



AI-based Market Intelligence Systems for Farmer Collectives: A Case Study from India

RONAK LADHAR and SOURAV SHARMA, IIT Delhi, India
SRIKANT TANGIRALA, BITS Pilani, Hyderabad Campus, India
NISHANT GUPTA, IIT Delhi, India
ABDUL AZEEM, BITS Pilani, Hyderabad Campus, India
ARJAV JAIN, BITS Pilani, Hyderabad Campus, India
BHUVAN CHAND KATAKAM, BITS Pilani, Hyderabad Campus, India
ADITYA BOMMAKANTI, IIT Delhi, India
C. SANKARAIAH, Center for Collective Development, India
HARI PRASAD PIRIDI, BITS Pilani, Hyderabad Campus, India
KAUSHALENDRA YADAV, Farmveda, India
KUMRA VITTALRAO, Center for Collective Development, India
MATIUR RAHMAN, Gram Vaani, India
RASHUL CHUTANI and RISHI SHAH, IIT Delhi, India
ROHAN KATEPALLEWAR, Gram Vaani, India
DIPANJAN CHAKRABORTY, BITS Pilani, Hyderabad Campus, India
AADITESHWAR SETH, IIT Delhi, India

Small and marginal farmers are unable to get a good price for their produce, because they face several challenges in market participation. Aggregation of produce via farmer cooperatives and the ability to delay sales (for non-perishable crops) to when market prices are high has emerged as a useful strategy to improve farmer incomes. We work with a network of farmer cooperatives in India growing soyabean and explore the potential of developing a machine learning-based price forecasting and sales recommendation system that produces suggestions on the best dates when harvested soyabean crops should be sold, e.g., whether to sell right away (if prices are likely to fall in the future) or to wait (if prices are likely to rise). We present an evaluation of different methods for price forecasting and a prospect theory-based method to produce sales recommendations. Experiments on historical data indicate that we can provide modest gains to farmers, and we build

We express our gratitude to the Google AI for Social Good program for providing funding for this project. Authors' addresses: R. Ladhar, S. Sharma, N. Gupta, A. Bommakanti, R. Chutani, R. Shah, and A. Seth, Department of Computer Science and Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi - 110016; emails: mcs212146@cse.iitd.ac.in, ssourav.iitd@gmail.com, {Nishant.Gupta, cs5190471, cs5180789, Rishi.Shah.cs119, aseth}@cse.iitd.ac.in; S. Tangirala, A. Azeem, A. Jain, B. C. Katakam, H. P. Piridi, and D. Chakraborty, Department of Computer Science and Information Systems, BITS Pilani Hyderabad Campus, Secunderabad, Telangana - 500078; emails: {f20200055, f20191234, f20191141, f20190142, p20210102, dipanjan}@hyderabad.bits-pilani.ac.in; C. Sankaraiah, K. Yadav, and K. Vittalrao, Center for Collective Development, 1444, 17th B Main Road, Marenahalli, 2nd Phase, J. P. Nagar, Bengaluru, Karnataka - 560078; emails: sankar.ptp@ccd.ngo, kaushalendra_yadav@farmveda.in, kumravittalrao@gmail.com; M. Rahman and R. Katepallewar, Gram Vaani Community Media, Plot No. 2, First Floor, 100 Feet Road, Ghitorni, MG Road, New Delhi - 110030; emails: matiur.rahman@oniondev.com, rohan.22march@gmail.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2834-5533/2023/09-ART6 \$15.00

<https://doi.org/10.1145/3609262>

and field test an Android application for this purpose. Early results indicate positive feedback. Our methods can be generalized to other agricultural commodities that can be stored for several months and help farmer cooperatives to compete effectively in agricultural markets.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → **Agriculture**; • **Computing methodologies** → **Machine learning**;

Additional Key Words and Phrases: Commodity prices, time-series forecasting, sales recommendation, soybean, prospect theory, cooperatives

ACM Reference format:

Ronak Ladhar, Sourav Sharma, Srikant Tangirala, Nishant Gupta, Abdul Azeem, Arjav Jain, Bhuvan Chand Katakam, Aditya Bommakanti, C. Sankaraiah, Hari Prasad Piridi, Kaushalendra Yadav, Kumra Vittalrao, Matiur Rahman, Rashul Chutani, Rishi Shah, Rohan Katepallewar, Dipanjan Chakraborty, and Aaditeshwar Seth. 2023. AI-based Market Intelligence Systems for Farmer Collectives: A Case Study from India. *ACM J. Comput. Sustain. Soc.* 1, 1, Article 6 (September 2023), 32 pages.
<https://doi.org/10.1145/3609262>

1 INTRODUCTION

More than 57.8% of the Indian population is engaged in agriculture, with 86.06% of farmer households being small and marginal farmers with land-holding below 2.5 hectares [12]. Any surplus produce beyond what is required by the smallholder farmers for subsistence, or cash crops grown by them, are mostly sold to local traders who collect produce at the farm gate and aggregate it for sales at nearby market yards [9]. Although state governments in India operate more than 6,946 such market yards across the country [22], the small amounts of produce by individual farmers makes it unviable for them to transport their produce and participate directly in auctions at the market yards. Working via local traders, however, reduces the profit margins for farmers and also makes it hard for them to negotiate better prices since the traders tend to operate as local monopolies and are often also lenders of credit to the farmers and thus are able to maintain significant power over them [19]. Many variations of such local capital structures exist in India with local traders replaced by market commission agents in some parts or millers of agricultural produce in other parts, but the broad dynamics of exploitation and powerlessness of small and marginal farmers is similar.

Electronic marketplaces cannot solve this problem unless they also address the transportation and credit needs of smallholder farmers [5]. However, farmer collectives such as cooperatives and farmer producer companies are a well acknowledged means of countering this power imbalance. Groups of small and marginal farmers organized into collectives are able to pool their produce and sell in bulk in the market, which gets them better prices than by selling individually to local traders [27]. They may also sell directly to large food processing corporations instead of going via local markets to get an even higher profit share. Mature collectives are also able to build their own corpus of funds or are able to raise loans from institutional lenders such as banks, which they deploy to set up their own food processing units. This vertical integration improves their margins even more by enabling them to sell value-added products than just the raw produce [27]. Other benefits of collectives include the farmers being able to learn from one another to adopt more productive agricultural practices, pool their requirements for agricultural inputs to be able to procure them at lower costs, and also collectively plan for more sustainable crop cycles and water conservation practices [27].

For our study, we work with **The Center for Collective Development (CCD)**,¹ a not-for-profit organization founded in 2004 that supports over 43,000 farmers to collectivize into

¹<https://www.ccd.ngo/about-us.html>

village-level farmer cooperatives and district-level federations² and has helped boost farmer incomes by at least 20%. CCD provides technical support to the cooperatives to manage their accounts, raise loans, and build market linkages.

In the district of Adilabad in Telangana in India, soyabean cultivation is widely practiced by CCD assisted cooperatives, who have evolved an interesting technique to improve profit margins for farmers. This technique, locally called *Anamat*, forms the primary focus of our study. Under this system, farmers, upon harvesting soyabean, leave some amount with the cooperative that the cooperative stores and sells in bulk at a later date. This helps improve farmer incomes, because when the entire soyabean harvest hits the market, typically in September each year, the prices that farmers get are low. Prices normally begin to climb after September, and since soyabean can be safely stored for two to four months in a shaded and enclosed space without the need for any special cold storage infrastructure, the cooperatives are able to pass much higher prices to the farmers. In our field visits, we found examples where a cooperative sold their soyabean harvest for as high as INR 6,500 per quintal,³ much higher than the typical price at the time of harvesting in September of INR 4,000 per quintal.

Prices, however, may not always rise as much, and fluctuations are also common. In the same year as described above, another cooperative in the area held their soyabean stock until prices climbed to INR 6,000 per quintal, but the prices dropped to INR 5,500 in between and the cooperative decided to sell, being more risk averse than the earlier cooperative. A key problem articulated by the federation board members was that of price forecasting—if they knew what prices were likely to prevail in the future, then they could guide the cooperatives on when to sell and compete more effectively with other participants in agricultural markets. Having been engaged in soyabean cultivation since several decades, the farmers are well aware of dynamics that influence prices—unseasonal or erratic rainfall damages the crop and impacts its quality, which fetches lower prices; high prices of soyabean in a particular year often lead to more land being shifted by farmers to soyabean in the following year, which results in lower prices in that year; good harvests in major soyabean-producing countries like Brazil results in a global glut by January, which is when Brazil soyabean hits the market; a higher exchange rate for Indian rupees also results in higher prices; and so on. This awareness of macro trends does not, however, translate easily to predict the prices to expect in the future.

The primary objective we identified, in consultation with CCD and their project staff, was as follows: Use various datasets such as time series of historical soyabean prices, satellite imagery, and information about news events, to forecast soyabean prices for the next month or so and, taking these forecasted prices and the uncertainty in the forecasts into account, recommend dates when sales of stored soyabean produce should be made. A secondary objective was also identified to create greater solidarity within the member farmers of the cooperatives and federation, and the mechanism decided was to provide an internal communication system for the federation to send periodic announcements, board meeting minutes, agricultural advisories, and so on, to the farmers. The system could also be used to relay price updates and sale recommendations directly to the farmers.

In this article, we describe the price forecasting and sales recommendation methods in detail. After extensive experimentation, we developed a method for price forecasting using multivariate **Temporal Convolutional Networks (TCNs)** and used the prospect theory of behavioral

² A farmer cooperative is a legal entity that is wholly owned and run by its member farmers. The cooperatives nurtured by CCD are at a village level with 50–100 farmers in each cooperative. The cooperatives are aggregated into a district-level federation that helps the cooperatives to find bulk buyers, negotiates on their behalf, and provides loans to farmers and especially to new cooperatives to strengthen them.

³ 1 quintal = 100 kgs, and a cooperative is typically able to store and sell 50–100 quintals of soyabean in a season.

economics to make sales recommendations taking into account the forecasted prices and their uncertainty. We also describe an Android application we built to provide the forecasts and recommendations to the federation and CCD field officers. Finally, we describe our field testing experience over five months (and ongoing) when we sought live feedback from CCD field officers on the recommendations produced by the app for tallying them with sales actions taken by the cooperatives to evaluate the readiness of the application for deployment. Our overall framework for price forecasting and sales recommendation can be generalized to other non-perishable commodities with similar dynamics, such as onion and groundnut, which can be stored easily, and to some extent potato, which requires a cold storage infrastructure. In our work so far, we assume that in the near future our tools will be used only by a few farmer cooperatives, and therefore we do not investigate equilibrium effects with widespread use of the system.

