., Tambe A., and Dhangar D, “Sales Prediction Using Effective Mining Techniques,” *International Journal of Computer Science and Information Technologies, Vol. 6 (3)*, 2015.
6. S. Cheriyan, S. Ibrahim, S. Mohanan, S. Treesa, “Intelligent Sales Prediction Using Machine Learning Techniques,” *IEEE*, 2018.
7. Open Government Data (OGD), India. <https://data.gov.in/>
8. Agricultural Meteorology Division. <http://imdagrimet.gov.in/>

1. [↑](#footnote-ref-1)