Increasing small holder farmer income by providing localized price forecasts

Anita Yadav[1], Arjun Verma[1], Dr. Anitha Govindaraj[2], Vikram Sarbajna[1]
[1*]Credible, Mumbai, India
[2†] AIGGPA, Bhopal, India

Abstract

Around 82% of India's farming community fall under the category of small holder farmers. These farmers are increasingly exposed to production risks from changing weather patterns and price risk due to volatility in domestic and international markets. Price forecasting can be used to manage price risks by helping the farmer with two main post harvest decision points: when to sell and which market to sell at. In this paper, we present a unique machine learning approach for forecasting crop prices at a high frequency (daily rolling forecasts) for local markets. We have been able to forecast prices for Soybean, Chickpea and Mustard in the state of Madhya Pradesh. An improvement of up to 19% (for Soybean prices) in forecast accuracy has been achieved when compared to traditional uni-variate forecasting methods. Informed selling using our forecasts can provide farmers an increase of up to 14.5% in their income. We shall be providing price forecasts to the farmers in partnership with the State Agricultural Marketing Board of MP through APIs for the upcoming winter harvest season.

1 Background and need for price forecasts

Agriculture plays a crucial role in the Indian economy, it is estimated that around 59% of the total workforce in the country earn their livelihood through agriculture [1]. An overwhelming majority (82%) of the farmers in India are categorized as small and marginal (landholding of less than 2.0 hectares). These farmers are vulnerable to myriad external risks [2] such as production risks arising from changing weather patterns, pests and diseases to market risks which are influenced by supply and demand factors from both domestic and international markets. Current machine learning (ML) applications are heavily tilted towards managing production risks (for example localized weather forecasts, image-based disease detection [3] etc.). Given this context, we have created crop price forecasting models that can be used by small holder farmers to mitigate price volatility and increase their income. Lack of localized and accurate information makes smallholder farmers in India especially vulnerable to these risks [4]. Together with the Atal Bihari Vajpayee Institute of Good Governance and Policy Analysis (AIGGPA) and the State agricultural marketing board (Government of Madhya Pradesh), we created and tested price forecasting models for selected crops in the Indian state of Madhya Pradesh (MP). AIGGPA is a public policy advisory institute for the state government of MP. It had recently concluded a study on farmer distress in MP[5] where it advocated for better risk management measures for price volatility.

Our work and forecasting models are unique in the following ways, firstly we demonstrate that by combining technical and fundamental factors we can outperform traditional uni-variate models in crop price forecasting. Secondly, while most forecasts are generalized and at a national or state level, we deliver forecasts for local markets that are relevant to small holder farmers (SHF).

[*]http://www.credible-india.com
[†]Atal Bihari Vajpayee Institute of Good Governance and Policy Analysis (http://www.aiggpa.mp.gov.in)

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

2 Delivering localized price forecasts for SHF

Most work on price forecasting [6, 7, 8] have used the Autoregressive Integrated Moving Average (ARIMA) model [9]. The forecast frequency is typically once a month and for a limited number of markets. This setting is well suited for supporting long term policy analysis but falls short in its usability for SHF, where the need is for higher forecasting frequency (rolling forecasts for each day of the week) and to cover all the relevant markets close to the farmer. In short, for a price forecast to be relevant for the farmer it needs to be localized.

2.1 Scope of the work

The crops were selected based on their relevance for MP (in terms of total production in Metric Tons). MP is the largest producer of Soybean and Chickpea in India accounting for around 51.3% and 37% of the total production for these crops respectively[5]. It is the second largest producer of Mustard with 11.5% of the total production. Crops were also selected based on high price volatility. Initially we had considered horticultural crops given their high price volatility, however we had to discard them due to limited historical data availability.

The forecasts had to be provided for all the major markets in the state (top 30 markets in terms of volume of arrivals for each crop). Finally, a forecast horizon of 14 days was set with a daily forecast frequency.

2.2 Training data set and models used

The training data set comprised of technical (historical price) and fundamental factors influencing the prices at a local market level. The fundamental factors were identified using a combination of literature research and expert interviews. Table 1 provides an overview of the common fundamental factors for all three crops.