Our key contributions are threefold as follows: To identify the specific problem of price forecasting that is relevant for farmers of many non-perishable commodities to make better profits, to develop and evaluate a methodology for price forecasting and to use it in making sales recommendations, and to document feedback from a field testing of the application to determine its suitability for deployment with the federation. We next describe related work in time-series modeling and prospect theory, followed by a description of the context, the dataset, and the forecasting and sales recommendations methods along with their evaluation. We finally present details of the Android application and findings from a field test of the application.

2 RELATED WORK

Time-series forecasting: Time-series modelling has many applications, such as in traffic forecasting, weather forecasting, census analysis, and stock price forecasting. Techniques such as artificial neural networks [24], deep learning methods such as **Long Short-Term Memory (LSTM)** [20], and multivariate statistical methods [30] have been shown to help with agricultural commodity price forecasting. Other applications have also been suggested, such as to select crops to sow based on expected price realizations predicted using machine learning [7].

More recently, borrowing from the success of convolutional neural networks in computer vision and language tasks, TCNs have been proposed to learn non-linear features on arbitrary dependencies on historical data for time-series forecasting [6]. This has found uses in wind power forecasting [43], stock trend prediction [11], spot price forecasting of natural gas [34], and carbon price forecasting [28], among others.

In our work, we use TCNs with multivariate inputs and show that it works better than other methods in our problem setting. Among the inputs to forecast soyabean prices in a particular local market, we experiment with using the past prices of soyabean in the market, satellite-based indexes calculated over geographical regions close to the market, relevant local, national and international news events mined from newspapers, and a novel method to also incorporate past prices of other local markets. This method uses the observation that prices in various national markets converge rapidly, with large markets typically providing a leading indicator of future trends [23], and can hence provide useful signals for price forecasting. News articles, however, are expected to incorporate information especially about exogenous events such as weather disruptions, economic crises, and foreign exchange markets [29]. Satellite data are expected to bring additional dimensions, such as cumulative weather or climate effects that may be observable through variables related to vegetation health, rainfall, humidity, and so on [3].

Prospect theory: Prospect theory was developed by the Nobel Prize winners in behavioral economics, Daniel Kahneman and Amos Tversky, in 1979 [25]. The theory models the *loss aversion bias* in humans, where humans tend to perceive losses as more severe than gains of a similar magnitude. An extension also models the *certainty effect*, where humans tend to favor more certain

outcomes over less certain ones. Several researchers have empirically obtained bias functions in different cultures and for different settings. We selected the functions proposed by Abdellaoui, et al. [1, 2] for the probability weight and value components since their setting involving two-outcome prospects is similar to our setting where cooperatives may sell either on the current day (reference) or on some day in the future. In the forthcoming sections, we describe these bias functions and the possible gains from following the recommendations.

3 CONTEXT AND SYSTEM DESIGN

Soyabean is among the most popular oilseeds in the world, and India is a major soyabean producer, producing 7 million tons (4% of global soyabean production) of soyabean annually [31]. The main production states are Rajasthan, Madhya Pradesh, and Maharashtra.⁴ Since a lot of agriculture in India is rain-fed, most soyabean is grown in the *Kharif* season from July to September, during the monsoons. At the time of harvest, a rapid increase in arrival quantities of soyabean in the markets leads to low prices. The prices gradually increase as arrivals reduce. Since harvested soyabean can be easily stored for several months without requiring any cold storage facilities, large traders are able to procure soyabean cheaply from small farmers and sell it later when prices climb up.

Cooperatives assisted by CCD use the same method under the *Anamat* system: They aggregate produce from the member farmers, store it in a local rental space, and sell it at an appropriate time in the coming months when the prices are higher. Cooperatives that have accumulated a corpus of funds are also able to provide an advance payment to farmers at a rate of 50% of the minimum price at which they expect to sell, so that farmers get some ready cash for their day to day expenses or loan repayments. The final balance after the sale is passed on to the farmers after deductions are made for transportation and other overheads.

The cooperatives may not always sell at the local market—in many cases they may sell directly to soyabean processing companies in the region. The rate given by the companies is somewhat higher than the market rates but follows the same trends. In our study, our aim therefore is to forecast prices in the local market of Adilabad and make recommendations of when to sell. We had to undertake many steps to get to this point. Figure 1 shows the overall design of the system we put in place.

First, we found that historical market prices of soyabean in Adilabad were sparingly available. The **Agricultural Marketing Information Network (Agmarknet)** website by the Government of India hosts an archive of historical prices of all commodities in all local markets of India, but a lot of data is missing. We therefore built a method for price forecasting for some of the main soyabean markets in India and used transfer learning to fine-tune this model to Adilabad. To obtain historical prices of these main markets, we put up a system to crawl data from Agmarket daily, shown to the left in Figure 1. We also put in place a data collection process using the ODK application for CCD field officers to provide daily market prices in Adilabad to build a local time series for Adilabad that could be used in the transfer learning process. Both data streams of price time series, from Agmarknet for the main markets and through ODK for Adilabad, had occasional missing values. We used the **Auto-Regressive Integrated Moving Average (ARIMA)** statistical method to impute this data based on past values. In prior work, we experimented with different imputation methods, such as ARIMA, last observation carried forward, linear interpolation, and Stineman interpolation, and found Stineman interpolation to work best to fill missing values in agricultural commodity price time series [18]. However, in our current work, we need to impute based on past values only without taking future values into account, and ARIMA worked best in this case.

⁴<http://www.sopa.org/statistics/soybean-hectares-planted/>

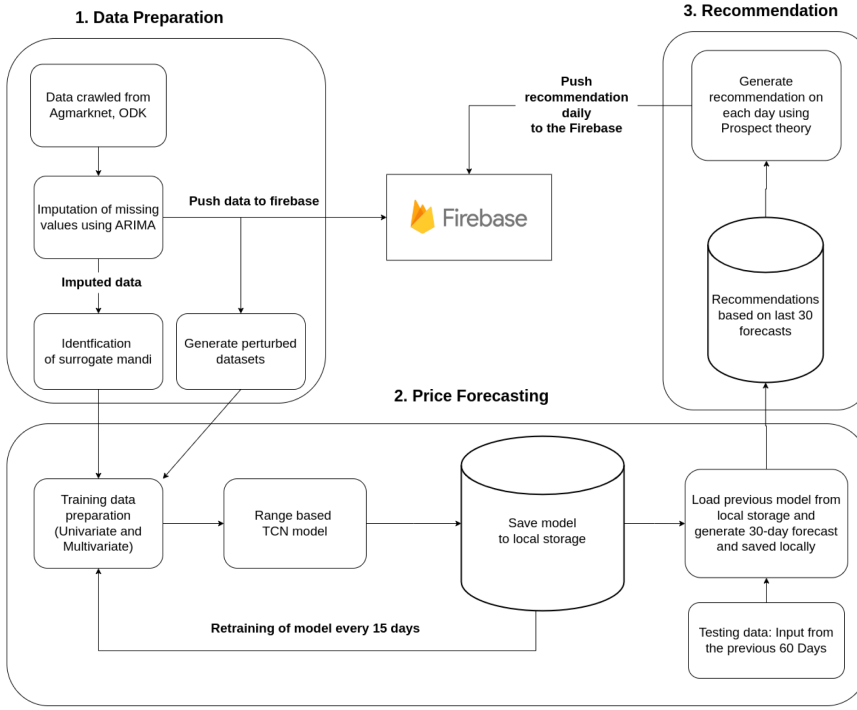


Fig. 1. System design.

Second, the interpolated time series for the main markets were used to train a TCN model. We train separate models for short-term daily (30 days) forecasts and long-term weekly (12 weeks) forecasts, with short-term forecasts expected to be useful for farmers who need cash quickly but may have the capacity to defer sales by a few weeks, and long-term forecasts may be useful for farmers who have greater resilience and can defer sales for longer time periods. In the pilot, we experimented with only short-term forecasts and discuss that next. This is shown to the bottom of Figure 1. The model was first trained using data from January 1, 2006 to December 31, 2013 and is then re-trained every 15 days. The most recent model is used to make forecasts for today and the next 29 days (total 30 days) for the main markets, and a model fine-tuned for Adilabad using the data collected via ODK is used to make forecasts for the Adilabad market.

Third, the forecasted prices are used to generate a recommendation for the three best dates to sell, today or in the next 29 days. This is shown to the right in Figure 1. We use prospect theory to make the recommendations. Prospect theory models two kinds of biases in humans: the *loss aversion* bias wherein humans tend to weigh losses more than gains and a *certainty effect* wherein humans tend to favor more certain outcomes over outcomes that are less likely [41]. To apply prospect theory, we therefore also require an uncertainty assessment over the price forecasts that we make. We do this by training additional models on the same data but after perturbing the data by introducing errors in it: These errors are generated through an error model that we describe in later sections, and thus a range of prices is forecasted for each day. We then make recommendations for the best three dates to sell within the next $k \in [0..29]$ days that can fetch better gains as compared to selling today. If, for example, prices are forecasted to fall in the coming days, then we recommend to sell immediately, or if prices are forecasted to rise and then fall, then we recommend to sell when the gains are likely to be the highest. New recommendations can then be obtained

Table 1. Percentage of Available Data for Main Markets during 2006 to 2021

State	Mandi	Available Date(%)
Rajasthan	Kota	73.45
	Bhawani	59.16
Madhya Pradesh	Sonkatch	50.99
	Ratlam	67.18
	Mahidpur	57.71
	Ganjbasoda	58.83

on this day on when to sell in the next upcoming 30 days, and thus the horizon can keep getting extended if the prices keep rising. Prospect theory helps make these recommendations by also factoring the uncertainty in the price forecasts.

Finally, the recommendations and forecasted prices, along with past prices, are pushed to Google Firebase and visualized on the Android application. The application allows users to also select different markets and past years to make comparisons and observe the trends by browsing data on their own.

