AN EXPEDIENT PREDICTION TECHNIQUE IN AGRICULTURE INTELLIGENCE

Dr. Tejaskumar R. Ghadiyali
Assistant Professor,
UCCC & SPBCBA & SDGH College of BCA & IT, Udhna, Surat, Gujarat.

Dr. Kalpesh B. Lad
Assistant Professor,
SRIMCA, Uka Tarsadiya University, Bardoli, Gujarat

ABSTRACT

India is the second largest producer in agriculture produce in the world. For the economic development of India, development of agriculture sector is must. Farmer community is the backbone of the agriculture system and well mannered sustainability of farmer community surely leads towards the well being of the country. In this study the researchers have tried to carry out critical analysis of prediction techniques as an exceptional case study with agriculture commodity. This study offers critical analysis of agriculture commodity market price prediction using Statistical and Soft Computing approaches and judge expedient prediction techniques out of it. This study basically contributes to the well being of the farmer community and thereby enhances their ability towards Intelligence agri-business.

Keywords: Prediction Techniques, Price Prediction, Agriculture Intelligence, Soft Computing

I. INTRODUCTION

Agriculture sector plays a vital role in the economic development of India. Approximately 19% GDP is contributed from this sector and more than 60% workforce of the total population of India is engaged in this sector [1]. Over 700 million people i.e.69% of the population depends upon the rural economy for their livelihood [2]. These facts show the importance of the Agriculture sector in India. To sustain the agricultural market in India, more than1756 APMCs across 28 states available. **A**gricultural **P**roduce **M**arket **C**ommittee is a place where an agriculture stake holder can do the business with traders. In the development of Agri-Business, Information Technology contributes a lot in terms of the various web portals, web services and Agriculture Market Information System (AMIS).

e-Chaupal[13], agriwatch[14], indiagriline[15], kishan[16], e-agriculture[17] are some of the examples of web services provided by the government or private agencies. In spite of several advancements in existing information availability, a farmer still remains in dilemma about the price of the agriculture product at the time of its submission in the APMC? It may be possible that due to social and some other economical factors the price may differ in various APMCs. So in this situation the farmer needs to know which APMC rate can be beneficial to him and up to what limit it can profit him. Considering such information and the importance of filling the information gap the researchers tried to predict the agriculture product market price as an ingredient of ‘Agriculture Intelligence’ system [3], before a farmer can deliver it to the market place and thus help the farmer community to increase their profitability and thereby do agri-business intelligently. ‘*Agriculture Intelligence*’ is neither a product nor a system. It is an architecture, which is a collection of integrated operational as well as decision-support components, technologies and databases that provide the agriculture community easy access to agricultural knowledge. Thus agriculture intelligence utilizes various prediction techniques and predicts agriculture commodity market price.

We can classify prediction techniques into two broad categories such as Statistical Techniques and Soft Computing Techniques. In this study the researchers have used statistical techniques such as Regression Analysis and Soft Computing Techniques such as Neural Network for market price prediction of selected agriculture commodity Green Chilly. After having accuracy analysis from both the techniques, the researchers have tried to enhance the accuracy by classifying entire year span into different season span (winter, summer and monsoon) and then monthly to perform accuracy analysis for that particular season for the selected agriculture commodity Green Chilly.

In this study the researchers have studied various prediction techniques; review literature related to it, design methodology for experiments, and offers experiment results and conclusion of the research study.

II. Supportive Prediction Techniques

The researchers have studied and utilized various supportive prediction techniques as and when required during this study. They are Time Series Analysis (Holt's Method for exponential smoothing), Neural Networks, Multiple Linear Regression and Principal Component Analysis. These different approaches were utilized to cope up with the generated problems during the research study.

a) Time Series Analysis

Agriculture commodity market price depends on time factor. So '*Time Series Analysis*' helps us out in predicting such future price. Time series problem analysis is composed of four elements Trend (T), Cycle (C), Seasonal effects (S) and Irregular fluctuation (I). Time series analysis usually involves predicting numeric outcomes, such as the future price of the individual stock or commodity. In such time series analysis what will be the next price value with respect to the new time factor is important, and here forecasting comes into picture. Forecasting is the science of predicting the future.