Table 1: Fundamental factors influencing prices

Identified factors
Modal prices at selected markets
Arrivals at selected markets
Weather (temperature, rainfall, humidity) at a district-level
Minimum Support Price (MSP)
Area under cultivation at a district-level
Yield at a district-level
Total production at a district-level
Currency exchange rates for major trading (Import/Export) countries
Area & Production of neighbouring states
Area & Production of major exporting countries
International consumption

Most of the data was publicly available. Historical prices at various markets were obtained from the Indian central government data-portal Agmarknet [10]. It provides daily prices and arrival volumes of commodities at various markets of India. AIGGPA sourced certain data sets such as district-level area, yield and production. For international production and consumption, we used the USDA data-portal [11].

For Soybean, the training set consisted of data from 1st of January 2014 to 30th of September 2018. The testing of the forecasts was done for the peak arrival season which is from 1st of October 2018 to 31st of January 2019. For Chickpea and Mustard the training set consisted of data from 1st of January 2014 till 1st of February 2019. The testing period was 1st of March 2019 till 31st of June 2019, which coincides with their peak arrival season at local markets.

Four classes of supervised ML models were tried out, these being random forest regressor[12], extreme gradient boosting (XGBoost)[13] & LASSO[14]. For comparison purposes we also trained the

ARIMA model. Recurrent Neural Networks (RNN) and Long short-term memory (LSTM) models were initially considered but ultimately discarded due to their lack of applicability to this particular data set.

3 Results and Evaluation

3.1 Prediction accuracy

We used root mean squared error (RMSE) as a measure of accuracy for the price forecasts. For Soybean we achieved an RMSE of Rs.155.38 during the period of October 2018 to January 2019. The actual prices during this period ranged between Rs.2726 to Rs.3930 per quintal. For Chickpea an RMSE of Rs.250.13 was achieved during the period of March 2019 to June 2019. The actual prices during this period ranged from Rs.2300 to Rs.4620 per quintal. Finally for Mustard we were able to achieve an RMSE of Rs.158.91 during the period from March 2019 to June 2019. Mustard prices during this period ranged from Rs.2700 to Rs.4550.

Table 2: Crop-wise comparison of algorithms

Crop	ARIMA	XGBoost	LASSO	Random Forest
Soybean	194.11	196.42	**155.38**	193.98
Chickpea	255.50	321.38	**250.13**	330.09
Mustard	174.07	162.07	188.99	**158.91**

Table 2 provides the RMSE achieved and a comparison of perfomance between uni-variate model (ARIMA) with multivariate models such as LASSO, random forest and XGBoost. The results demonstrate that multivariate models consistently out perform ARIMA. The forecasts were provided for 30 markets during the peak arrival season. Tables 3, 4 & 5 provides a further breakdown of the achieved RMSE for a subset of these markets.

Table 3: RMSE for Soybean

Market	RMSE
A lot	106.13
Badnagar	117.07
Badnawar	127.80
Betul	128.18
Damoh	97.86
Dhar	104.73
Itarsi	124.85
Mahidpur	130.70
Ratlam	120.59
Ujjain	133.45

Table 4: RMSE for Chickpea

Market	RMSE
Ashoknagar	429.13
Badnagar	229.67
Damoh	218.16
Gadarwada	158.16
Ganjbasoda	220.41
Guna	169.66
Jabalpur	142.47
Katni	150.39
Sagar	134.55
Sironj	196.65

Table 5: RMSE for Mustard

Market	RMSE
Kailaras	151.34
Lahar	100.22
Lashkar	119.52
Morena	118.83
Neemuch	84.52
Piplya	189.67
Rewa	222.27
Sabalgarh	156.47
Sagar	258.55
Satna	180.75

3.2 Prediction utility for farmers

By providing 14 day price forecasts we help the farmer with two major decisions i.e. when to sell and which market is the most attractive to sell to. Taking Soybean as an example, figure 1 shows how the prices fluctuate for 3 different markets within the same district of Ujjain for 14 days. The prices can vary anywhere between Rs.3362 to Rs.3852 per quintal. Acting on our forecasts and selling at the market with the highest price would mean an increase of up to 14.5% in the farmer's income.