4 DATA PREPARATION

4.1 Dataset for Soyabean Prices and Arrivals

We obtain a time series of soyabean prices from the Agmarknet website operated by the Government of India, which makes available data about the daily prices and arrival volumes for more than 2,000 varieties of about 300 commodities traded in 2,700 small and large markets in India. We scrape this data from January 1, 2006 to December 31, 2021. Since a lot of the data is missing, we were able to choose only a few markets for building time forecasting models. We first identified the largest states for soyabean production during the Kharif season, which are Rajasthan and Madhya Pradesh, responsible for 49.91% of Kharif soyabean production.⁵ From these states, we then selected markets for which at least 50% of the data was available, and there was no continuous period of missing data for 60 days or more, as shown in Table 1.

4.2 Imputation through ARIMA

We use ARIMA to impute missing data for the time series of the above markets. ARIMA is a statistical time-series forecasting method that includes an auto-regressive component based on its own lagged values and a moving average component based on the error in its previous forecasts. We chose ARIMA instead of standard interpolation methods so as to not take future values into account while imputing. The imputation happens as follows. A time series can be decomposed into alternating intervals of available and missing data as follows: $(A_1, M_1, A_2, M_2, \dots)$, where A_i is a consecutive sequence of days with no missing data and M_i is an interval of consecutive days of missing data. We start with using A_1 as an input to the ARIMA model to impute values for the M_1 interval, then the entire sequence of A_1 followed by the interpolated M_1 interval and the actual A_2 interval to impute values for M_2 , and so on. Figure 2 shows an example of an original time series with missing data, and the ARIMA interpolated time series. The same method is used for imputing both the price and the arrival time series for various selected markets.

⁵<https://www.sopa.org/statistics/soyabean-production-by-state/>

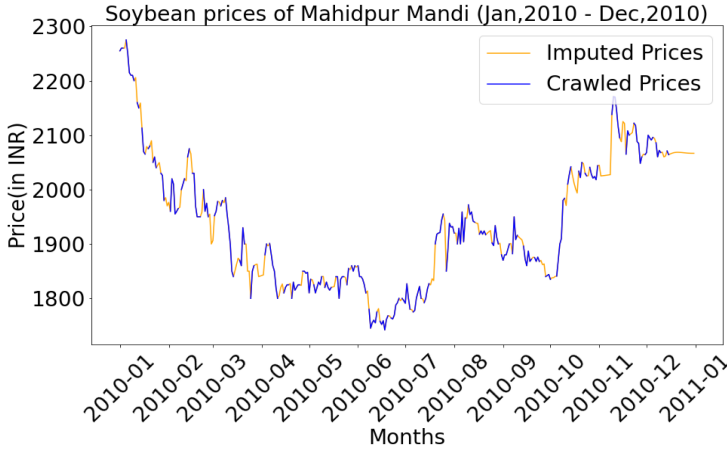


Fig. 2. Imputation of missing prices using ARIMA for Mahidpur Mandi (2010).

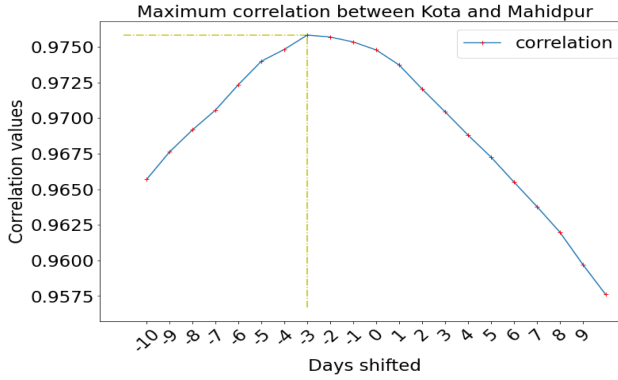


Fig. 3. Relation between lag days and correlation between Kota and Mahidpur.

4.3 Surrogate Market Analysis

Prior work on an analysis of onion prices in India identified a high correlation between prices at different onion markets in India [30]. Such reduced price dispersion across markets due to better communication between market participants is a well-documented feature in agricultural markets [23]. A similar behavior persists in soyabean markets, and we use it to develop a multivariate price forecasting strategy by using data from another market to help forecast prices for a given target market. For a target market, we identify a surrogate market that typically leads it and use the data of this surrogate market as an exogenous variable to forecast prices for the target market. We identify such market-surrogate pairs as follows. We compute correlations between a pair of markets by shifting one of their time series in a range of -10 to $+10$ days, as shown in Figure 3 for an example of the Kota and Mahidpur pair of markets. The correlations peak when the Kota market was shifted back by 3 days, showing that the Mahidpur market leads it by 3 days, and can be used as a surrogate market to forecast prices for Kota. We compute peak correlations for all pairs of markets, and the shift in days corresponding to this peak correlation, to identify (target market, surrogate market) pairs. Table 2 shows these peak correlation values, and Table 3 shows the lead/lag number of days. A negative value in position (i, j) indicates that market j (column) leads market i (row), and thereby the time series of market j can serve as the surrogate market for market i .

Table 2. Peak Correlations between All Pair of Selected Markets, Computed over the Period of 2006–2021

Market i Market j	Ganjasoda	Mahidpur	Ratlam	Sonkatch	Bhawani	Kota
Ganjasoda	1	0.98	0.97	0.95	0.98	0.98
Mahidpur	0.98	1	0.97	0.94	0.97	0.98
Ratlam	0.97	0.97	1	0.95	0.97	0.98
Sonkatch	0.95	0.94	0.95	1	0.94	0.95
Bhawani	0.98	0.97	0.97	0.94	1	0.98
Kota	0.98	0.98	0.98	0.95	0.98	1

Cells in bold indicates market i corresponding to each market j with highest correlation value.

Table 3. Lead/Lag Values between All Pair of Selected Markets, Computed over the Period of 2006–2021

Market i Market j	Ganjasoda	Mahidpur	Ratlam	Sonkatch	Bhawani	Kota
Ganjasoda	0	0	1	2	0	−4
Mahidpur	0	0	0	0	−3	−3
Ratlam	−1	0	0	1	−3	−5
Sonkatch	−2	0	−1	0	−4	−8
Bhawani	0	3	3	4	0	−2
Kota	4	3	5	8	2	0

We thus identify three market-surrogate pairs for subsequent analysis: Kota (Rajasthan) and Mahidpur (Madhya Pradesh), Bhawani (Rajasthan) and Ganjasoda (Madhya Pradesh), and Ratlam (Madhya Pradesh) and Sonkatch (Madhya Pradesh).

4.4 Satellite-based Data

Along with time series of prices and arrivals, we also performed experiments by incorporating time series of multiple satellite data-based indexes into our forecasting models.

As specific plot markings of fields where soyabean is grown are not available in India, we compute aggregate satellite-based indexes over the entire sub-district⁶ in which a market is located. Since the temporal resolution of most satellites is of the order of a few days, we use linear interpolation to obtain daily time series. The time series are computed on the Google Earth Engine [17] and downloaded for local computation in the forecasting models.

Table A.1 lists the indexes that we evaluated. We used standard vegetation indexes such as the **Normalized Difference Vegetation Index (NDVI)** and the **Enhanced Vegetation Index (EVI)**, which reflect the health of the vegetation; *Soil Moisture*, which indicates the wetness of the soil at different depths; *Temperature*, indicating the day and night land surface and soil temperatures; *Precipitation*, which gives the daily precipitation; and so on.

4.5 News Data

Our reading of reports on soyabean markets and discussions with CCD brought forth several factors that influence the market prices, such as unseasonal or erratic rainfall, which can damage the

⁶Sub-districts span about 15 km × 15 km each, and there are approximately 6,000 sub-districts in India, corresponding roughly to one market per sub-district.

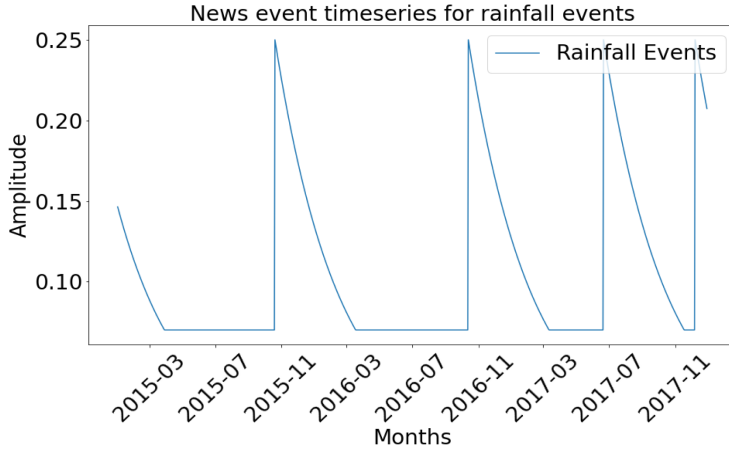


Fig. 4. Rainfall event time series over the years January 2015 to December 2017.

crops; larger area sown under soyabean that can result in supply gluts; global soyabean production and currency exchange rates; and so on. To build an exhaustive list of such factors, we used a news article crawler that has built a corpus of all articles published since 2011 in six major English national newspapers in India [32, 36] and searched it for all articles related to soyabean prices. We then coded a random sample of 180 articles and identified five types of events that tend to have a short-term effect (approximately over a month) on soyabean prices and another five types of events that have a longer-term effect. For example, events such as hoarding are typically considered as causing short-term disruptions, whereas events such as deficient rainfall have longer-term repercussions. Table A.2 shows these event types. We further used this as ground truth to build a document classifier to identify events. We experimented with a simple keyword-based classifier we constructed based on domain knowledge, an SVM classifier, and a BERT-based model. The classifiers performed similarly, largely due to the small size of our dataset, and hence we use keyword-based classification in our work. This can be substituted with better event detection models on streaming text data if adequate training data are available [13].

For each event, we also marked the onset date based on the first time the event was mentioned in the news. We then modeled the effect of each event as a sudden shock onset that then decays exponentially over its effect duration. This allowed us to construct time series for each type of event with the effect of each event modeled through a *decay function* $(a, \alpha, t) = a \times e^{-\alpha \times t}$, where a represent the maximum amplitude of the event at the time of onset, α decides the decay duration⁷ based on an event type, and t is the t th day from the date of occurrence of the event. For illustration, Figure 4 shows the event time series for rainfall. For simplicity, we assumed the maximum amplitude of each event onset to be the same.