To forecast the future price '*Exponential Smoothing*' techniques such as Holt's Method also helps us out. Holt's method [4] for exponential smoothing is used to weight data from previous time periods and it shows exponentially decreasing importance in the forecast. Exponential smoothing is achieved by multiplying the actual value for the present time period Xt by a value between 0 and 1 (the exponential smoothing constant) referred to as α and adding that result to the product of present time period's forecast Fp and $(1 - \alpha)$.

$$F_{t+1} = \alpha * X_t + (1 - \alpha) * F_t$$

F_{t+1} = the forecast for the next time period (t + 1)

F_t = the forecast for the present time period (t)

X_t = the actual value for the present time period (t)

α = a value between 0 and 1 as the exponential smoothing constant

Another exponential smoothing technique which includes trend effects is Holt's two-parameter method. Holt's technique uses weights to smooth the trend and it uses the following equations to accomplish this:

Smoothed Values : $E_t = \alpha X_t + (1 - \alpha)(E_{t-1} + T_{t-1})$

Trend Term Update : $T_t = \beta (E_t - E_{t-1}) + (1 - \beta)T_{t-1}$

Forecast for Next Period : $F_{t+1} = E_t + T_t$

For k periods in the future: $F_{t+k} = E_t + kT_t$

In simple exponential smoothing, the smoothed values were the forecasts. The next smoothed value was determined by weighing the actual value by α and the previous forecast (the last smoothed value) by $1 - \alpha$. From the last two equations we observe that forecast is now a function of both the smoothed value (E) and the trend (T). With simple exponential smoothing there is only one smoothing constant, α. Since trend is part of the forecasting process, here we have another parameter (weight) for trend, β.The weights determine how much emphasis is to be placed on recent values versus previous values. Higher values of α and β place more emphasis on recent values. Another improved part in this type of statistical method is Exponential Smoothing with both Trend and Seasonality. Seasonality in our case determined for every month separately and seasonal factor has been calculated for future process. Thus the final forecast value becomes the product of the trend factor and the seasonality factor.

b) Neural Network

In Soft Computing Prediction Techniques *Neural Networks* [5] have a proven track record in predicting numeric or continuous outcomes. It is basically useful in various types of applications like, supervised classification & pattern recognition, un-supervised clustering, curve fitting and dynamic time series analysis. Neural Networks work well with datasets containing large number of noisy input data. Neural Network evaluation function such as 'Sigmoid function' naturally smoothes input data variation caused by outliers and random errors. Neural Networks offer a mathematical model that attempts to try to be like the human brain. The knowledge is often represented as a layered set of interconnected processors. These processor nodes are frequently referred to as 'neurodes', which indicate a relationship of the neurons of the brain. Neural Network learning can be supervised or unsupervised. Learning is accomplished by modifying network connection weights while a set of input instances is repeatedly passed through the network. Once trained, an unknown instance passing through the network is classified according to the value(s) seen at the output layer.

Probably the biggest criticism of neural networks is that they lack the ability to explain their behavior. However we can use the experimental process to help us to achieve desired result. There are several Neural Networks tools which can act as per the theory of the neural network and one can achieve practical results through those tools. The examples of Neural Networks tools are 'nntool' of Matlab[18], Easynn[19], Neuroxl[20], Neurosolutions[21], WEKA[22], ExcelMner[23], IntelligentMiner[24] and cortex[25].

In this research study, the researchers have utilized one of these tools (MATLAB – nntool) and have achieved experimental results.

c) Multiple Linear Regressions

Multiple Linear Regressions [6] attempts to model the relationship between two or more explanatory variables or independent variables (β1, β2, β3 … etc.) and a response variable or dependent variable (Y) by fitting a linear equation to the observed data. The model is expressed as DATA = FIT + RESIDUAL, where the "FIT" term represents the expression β1X1 + β2 X2 + β3 X3 + ... βn Xn. The "RESIDUAL" term represents the deviations of the observed values Y from their means y, which is normally distributed with mean 0 and variance σ. The notation for the model deviations is ε. Formally, the model for multiple linear regressions, given n observations, is:

$$Y_i = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots \beta_n X_n + \varepsilon_i \text{ for } i = 1,2, \ldots n.$$

In the least-squares model, the best-fitting line for the observed data is calculated by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared, then summed, there are no cancellations between positive and negative values. The least-squares estimates β1, β2, β3 … βn are computed by developed program for selected parameters X1, X2, X3 …… Xn . The researchers conducted experiments of multiple linear regressions using statistical software WESSA.