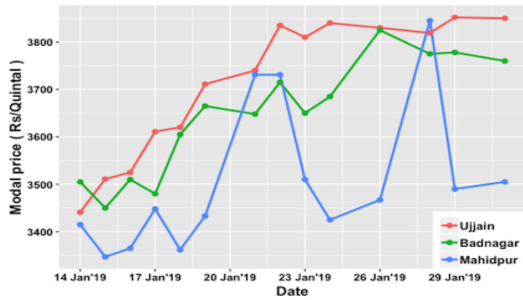
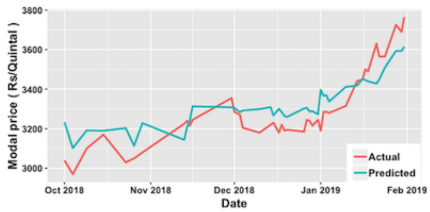
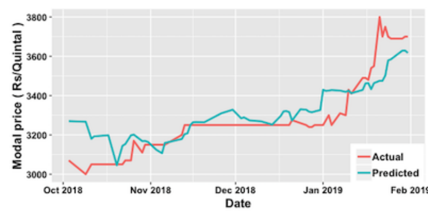


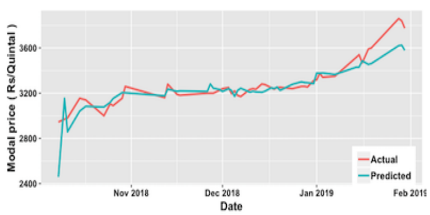
Figure 1: Price fluctuation for 3 markets within the same district in a 14 day period



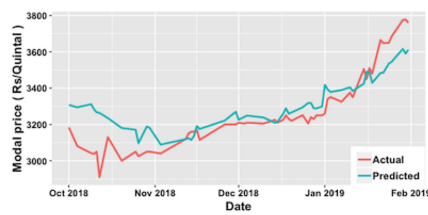
(a) Damoh



(b) Dhar



(c) A Lot



(d) Badnagar

Figure 2: Actual and predicted prices for Soybean across 4 markets

4 Discussion and Conclusion

We have successfully created price forecasting models based on a combination of technical and fundamental factors for three crops for farmers in the state of MP, India. In the process, we have also demonstrated in Table 2 that our approach outperforms traditional uni-variate price forecasting methods. By providing simultaneous forecasts for multiple markets, we have improved its utility for SHF.

4.1 Limitations, risks and negative outcomes

An obvious limitation of the models would be their inability to account for sudden and one-off events such as demonetization [5]. A risk and a negative outcome of price forecasting is the possible distortion of market prices. If all the farmers who receive the forecasts decide to sell at one particular market, this could lead to over-supply and subsequently lower prices with long waiting period for the farmers to sell their produce. Contingencies for such an outcome needs to be planned while preparing for deployment in real world.

4.2 Future work

Going forward, we intend to increase the number of crops, increase the forecast horizon from 14 days to 90 days (providing support for deciding which crops to grow) and provide the price forecasts to farmers in MP using APIs, through the State Agricultural Marketing Board of MP.

References

[1] FAO. India at a glace. http://www.fao.org/india/fao-in-india/india-at-a-glance/en.

[2] NITI Aayog. Raising Agricultural Productivity and Making Farming Remunerative for Farmers. An Occasional Paper, 2015.

[3] S. R. Huddar, S. Gowri, K. Keerthana, S. Vasanthi, and S. R. Rupanagudi. Novel algorithm for segmentation and automatic identification of pests on plants using image processing. In *2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12)*, pages 1–5, 2012.

[4] Nitin Bhagachand Bachhav. Information needs of the rural farmers : A study from maharashtra, india: A survey. *Library Philosophy and Practice LPP*, 2012. ISSN 1522-0222.

[5] Anitha Govindaraj. Agriculture Distress and Farmers' Unrest in Madhya Pradesh: An Exploratory Study. Project Report, Atal Bihari Vajpayee Institute of Good Governance and Policy Analysis, 2018.

[6] Ashwini Darekar and A Amarender Reddy. Cotton Price Forecasting in Major Producing States. *Economic Affairs*, 62(3):373, 2017. ISSN 0424-2513, 0976-4666.

[7] Ashwini Darekar and Amarender A. Reddy. Forecasting of Common Paddy Prices in India. *Journal of Rice Research*, 10:71–75, 2017.

[8] Dinesh Meena, O.P Singh, and Rakesh Singh. Forecasting mustard seed and oil prices in India using ARIMA model. *Annals of Agri Bio Research*, 19(1):183–189, 2014.

[9] George Edward Pelham Box and Gwilym Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, Inc., San Francisco, CA, USA, 1990. ISBN 0816211043.

[10] AGMARKNET. http://agmarknet.gov.in.

[11] USDA. Foreign agricultural service. https://www.fas.usda.gov/data.

[12] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 1573-0565.

[13] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pages 785–794. ACM Press, 2016. ISBN 9781450342322.

[14] Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246.