5 METHODOLOGY AND EVALUATION FOR MAIN MARKETS

5.1 Price Forecasting

We first experiment with different price forecasting methods and choose the one that gives the least RMSE over prices forecasted for the next 30 days. Prior work has shown that neural network architectures such as LSTM work better on price forecasting tasks than statistical auto-regression-based methods [24]. We therefore evaluate two architectures, LSTM and TCN, in univariate and

⁷The decay duration is assigned as 30 days for short-term events and 150 days for long-term events.

Table 4. RMSE Scores for LSTM and TCN Models (Computed over All 30-day Intervals from 2014 to 2020)

RMSE score (Years 2014-2020) Mandi	Univariate		Multivariate							
	LSTM	TCN	LSTM				TCN			
			Surrogate Mandi		Arrivals	Cyclic Feature	Surrogate Mandi		Arrivals	Cyclic Feature
Kota	317.78	218.95	Mahidpur	255.81	303.56	528.73	Mahidpur	187.05	238.31	379.98
Raltam	326.38	207.94	Sonkatch	231.61	299.96	495.42	Sonkatch	197.71	214.09	339.51
Bhawani	272.58	252.53	Ganjbasoda	232.15	288.44	380.92	Ganjbasoda	238.38	246.93	460.99

Cells in bold indicates which model variant performs best in terms of RMSE score.

multivariate settings. Both the LSTM and TCN architectures take the last 60 days of time series as input data and forecast for the next 30 days. The univariate architectures work on the price time series of the same market, whereas the multivariate architectures use exogenous time series such as arrival volumes in the same market, prices in surrogate markets, satellite data-based time series in the market sub-district, news data, and so on. We also evaluate a variant where we include an additional feature of the day of the year for which the forecast is being produced. Since the prices follow seasonal trends, the day of the year is converted to cyclic features by distributing the 365 days of a year over 360° or 2π radians and representing it with the corresponding sine and cosine values [26]. This method allows us to efficiently incorporate seasonality in the forecasting model without having to include long time series of historical data that can result in very large training times.

The LSTM architecture we use has one input layer, then a hidden layer, followed by a dense output layer. The input layer has 32 units with *tanh* activation functions and a dropout rate of 0.3%. The hidden layer has 32 units with a dropout rate of 0.1%. The output layer is densely connected with the previous hidden layer and outputs forecasted prices for the next 30 days.

TCN can be considered as an improvement over LSTM in the sense that it allows the model to learn arbitrary dependencies, over different timescales, on different parts of the input sequence. A TCN network consists of three parts: an input tensor (of multiple time series of past data, in our case), a series of residual blocks with varying dilation factors, and a final output tensor that connects to a dense fully connected layer to produce the final output (30-day forecasts into the future, in our case). Figure 5(a) shows this architecture. Figure 5(b) explains the notion of dilations. For forecasting, convolutions need to be done causally, i.e., any convolution at time t should only use past data until t . A dilation factor controls how far back in time should these convolutions range and a kernel parameter controls the length of the convolution. In Figure 5(b), the dilation factors are increased exponentially, with a kernel size of 2.

The dilations are done within residual blocks shown in Figure 6(a)–(c). Each residual block performs a dilated convolution, weight normalization, activation, and dropout; this is done twice for stability, and the output is fed to the next residual block. In our model, we have used a kernel size (k) of 3, number of filters (n_f) as 64, and dilations as (1, 2, 4, 8, 16, 32). Input channels ($n_i - 1$) are equal to the number of exogenous variables used with the models. For example, with one additional exogenous variable, input to the model will be of shape $1 \times 60 \times 2$ for the past 60 days as input, and the output generated from the model will be of shape $1 \times 30 \times 1$ for the next 30-day forecast.

The dataset we use for evaluation consists of a total of 5,478 data points (January 1, 2006 to December 31, 2020) of soyabean market prices time series. For each architecture, we first start with fitting a model over data from January 1, 2006 to December 31, 2013, i.e., over 2,921 data points. We take intervals of 90 days as our training samples, where the first 60 days serve as input and the next 30 days serve as output. The trained model is then used to produce a final forecast for the 30-day period of January 1, 2014 to January 30, 2014, and the RMSE is calculated against the actual prices during this interval. We then update the model by training it on this next 30-day

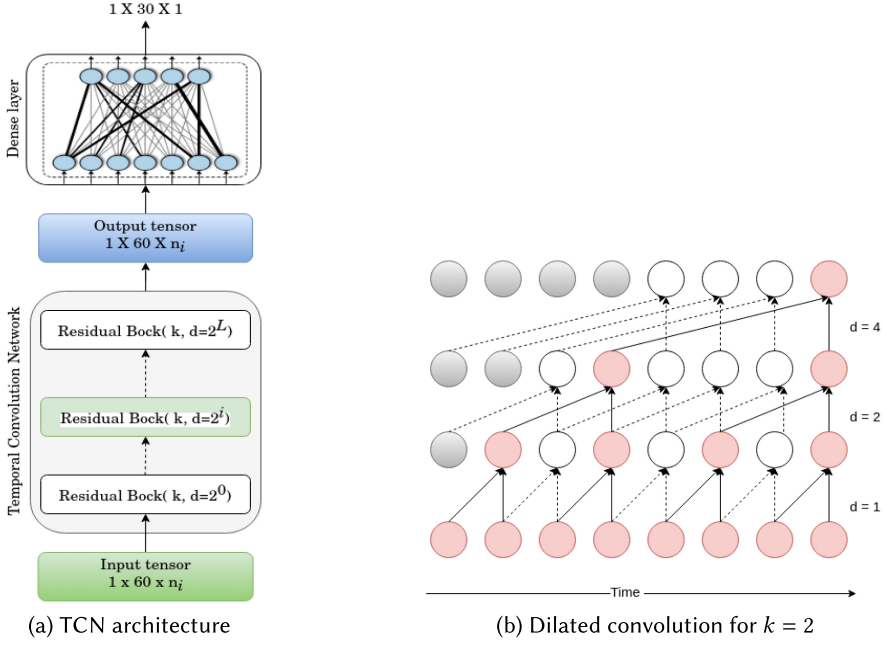


Fig. 5. TCN architecture and dilated convolution.

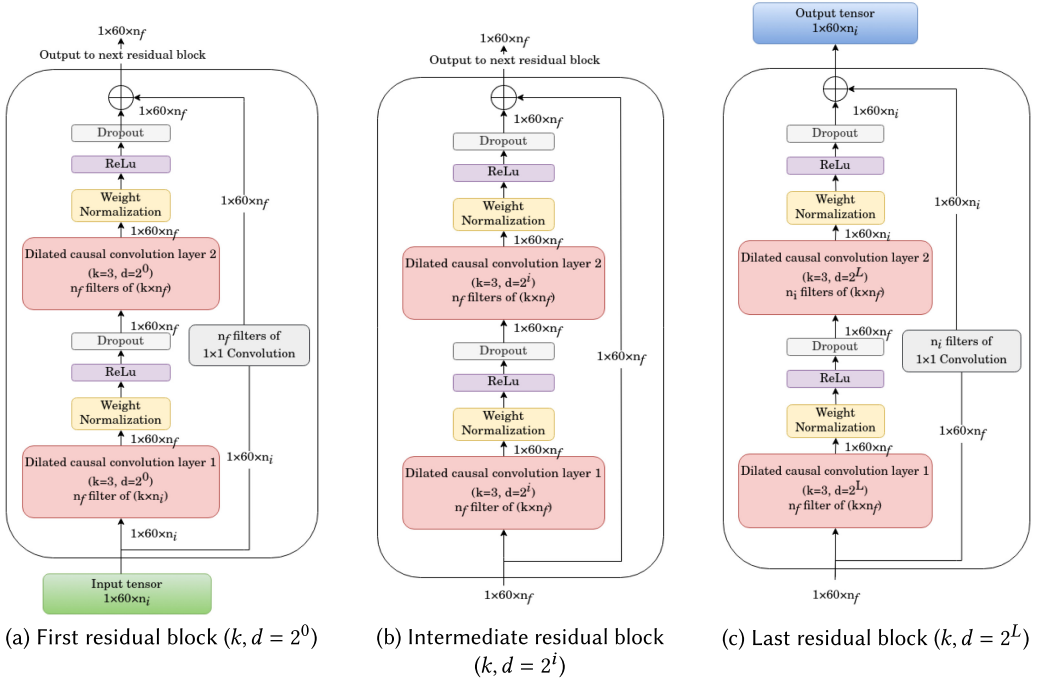


Fig. 6. Residual blocks in temporal convolution networks.

Table 5. RMSE Scores for LSTM and TCN Models (Computed over 30-day Intervals during the Soyabean Season (September to January) from 2014 to 2020)

RMSE score (Seasons 2014-2020) Mandi	Univariate		Multivariate							
	LSTM	TCN	LSTM				TCN			
			Surrogate Mandi	Arrivals	Cyclic Feature		Surrogate Mandi	Arrivals	Cyclic Feature	
Kota	273.82	214.16	Mahidpur	225.56	286.18	479.12	Mahidpur	194.79	222.34	381.71
Raltam	360.65	213.85	Sonkatch	243.90	291.67	424.12	Sonkatch	228.49	228.89	307.97
Bhawani	229.39	245.33	Ganjbasaoda	182.54	264.96	313.70	Ganjbasaoda	226.05	235.12	457.78

Cells in bold indicates which model variant performs best in terms of RMSE score.

period until January 30, 2014, i.e., $2,921 + 30$ data points. The learning and output process is then repeated: A forecast is produced for the next 30 days, from January 31, 2014 to March 1, 2014, and the RMSE is calculated. The dataset is then again extended by 30 days, followed by updating the model to now include $2,921 + 30 + 30$ data points, and so on. An average RMSE is calculated over all the 30-day forecasts that are made.