WESSA is freely available Web Enabled Scientific Services & Applications software [115]. It performs statistic calculations and plots mathematical equations. WESSA offers descriptive statistics, regression analysis, statistical distributions, time series analysis and statistical hypothesis testing.

d) Principal Component Analysis (PCA)

PCA is a procedure for identifying a smaller number of uncorrelated variables called "principal components" from a large set of data. The goal of principal components analysis is to explain the maximum amount of variance with the fewest number of principal components. Principal component analysis is commonly used as one step in a series of analyses. The principal component analysis is used to reduce the number of variables or when too many predictors relative to the number of observations exist.

III. LITERATURE REVIEW

The farmer is the backbone of the agriculture system. The core aim of the farmer is to have good earnings through agriculture. This can be possible by three major aspects i) By increasing production, ii) By improving quality and iii) By inculcating intelligence in agribusiness. There are ample of examples of past researches in the domain of agriculture that majorly focused on the first and the second aspect. The combination of the first and the second with the third provides assurance of good earnings to the farmer. The researchers emphasize on the third aspect and have reviewed the literature of prediction systems in the agriculture domain.

The study was initiated by identifying agriculture commodity price data availability followed by price prediction using archive price data.

a) Data Availability

To identify agricultural market price information providers to the farmer community and its related stake holders, the researchers found that there are many web-services [13-17]. Moreover, many portals or web services also provide information about several rural services of Government of India such as E-post [31], E-seva [32]. Also there has been contribution from many private portals and universities such as E-sagu [33], StarAgri [34], MahaAgri [35] and Lokmitra [36].

Agri.Gujarat [27] is a government web service handled by the Department of Agriculture and Co-operatives, Government of Gujarat. It is a common platform to transfer agricultural information of Gujarat state. It provides information related to policy declaration for the farmer community of Gujarat. Subsidy declaration detail with local language support is one of the major features of this service. It also has direct linkage to all four agriculture universities of Gujarat to have agricultural information with regards to that specific geographical area.

I-khedut [28] is also an initiative of Government of Gujarat, designed to provide agricultural information like online application to gain the benefit of Government Policies, Agriculture Loan information, Weather Details, APMC product Rate etc. It is also connected to PMFBY (Pradhan Matri Fasal Bima Yojana) and allows having information from it. Moreover, this portal provides mobile applications to avail static information with ease. There are many web portals designed by the state government to give information benefits to the famers of the state. For example E-Mitra [29] (Rajeshthan), Bhoomi [30] (Karnataka), Lokmitra [36] (Himachal Pradesh) are some of the web portals designed to give information to the farmer community of their states. Agmarknet [12] is one of the authentic data sources governed by the Government of India to avail agriculture crop past market price.

All these agriculture applications create information overflow without proper integration of correlated agricultural information. Moreover, they exist in scattered manner at heterogeneous resources that do not suffice knowledge for agriculture stake holder. To deal with this problem, the researchers started price prediction studies using archival price as discussed in the next topic.

b) Prediction Studies using Archival Price

The researcher have reviewed the literature related to knowledge extraction such as prediction of the future market value of agricultural crop before crop cultivation. There were several studies conducted which determine future market price based on archive price data. The volatility of an agriculture commodity price is very high and therefore price forecasting for decision makers in this domain has become more sensitive and challengeable compared to non-agricultural domain. The approaches used by different researchers are time series analysis with various statistical techniques and soft computing techniques such as neural network.

Govardhana Rao et.al [07] studied the seasonal variation and forecasting in wholesale price of rice in Guntur district of Andhra Pradesh. The study analyzed trends and seasonal variations of rice wholesale price from 1991 to 2010 using price forecasting model ARIMA for Guntur and Andhra Pradesh market and predicted the future trend of rice price. Their result says that for both these markets, the contribution of time to change in prices was to the extent of 73% and 80% respectively. Based on this fact, they have recommended the sale of rice. The major limitation of this study is that, it considers only percentage change of price (annual increase) in Rice and offers suggestions based on it.

Ticlavilca et.al. [08] have performed multiple predictions for agricultural commodity prices before one, two or three months. In order to obtain multiple-time-ahead predictions, the research applies the Multivariate Relevance Vector Machine (MVRVM) that is based on a Bayesian learning machine approach for regression. This study uses data of monthly agricultural commodity prices to evaluate the results. In the volatile agriculture market, monthly agriculture commodity price data has very low impact to justify the prediction. Moreover, this study predicts price before one, two or three months only. In general, agriculture crop production period is more than three months. So this study does not predict price before crop cultivation and hence it is not much useful in risk management activities of agribusiness.

Manpreet Kaur et.al. [09] have discussed applications and techniques of Data mining in agriculture considering the problem of price prediction of crops. To increase the accuracy percentage of price prediction, the researchers utilized back propagation neural network prediction model using the Coimbatore market price of tomato as an example and simulated the result in MATLAB. This study attempted to forecast the prices of vegetables considering the single factor 'changes in the price of crude oil'. Other influencing parameters such as meteorological and fiscal could also have had much impact in forecasting vegetable prices which were not considered in this study. Moreover, sample size considered to conduct the study is only of previous three years, which is a very short period for agriculture price prediction application.

Linwood Hoffman [10] has developed a model to forecast the seasonal average of the agriculture commodity 'Corn' using future prices. The research has used parameters such as basis value, marketing weights and composite of monthly futures and cash prices instead of only archival price, to forecast the seasonal average U.S. corn farm price. There is no significant difference found between result of this study and the projection of U.S. Department of Agriculture. The main problem faced in this study is to have the value of parameter 'cash price received' and marketing weights. Cash price data is an aggregation of monthly data and marketing weights data is of 5-years moving average of these monthly weights. So, daily basis forecasting is not possible. In this way, this study gives us a superficial prediction analysis instead of in-depth prediction analysis.

Wen HUANG et.al. [11] checked the role of high frequency data in forecasting agricultural commodity futures. The study used Realized GARCH model developed by Hansen, Huang and Shek (2012) to estimate and forecast price volatility for agricultural commodity futures. The main focus of this study is to compare realised GARCH model with GARCH and EGARCH model. To improve forecasted result, this study did not apply any evaluation criteria and forecasting methodology.

All the above narrated studies predict price in agriculture domain on their archival price data. After studied all such literature, the researchers designed experiments to judge better techniques that will help every other researcher in market price prediction of an agriculture commodity.

IV. EXPERIMENTS DESIGN

To perform the agriculture commodity market price prediction, the researchers have downloaded data from the authentic Indian government web portal 'Agmarknet'. The scope of the study is for the last six years of calendar data i.e. from 01/01/2012 to 31/12/2017 of agriculture commodity Green Chilly for 'Surat' APMC. The available data is on daily basis and the researchers have performed required pre-processing to get the cleaned data. In this experiment the researchers have used first five years (2012-2016) archival market price data as training data and last year (2017) market price data as testing data. After having cleaned archival market price data for agriculture commodity Green Chilly, the researchers have designed experiments using two different prediction techniques. Market Price Prediction was done using Soft Computing Techniques - Neural Network (using nntool of MATLAB) and Statistical Technique - Linear Regression (using WESSA tool).

Soft Computing Techniques - Neural Network

- Input training data to neural network (nntool of MATLAB) to train it.
- Compared result (predicted market price) with actual market price (testing data) and find out difference.
- Error measure used here was Mean Absolute Percentage Error (MAPE).
- Found out accuracy in percentage using formula. (100 – MAPE).

Statistical Techniques – Linear Regression

- Input training data to linear regression (WESSA tool) to generate linear regression equation.
- Applied Linear Regression equation to Testing data and generate predicted data.

- Compared result (predicted data) with actual market price (testing data) and find out difference.
- Error measure used here was Mean Absolute Percentage Error (MAPE).
- Found out accuracy in percentage using formula. (100 – MAPE).

V. EXPERIMENT RESULTS AND DISCUSSIONS

Based on above design experiments were conducted and result produced for both the prediction techniques for agriculture commodity market price prediction.

Table-1 Yearly data Accuracy% using both prediction techniques

Agriculture Commodity (Green Chilly)	Using Soft Computing Techniques Neural Network (nntool of MATLAB)	Using Statistical Technique Linear Regression (WESSA tool)
MAPE	34.43	33.77
Accuracy %	65.57	66.23

As shown in above table, the accuracy achieved using Linear Regression is somewhat better than the accuracy achieved by Neural Network of soft computing. But still the accuracy did not suffice the overall goal of market price prediction. So to enhance the accuracy level, the researchers again conducted experiments using linear regression technique to predict market price by classifying year data to seasonal (monthly) data for selected agriculture commodity Green Chilly and result was generated. Such generated results of Table – 2 were comparatively better than the results achieved as per Table – 1. In this approach each month linear regression equation (model) is generated separately. And then each month training data is compared with that month's generated test data (using linear regression equation of that month) and accuracy will be generated as per Table-2. For visual convenience of the reader, the researchers also generated graphical representation of its season wise analysis as shown in Figure-1.