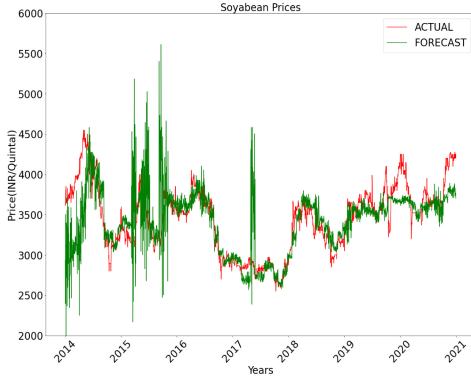
Outputs for a few models are shown in Figure 7, and Tables 4 and 5 show the performance of different architectures. Table 4 shows the RMSE scores computed over all 30-day intervals, and Table 5 shows the scores computed only over intervals in the soyabean season (September to January), i.e., the actual period of interest to us since soyabean is only planted during the Kharif season in Adilabad and is sold by January. We compare the univariate LSTM and TCN models with a range of multivariate models using exogenous time series: surrogate market prices, the arrivals volume, and cyclic features for seasonality. In general, we find that TCN multivariate models using surrogate markets work the best. We also show in Figure 8 the yearwise performance of annual and seasonal RMSE scores. Here, too, we find that the TCN multivariate models with surrogate markets work well for most years, and we use this architecture for subsequent analysis. We further examine the robustness of these models for different months (shown in Figure A.1 in the Appendix) to determine if there are any particular months when they do not work well and find the TCN multivariate model to be reasonably robust.

We also evaluated the models with using satellite data and news data time series. The results are shown in Table A.3. We found that these time series were not able to improve the performance beyond that given by using surrogate market prices. We believe that our strategy of surrogate markets works well, because their time series would be capturing several more unobserved variables beyond the ones such as satellite data, news data, or seasonality using cyclic features that we tried to deliberately add. We show later that the satellite, news data, and cyclic seasonality related time series do indeed help but in making longer-term forecasts over 12 weeks rather than for the short-term 30-day forecasts that we have discussed here.

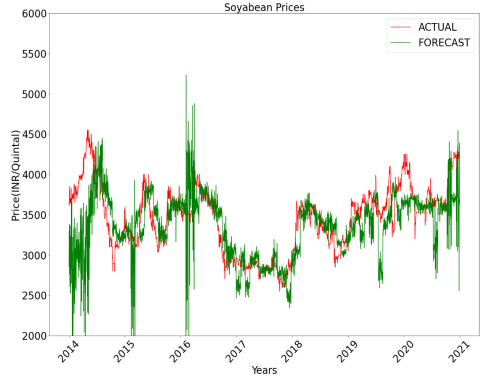
5.2 Sale Recommendations

5.2.1 Recommendation Methods. We next outline different methods to produce sale recommendations using the 30-day price forecasts generated by the TCN multivariate time-series model. We experiment with two strategies: a simple strategy that recommends the day with the highest price forecast in the coming days and a method using prospect theory. With both strategies, we try two variants with and without perturbed models, i.e., we train multiple models by introducing errors in the training data to get uncertainty estimates for the forecast of each day. For evaluation purposes presented here, we compare these methods on the best sale recommendation, but in the implementation we generate the top three sale recommendations over the next 30 days. We next describe these methods and then outline evaluation metrics to compare them with one another.

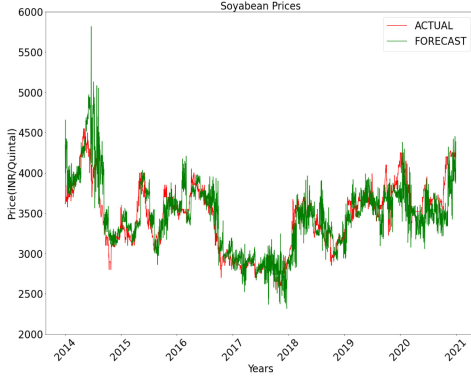
Note that since we produce forecasts each day, for any day $k \in [0..29]$ in the future, there are $30 - k$ forecasts that are produced. For example, for $k = 10$, the first forecast would have been



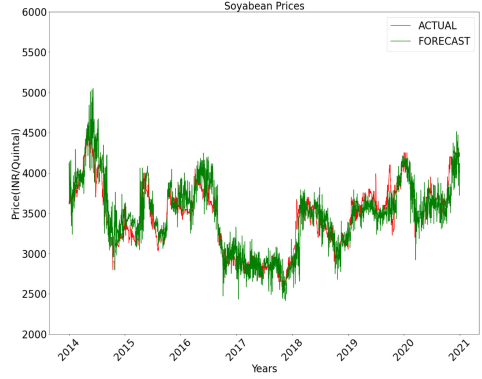
(a) LSTM Univariate



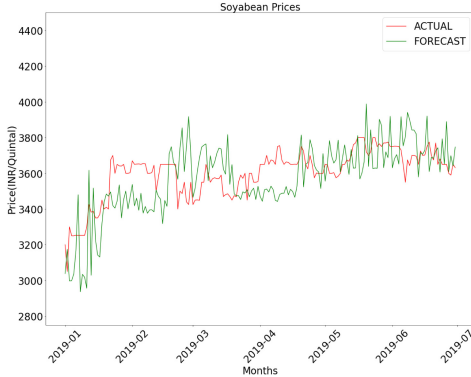
(b) LSTM Multivariate



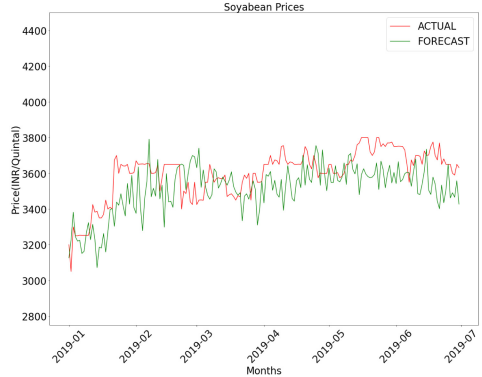
(c) TCN Univariate



(d) TCN Multivariate



(e) TCN Univariate



(f) TCN Multivariate

Fig. 7. (a)–(d)) Forecast plot for LSTM Uni, LSTM Multi, TCN Uni, and TCN Multi for Kota market 2014–2020. (e) and (f)) Zoomed-in forecasts for TCN Uni and TCN Multi for Kota market (January to June 2019).

produced 20 days ago, another forecast 19 days ago, and so on, with the final forecast produced yesterday using the actual price of yesterday. This enables us to obtain a range of forecasts for each day in the future as shown in Figure 9, with earlier days having more forecasts, and we take the mean forecasted price of each day to produce the sale recommendations.

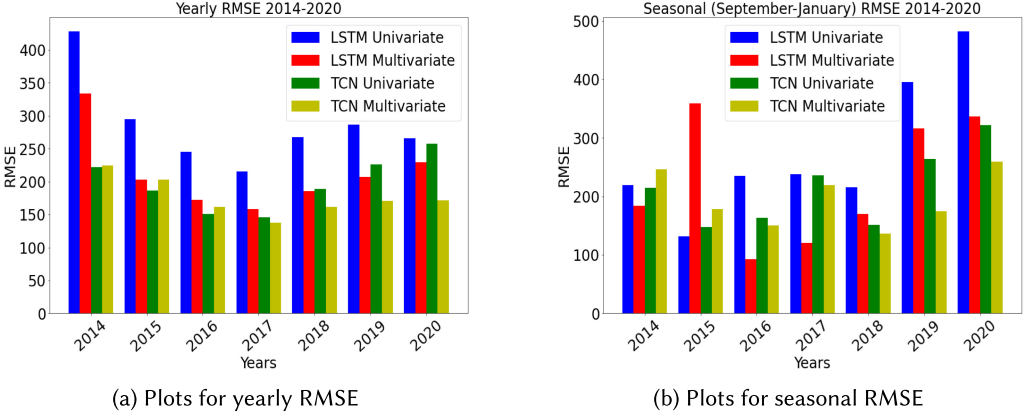


Fig. 8. Annual and Seasonal RMSE scores (computed from 2014 to 2020).

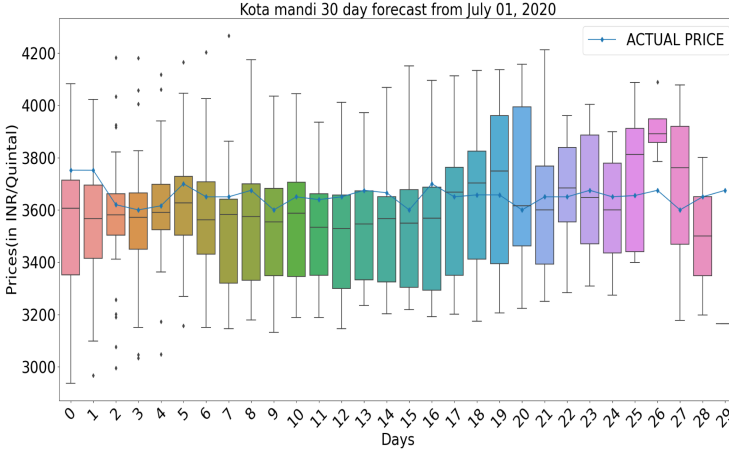


Fig. 9. TCN multivariate range-based forecast for 30 days.

Maximum price strategy: This simple strategy simply recommends the day that is forecasted to get the best price over the next 30 days, including today. As described above, the mean price for each day calculated over the range of forecasts for that day is used here.