Table-2 Seasonal Accuracy Analyses for: Green Chilly using Linear Regression Technique

GREEN CHILLY			
Model Name	AVG ACCURACY	MAX ACCURACY	MIN ACCURACY
GC_NOV	76.79	99.03	42.98
GC_DEC	84.77	99.50	58.84
GC_JAN	85.14	99.52	78.02
GC_FEB	91.22	99.39	79.70
Winter	**84.48**	**99.36**	**64.89**
GC_MAR	83.73	99.97	56.59
GC_APR	82.97	98.29	50.82
GC_MAY	77.03	99.36	44.59
GC_JUN	77.73	97.68	53.72
Summer	**80.37**	**98.83**	**51.43**
GC_JUL	72.61	98.01	8.21
GC_AUG	86.16	99.89	71.32
GC_SEP	88.19	99.94	60.71
GC_OCT	89.02	99.29	73.09
Monson	**84.00**	**99.28**	**53.33**
Yearly	**82.95**	**99.16**	**56.55**

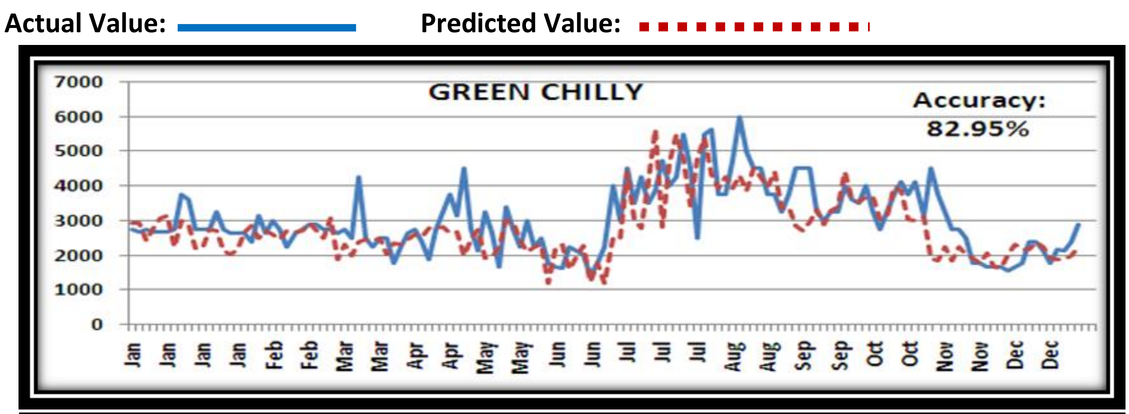

Figure-1 *Comparison of Actual V/s Predicted Market Price Analysis using seasonality concept: Green Chilly*

As shown in above Table-2 the researchers have achieved more accuracy compared to earlier approach. Here the researchers achieved 82.95% accuracy in market price prediction of agriculture commodity – Green Chilly. It means that the predicted future market price for this commodity in next year will be ranging from ±17.05% than its actual price in that period. Here the beneficiary can also take benefit of each month average accuracy level and predict accordingly for that month.

VI. CONCLUSION

In this study the researchers perform critical analysis of prediction techniques with a special reference to an agriculture commodity Green Chilly. The researchers also perform comparative analysis of statistical techniques-Linear Regression with soft computing techniques-Neural Network and suggest that for agriculture market price prediction with archival price data, Linear Regression is better. Such conclusion will also help those researchers who would like to choose prediction techniques in agriculture commodity market price prediction. Such predicted price in the form of knowledge will be helpful to the policy makers such as farmer community, APMC brokers and other agriculture stake holders, to take significant economic decision. The potential benefit of this prediction study lies in assisting producers in making better-informed decisions and managing price risk, which surely helps them in achieving their ultimate goal of profit making.