Prospect theory-based strategy: Prospect theory [41] is a behavioral model that approximates how people decide between alternatives that involve risk and uncertainty. Risk is modeled based on the *loss aversion* bias in humans wherein people tend to weigh losses more than gains. Uncertainty is modeled based on the *certainty effect* wherein people tend to favor more certain outcomes over outcomes that are less likely. This is shown diagrammatically in Figure 10. The value function $v(x)$ rises less slowly for positive outcomes than it drops for negative outcomes, to illustrate the loss aversion bias. The function is also assumed to be concave in the region of gains to illustrate diminishing marginal utility but convex for losses, indicating that people view higher losses as more significant than higher gains. The weighting function $w(p)$ related to the probability of occurrence of an event illustrates the certainty effect wherein people prefer more certain outcomes even if they may result in a lower gain over less certain outcomes that may lead to a higher gain. The final value function in prospect theory is then defined as $V(x, p) = w(p)v(x)$. Behavioral

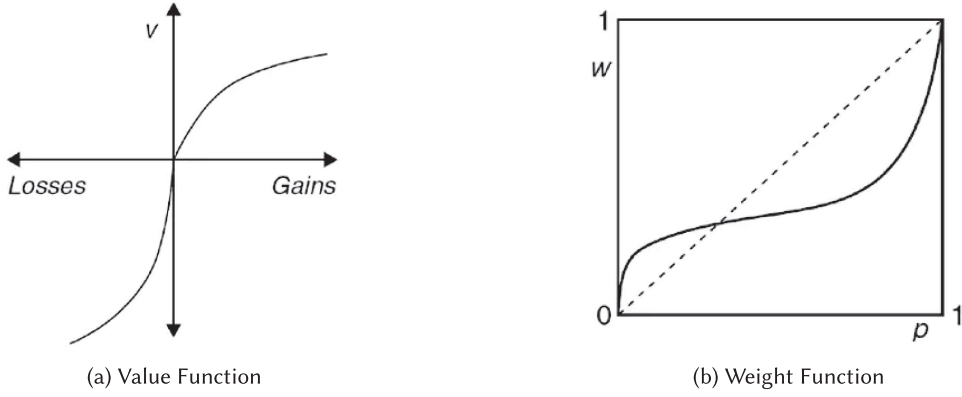


Fig. 10. Value and weight function (prospect theory).

experiments have been conducted in numerous geographies and cultures, in different domains, to empirically determine the weight and value functions. We use the function definitions determined by Abdellaoui et al. [1, 2] as their modeled scenario is similar to our context to evaluate prospects defined by (g_k, p_k, l_k) , where a positive outcome g_k can happen with probability p_k and a negative outcome l_k with probability $1 - p_k$. In our case, the prospect of selling on any day $k \in [0..29]$ needs to be compared with the prospect of selling on day $k = 0$, i.e., today. Abdellaoui et al. suggest comparing these prospects using $w(p_k)v(g_k) + w(1 - p_k)v(l_k)$, where g_k is the expected gain of selling on day k over today, l_k is similarly the expected loss, and p_k is the probability for the price on day k to be more than the price today. p_k can be calculated simply by counting the number of forecasts for day k that are higher than the price today. Similarly, g_k can be calculated as the mean price of forecasts for day k that are higher than the price today and conversely for l_k as the mean price of forecasts on day k that are less than or equal to the price today. Abdellaoui et al. then present empirical results to suggest the following value and weight functions:

$$v(x) = \begin{cases} x^\alpha & x \geq 0 \\ -\lambda * (-x)^\beta & x < 0. \end{cases} \text{ where } \alpha = 0.86, \beta = 1.06 \text{ and } \lambda = 2.61, \quad (1)$$

$$w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1 - p)^\gamma} \text{ where } \delta = 0.98, \gamma = 0.83. \quad (2)$$

We use these function definitions to identify the best prospect, i.e., day k on when to sell. Figure 11 demonstrates the overall process. To the top is shown an illustrative example of the range of forecasted prices on each day. The reference price is that of today; day-1 is an example where the price is likely to be near that of today; on day-2 the spread for prices is wider; day-3 is when prices are likely to be higher than today; day-4 is similarly likely to see a higher price but with a wider spread and similarly for day-5 and day-6 when the prices are likely to be less than today. The probability p_k for each day is calculated as the probability of the price being higher than today, using which the weight function $w(p_k)$ is calculated based on Equation (1). Similarly, g_k and l_k are calculated as the expected gain and loss respectively of the prices being higher or lower than today. These are used to compute $v(g_k)$ and $v(l_k)$ based on Equation (2). Finally, the prospect value is computed, shown to the bottom left of Figure 11, and a graphical representation using green and red is developed that is displayed in the smartphone application. Here green indicates gain and red indicates loss, with the size of the bars indicating the expected value of the

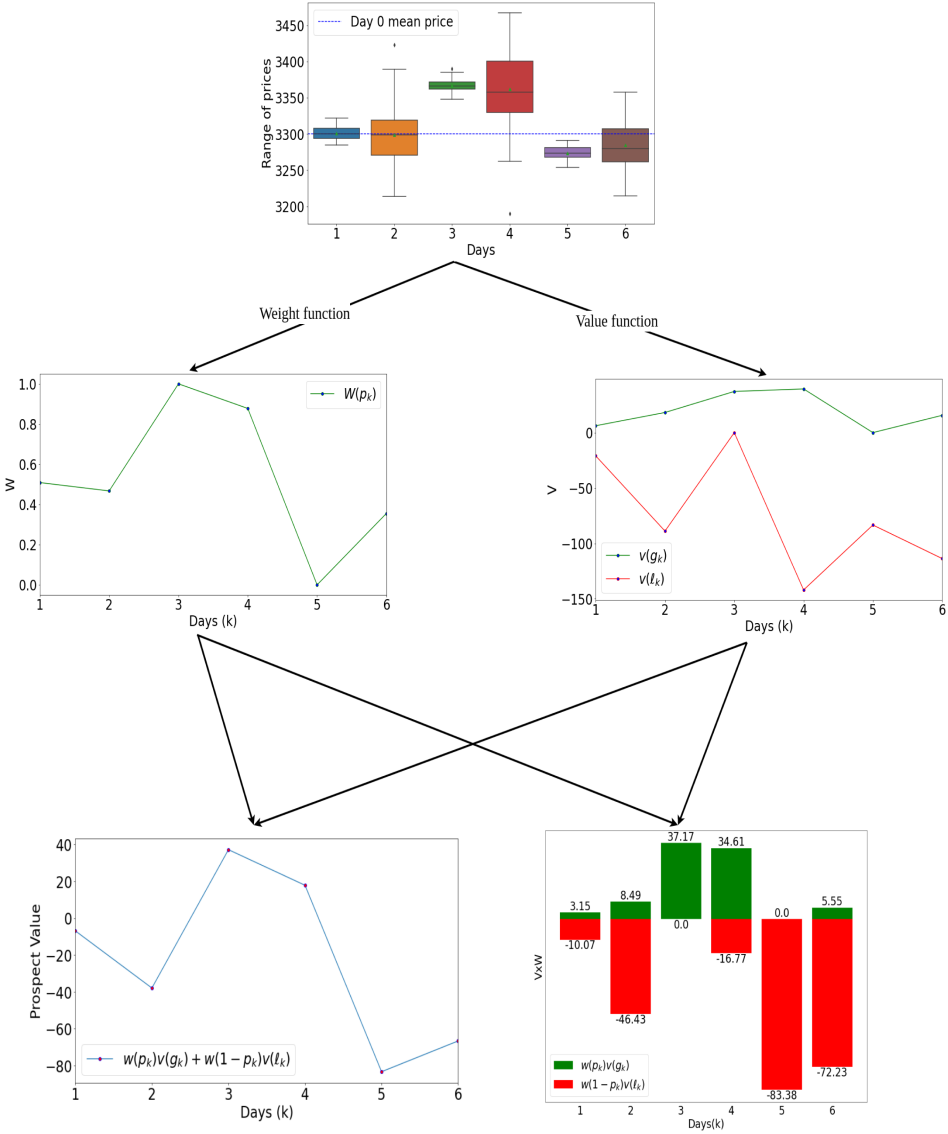


Fig. 11. Example to demonstrate the computation of prospect values.

gain or loss. A presence of both red and green bars indicates that the model is less certain of the gain or loss.

Perturbed models: To improve the robustness of the recommendations produced by the above methods, we further model uncertainty as follows. In addition to the TCN multivariate price forecasting model trained using all historical data as described above, we train four additional models with errors deliberately introduced in the training data. Recall that the training data are time series of a sequence of available and missing data intervals $(A_1, M_1, A_2, M_2, \dots)$, where the missing data intervals M_i are imputed using ARIMA forecasts produced from previous data $(A_1, M_1, \dots, A_i)$. We randomly shuffle the available and missing data periods on the imputed time series. We do not alter the time series itself; we only move around the periods when data are available or missing so

that the distribution of available and missing data remains invariant. We then remove data from the newly assigned missing periods and impute the values. We do this four times independently to develop four new training datasets and learn models for each one. We then use these additional perturbed models to also generate forecasts. In this way, for any day k in the future, we now have $5 \times (30 - k)$ forecasts. We use these forecasts in both the recommendation methods above.

5.2.2 Evaluation Metrics. We first outline some notation that we use to develop evaluation metrics for the various recommendation methods. For each day in our test-set, we generate recommendations for the next 30 days, and thus for n days in the test-set we examine the quality of the recommendations produced for all these days over their next 30-day intervals. For each interval i , $a_{i,k}$ is the actual price on day k of the interval, $r_i \in [0..29]$ is the day that was recommended for the sale in the i th interval, and thus $a_{i,0}$ represents the price on the first day of the interval and a_{i,r_i} is the price that was realized on the recommended day. δ_i is an indicator function equal to 1 if $a_{i,r_i} > a_{i,0}$, i.e., the actual price on the recommended day was higher than the first day, and 0 otherwise. We define three primary recommendation metrics as follows.

Probability of Accurate Predictions (PAP): This is the percentage of times out of n intervals that a method gave good recommendations, i.e., $\delta_i = 1$, indicating the chance for cooperatives to get a better price if they followed the recommendations instead of selling today, i.e., at the start of each interval,

$$PAP = \left(\frac{1}{n} \sum_{i=1}^n \delta_i \right) \times 100.$$

Net Gain (NG): Measured in INR per quintal, this measures the potential gain for cooperatives if they followed the recommendations instead of selling today,

$$NG = \sum_{i=1}^n \frac{a_{i,r_i} - a_{i,0}}{n}.$$

Net Gain by an Oracle (NG-Oracle): This measures the maximum possible gain for cooperatives if they knew the future, i.e., the day the actual price would be the highest. The difference between NG-Oracle and NG would thus represent a potential scope for improvement in the recommendation and forecasting methods to do better,

$$NG - Oracle = \sum_{i=1}^n \frac{\max_{k=0..29} (a_{i,k}) - a_{i,0}}{n}.$$

5.2.3 Results. Table 6 compares the different recommendation methods using the three metrics described above for the Kota market during the soyabean season. Results for other markets are given in Tables A.4 and A.5 in the Appendix. In general, we find that prospect theory using perturbed models work best, and since it is a behavioral model it is also likely to be appreciated more by the users. PAP scores are often more than 90%, indicating that most times the sale recommendations are able to suggest a date that actually fetches a better price than today. The difference between NG-Oracle and NG, however, indicates a wide scope for improvement. NG sometimes goes up to INR 100 per quintal and mostly remains positive but can also be negative at times and typically is between 20% to 50% of NG-Oracle. We also see that NG performance in some years is exceptionally poor for all methods. We therefore examine a few such years in more detail.

Figure 12 shows the forecasted and actual prices for 2015–2016 and 2020–2021 when the recommendations are poor and for 2014–2015 and 2017–2018 when the recommendations are better. The horizontal bars show the recommendations—the start of the bar indicates the date when the

Table 6. Seasonal (September–January) Recommendation Metrics for Kota, Rajasthan 2014–2020

Years	Recommendation Metric	With Perturbation		Without Perturbation	
		Prospect Theory	Max Mean Forecast	Prospect Theory	Max Mean Forecast
2014	PAP	92.81	77.12	81.05	67.97
	NG	26.4	26.3	7.07	7.86
	NG Oracle	157.69			
2015	PAP	77.12	42.28	49.67	45.1
	NG	20.41	29.32	22.05	18
	NG Oracle	195.98			
2016	PAP	73.86	55.56	68.63	52.29
	NG	−13.42	−17.4	−21.05	−27.23
	NG Oracle	67.39			
2017	PAP	92.16	75.16	78.43	72.55
	NG	92.48	123.96	108.24	124.34
	NG Oracle	221.79			
2018	PAP	84.97	60.48	81.7	67.97
	NG	28.59	20.65	63.54	33.04
	NG Oracle	192.73			
2019	PAP	94.12	59.48	71.9	53.59
	NG	40.97	−6.36	−3.42	−35.53
	NG Oracle	253.21			
2020	PAP	99.35	79.61	90.13	84.21
	NG	−0.01	76.46	110.9	130.66
	NG Oracle	268.87			

Cells in bold indicates the model with the top score for the specified metric within the particular year.

recommendation was made, and the end of the bar indicates the recommended date. Short bars imply that sales should be made soon, in cases where prices are likely to fall. Long bars imply that sales should be delayed, when prices are likely to rise. The poor recommendation performance for 2015–2016 and 2020–2021 seems to be arising from a poor quality of the price forecasts. During 2015–2016, news articles point to a shortage in soyabean arrivals of up to 48% in some markets due to coordinated hoarding of stocks by large farmers and traders [15], which resulted in a rapid increase in soyabean prices during September and October itself, which is rare. During 2020–2021, crop damage in Maharashtra similarly led to an increase in prices in October itself [8], combined also with dry weather in Brazil and Argentina, other countries that also produce a lot of soyabean, which led to high prices globally. In contrast, for years like 2014–2015 and 2017–2018, the TCN models were able to follow the trends much better and were able to produce better sales recommendations. News articles did indicate some interesting events that occurred during these periods. Rainfall was low in 2014–2015 [42], but it did not cause any anomalous price increases, because the Indian rupee strengthened at the same time, and therefore exports fetched better prices [4]. In a similar way, 2017–2018 had lower production but substitute imports were brought in as well, which suppressed the prices [14].

The recommendation performance clearly depends heavily on the forecasting performance. Further, since prospect theory requires good uncertainty estimates, and results on the perturbed models are better, we also evaluate a suitable re-training frequency for the models to minimize

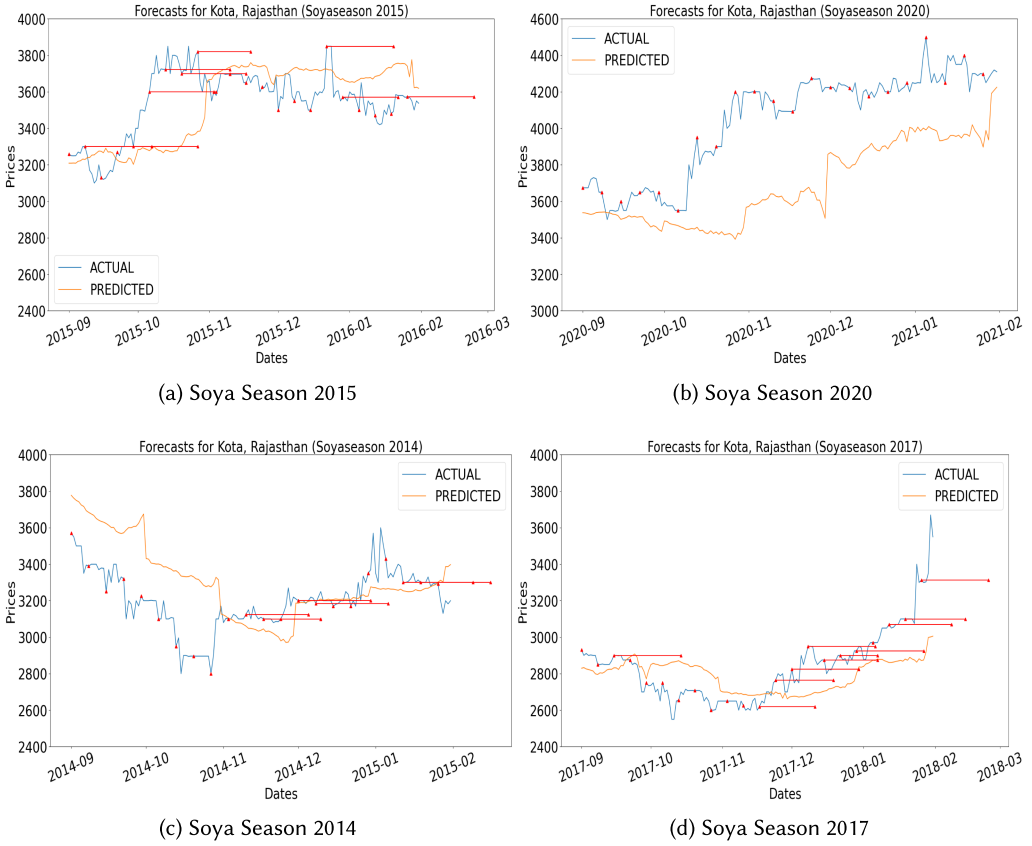


Fig. 12. Forecasts of soyabeen prices for Kota, Rajasthan over various years.

computational overhead. We find that a re-training frequency of 15 days is suitable. The NG and PAP scores are close enough to those obtained with retraining the models each day or every 7 days.

5.3 Transfer Learning

Finally, given the lack of data for the Adilabad market on the Agmarknet website, we had to develop a method to use the models trained on the main soyabeen markets to produce forecasts for Adilabad. During the first year of our study in 2021, we were able to deploy an ODK-based data collection application for the CCD field officers to provide daily market prices for Adilabad from November 2021 onward. This dataset spanned 90 days and was supplemented with additional data of 305 days from previous years available for Adilabad on the Agmarknet website. We used this data for a transfer learning approach as follows.

We were able to divide this data into 39 non-overlapping intervals of 90 days each, none of which had missing data for more than 15 consecutive days. The missing data were imputed using the ARIMA method as describe earlier. These 39 intervals were divided for a k -fold cross-validation with $k = 5$, where four folds were used to fine-tune the models trained on the main markets and the fifth fold was used for testing. Table 7 shows the results for the three market-surrogate pairs, with and without fine-tuning. The Bhawani-Ganjbasoda model after fine-tuning on the Adilabad data gave the least RMSE score for Adilabad. We therefore used this model for the field implementation.

Table 7. RMSE Scores for the Adilabad Market Produced on Models Trained for the Main Soyabean Markets

Metric Surrogate Market Pairs	RMSE	
	Without Finetuning	With Finetuning
Kota + Mahidpur	203.00	190.50
Ratlam + Sonkatch	215.20	218.43
Bhawani + Ganjbasoda	215.37	187.41

Cells in bold indicates the pair of market for which the multivariate TCN model has the lowest RMSE score.

6 FIELD TESTING

Our study progressed in two phases. During the first year of our study, we spent time understanding the operations of CCD, the federation, and the cooperatives to clearly define the sales recommendation use-case. Since price data for Adilabad was scantily available, we also commissioned an ODK application for the CCD field officers to collect these data. An additional use-case that emerged was the need for an internal communication system for the federation to communicate with its cooperatives and their member farmers. Since many farmers did not have smartphones, we deployed a tried and tested **Interactive Voice Response (IVR)** system for this purpose [38]. During the second year of our study, we deployed an Android application for the CCD field officers to review the sales recommendations produced through the end-to-end system described in detail in the previous sections. Based on the usefulness of the application assessed by the field officers, we will deploy it for direct use by the federation in the following year. We next discuss our deployment experience through the course of the study.

6.1 IVR-based Communication System

The IVR system allows a white-listed set of users who are identified by their caller-id to record voice messages that are then automatically pushed out to all the registered users through outbound phone calls. Phone numbers of the federation board members and CCD field staff were added to the white-list to send messages, and another ODK form was used to collect phone numbers of the member farmers to register them to the system. Close to 3,000 farmers were registered over the course of the study, and 115 messages were pushed over approximately 100,000 successful calls carrying information about the activities of the federation, minutes of their board meetings, weather advisories, cropping advice, and so on.

Field interactions with several farmers indicated a high degree of satisfaction with this system. Many of them reported that the voice announcements made them feel more connected with the federation and proud of what they had collectively accomplished. The IVR system also supports a feature through which farmers can ask questions that can be answered asynchronously by experts, but this feature was rarely used, because it was not actively publicized by the CCD field officers during their community interactions. Since our primary interest is to evaluate the sales recommendation application, which we describe next, we did not spend a lot of effort on rigorously studying the IVR system.