REFERENCES

1. Government of India Planning Commission (2014). Report of the Working Group on Agriculture Research and Education for the 12th Five Year Plan 2013-2018, pp.16-17.
2. Planning Commission Govt. of India (2014). "Twelfth Five Year Plan 2014–18",Agriculture, Rural development, Industry, Service and physical infrastructure, Vol. III.
3. Ghadiyali, Tejas; Lad, Kalpesh and Patel Bankim (2011). "Agriculture Intelligence: An emerging tool for farmer community", Proceedings of Second International Conference on "Emerging Application of Information Technology" EAIT 2011.18–20 Feb, Kolkata. 2011 IEEE DOI 10.1108\9/EAIT 2011.36, Vol.2 pp 313-316.
4. Ken Black (2009), "Business Statistics for Contemporary Decision Making" 4th edition, Wiley Publication. ISBN:978-81-265-0809-9 , 2009 Printed at 'Sheel Print-N-Pack, Noida, Delhi, India.
5. Richard J. Roiger & Michael W. Geats (2008). "Data Mining, A Tutorial-based Primer", Second Imression, Publisher - Pearson Education, ISBN 978-81-317-1512-3.
6. Ken Black (2009), "Business Statistics for Contemporary Decision Making" 4th edition, Wiley Publication. Printed at 'Sheel Print-N-Pack, Noida, Delhi, India, ISBN:978-81-265-0809-9.
7. Govardhana Rao G., Solmonrajupaul K., Vishnu Sankarrao D., Dayakar G. (2014). Seasonal Variations and Forecasting in Wholesale Prices of Rice (Paddy) In Guntur District of Andhra Pradesh In: International Journal of Development Research, Vol.4(11), pp.2418-2422.
8. Ticlavilca, A. M., Dillon M. Feuz and Mac McKee (2010). Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression In: Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St.Louis, MO. April 19-20.
9. Manpreet Kaur, Heena Gulati, Harish Kundra (2014). Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications In: International Journal of Computer Application Vol.99(12), pp. 1-3.
10. Linwood Hoffman (2004). Using Futures Prices to Forecast the Season-Average U.S. Corn Price In: Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management St. Louis, April 19-20.
11. Wen HUANG, Zhuo HUANG, Marius MATEI, Tianyi WANG (2012). Price Volatility Forecast for Agricultural Commodity Futures: The Role of High Frequency Data In: Romanian Journal of Economic Forecasting Vol.4, pp.83-103.
12. www.agmarknet.nic.in visited on 24th February 2018 at 6.30 p.m.
13. www.itcportal.com visited on 18th February 2018 at 10.15 a.m.
14. www.agriwatch.com visited on 18th February 2018 at 11.30 a.m.

15. www.indiagriline.com visited on 20th February 2018 at 4.00 p.m.
16. www.kishan.com visited on 20th February 2018 at 6.30 p.m.
17. www.e-agriculture.org visited on 20th February 2018 at 7.45 p.m.
18. MatLab (www.mathworks.com).
19. EasyNNPlus (www.easynn.com).
20. PredictorXL (Author:Yaskov Anatoliy, Company:OLSOFT-LLC, Ver: 3.2.1)
21. Neurosolutions (www.neurosolutions.com)
22. WEKA (www.cs.waikato.ac.nz/ml/weka/)
23. ExcelMiner (www.kdnuggets.com)
24. IntelligentMiner (www.ibm.com/software/data/iminer/)
25. Cortex (www.cortex.snowcron.com)
26. Wessa, P. (2016), Free Statistics Software, Office for Research Development and Education version 1.1.23-r7, URL http://www.wessa.net.
27. http://agri.gujarat.gov.in/ Last visited on 2nd February 2018 at 9.30 p.m.
28. https://ikhedut.gujarat.gov.in Last visited on 1st January 2018 at 5.30 p.m.
29. www.emitra.gov.in Last visited on 15th January 2018 at 3.45 p.m.
30. www.bhoomi.karnataka.gov.in Last visited on 21 January 2018 at 7.15 p.m.
31. indiapost.nic.in Last visited on 15th February 2018 at 4.00 p.m.
32. india.gov.in/e-seva-portal-online-citizen-services-andhra-pradesh visited on 15th Feb.18 at 5 p.m.
33. www.agriculture.iiit.ac.in (esagu) Last visited on 18th January 2018 at 3.30 p.m.
34. www.staragri.com Last visited on 11th January 2018 at 7 p.m.
35. www.mahaagri.gov.in Last visited on 13th January 2018 at 4.15 p.m.
36. www.hp.gov.in (Lokmitra Kendra) Last visited on 13th January 2018 at 6.30 p.m.