6.2 Android Application for Sales Recommendations

We built an Android application for the CCD field officers with a view to eventually deploy it for the federation to help cooperatives make better sales decisions. The application was released in October 2022, which is toward the end of the Kharif season when soyabean is harvested. Three groups of authors conducted separate field visits prior to the actual deployment to understand the

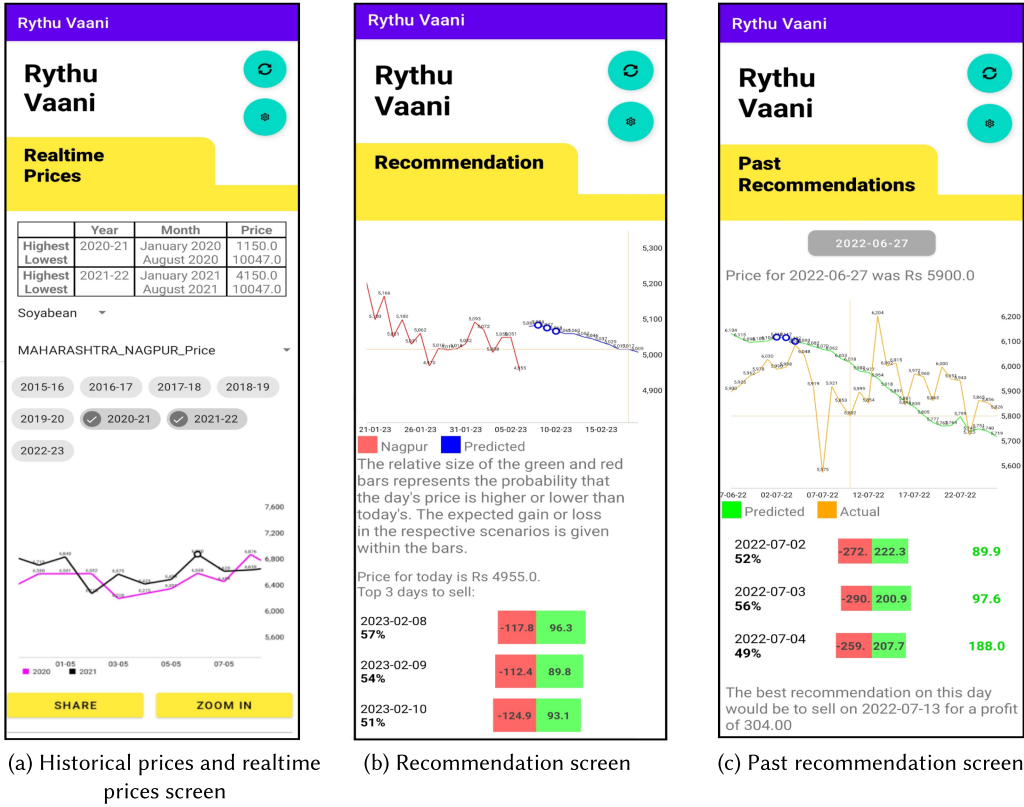


Fig. 13. Rythu Vaani an android application for visualizing realtime prices.

use-case thoroughly and assess the capacity of the federation and CCD field staff to use the envisioned application. We were anxious about the risk of loss to cooperatives because of errors made in the machine generated recommendations, but our interactions convinced us that the field partners had sufficient maturity to make sales decisions without attaching overly authoritative beliefs to the recommendations. The CCD field staff especially had good digital exposure and, as outlined below during the field testing when they were closely involved, they were able to point out several gaps in the application outputs, which assured us that risks arising from incorrect recommendations made by the application were minimal. The intended users seemed capable of factoring these recommendations into their decision making process but not to rely exclusively on them.

The system design for the application is as described earlier in Figure 1, where daily prices are crawled from Agmarknet for the market-surrogate pair of Bhawani-Ganjbasoda, ODK is used to report prices for Adilabad, and the collection of five TCN multivariate models for the Bhawani-Ganjbasoda pair followed by transfer learning to Adilabad are retrained every 15 days to produce daily forecasts for the next 30 days. The prospect theory recommendation method is then used to provide the best three sale recommendations, which are displayed by the Android application. The following main screens were developed, with screenshots also shown in Figure 13:

- *Historical prices:* Users can select to display time series of prices of Adilabad (collected via ODK) and the main soyabean markets (collected from Agmarknet) in different years. Multiple time series are overlayed on the same screen with a view to help users understand historical trends and compare them across markets.

- *Sale recommendations*: A dedicated screen displays forecasted prices for Adilabad for the next 30 days, along with the top-3 recommended dates for sales and the corresponding expected values of the gain and loss on these dates. Green and red bars sized in proportion to the expected gain and loss respectively are meant to provide an intuitive indication about the suitability of the recommended dates, as explained in Figure 11.
- *Past recommendations*: Users can also select any date in the past to review the forecasted prices and recommendations that were given at that time and compare with the actual prices that were seen. This was done with a view to enable users to review the accuracy of the forecasts and recommendations and build their own intuition on how to factor machine generated outputs in their decision making process.

These screens were designed in consultation with the CCD field officers. Inputs from the federation were also sought on a pre-release version of the application during a field training exercise for the CCD field officers. Several suggestions were incorporated, such as to show summary statistics of the highest and lowest price in the month, the ability to share the screens via Whatsapp, switch the display scale between daily and monthly, and so on. The CCD field officers used the application extensively and engaged in weekly review calls to provide feedback to our team about the application outputs. Two sets of feedback were sought from the field officers: whether the application recommendations on a particular day agreed or disagreed with the sales actions taken by the cooperatives on that day and whether the field officers agreed or disagreed with the application recommendations. The field officers closely monitored the activities of 16 cooperatives to report back their actions to us.

Simple visualizations of the time series of prices this year in Adilabad (obtained from the ODK data), the movements in different markets, and being able to compare the trends across markets and years, was reported to be useful by the field officers. In terms of the forecasts and recommendations though, the year 2022–2023 was unique, because soyabean prices remained flat all the way from October 2022 until early February 2023. Figure 14 shows the prices in various markets: Prices were low during September and October and climbed up from November onwards but have stayed flat since then. During these months, the recommendations generally suggested that sales should be made on that day itself. Some cooperatives made partial sales during this period to provide cash to their member farmers, whereas most others decided to wait for the prices to rise. Discussions revealed that most farmers in these cooperatives had planted a cotton crop prior to soyabean, and that had fetched them good prices. They therefore had the wherewithal to wait for the soyabean prices to rise in the long term. Until the end of January 2023, of 16 cooperatives only 7 had sold small amounts of soyabean. For this reason, our feedback discussions with the field officers were not able to bring significant insights about the usefulness and safety of the forecasts and recommendations, because with the prices being flat the decision making of the cooperatives was driven simply by choosing to sell if farmers needed immediate cash and otherwise to keep holding back. This unique experience of 2022–2023 brought forth the relevance of a new use-case, to forecast prices over the next several months. We discuss this in the next section.

News reports vaguely point to some reasons for the soyabean prices remaining sluggish this year. Unseasonal rain led to some amount of crop damage due to which the quality of soyabean produce has been average although the overall quantity did not see any shortfall [35]. Futures trading in soyabean was also banned last year and this year, along with that of several other crops [10], hinting that futures trading by speculators may have inflated crop prices or caused price fluctuations in the past because of which the government decided to impose a ban on it [21]. Although cooperatives such as the ones in our study do not participate in futures trading, large traders and buyers do engage actively but there is heavy debate on its usefulness for farmers and consumers [33].

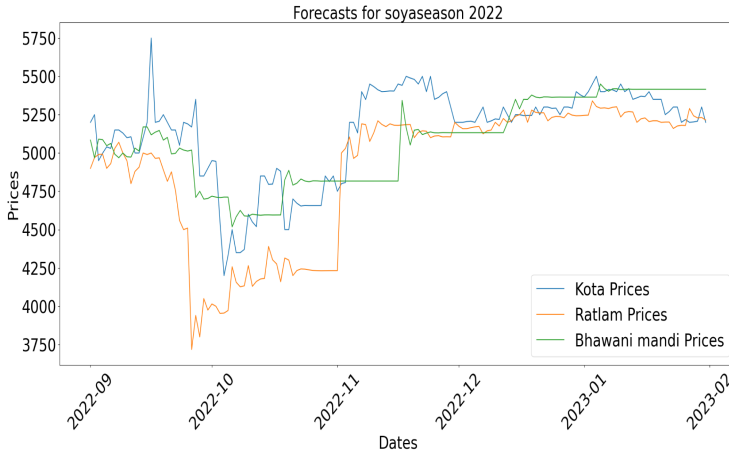


Fig. 14. Soyabean prices across mandis for the 2022–2023 season.

6.3 Long-term Forecasting

With long-term forecasting, we chose to forecast weekly average prices instead of daily prices, for the next 12 weeks. This horizon of approximately 4 months was felt to be reasonable to provide coarse recommendations of how long to wait and then to use the short-term 30-day forecasting model closer to these recommended dates. We used similar LSTM and TCN architectures as for 30-day price forecasting but used for input the time series of the last 52 weeks of weekly average prices to output the weekly average prices for the next 12 weeks. We also used time series of satellite data–based indexes, news data, and cyclic seasonality features, with the hope that these inputs will have a greater bearing on long-term forecasts than what we obtained on short-term forecasts. To compare models, as before we used RMSE for forecasting and PAP, NG, and NG-Oracle for sales recommendations.

Table A.6 in the Appendix shows the results for the Kota market. The following models are evaluated: LSTM/TCN univariate using only the time series for the target market, LSTM/TCN multivariate with surrogate market prices as an exogenous time series, LSTM/TCN multivariate with surrogate market prices and cyclic seasonality features, LSTM/TCN multivariate with surrogate market prices and satellite data–based indexes, and LSTM/TCN multivariate with surrogate market prices and news data–based indexes. We find that although a single model does not dominate but good PAP scores and NG scores are obtained with long-term forecasting, and in many cases the use of satellite data, news data, and cyclic seasonality features do help. The NG scores are more often closer to 50% of the NG-Oracle scores that is better than what we obtained with short-term forecasting. Performance was poor in the same odd years of 2016–2017 and 2019–2020 when soyabean hoarding and crop damage took place. The long-term price forecasting can thus be a viable addition to the Android application: At any time we can present a joint forecast on a daily-basis for the next 30-days along with a dilated forecast on a weekly basis for the next 12 weeks. We are currently in the process of integrating this into the Android application.

7 DISCUSSION AND CONCLUSIONS

We have discussed so far the problem we identified of price forecasting to support cooperatives of small and marginal farmers to make better margins from their soyabean produce. Our price forecasting and sales recommendation methods suggest that modest gains can be achieved in earnings if these machine generated recommendations are followed. Feedback from field testing

to evaluate the usability of the Android application has been positive and also led to discovery of a new use-case for long-term forecasting as well. We discuss next some reflections from our work.

First, we note that the timeline for conceptualization, development, field testing, and safe deployment of solutions can be quite long in the context of agriculture. We started the project during the soyabean season of 2021 and initially spent time understanding the problem statement through field visits. We also deployed early-on the ODK-based data collection application since we realized that data for the target market was not available publicly. We completed building the forecasting and recommendation methods, and the prototype Android application, by mid-2022 but could not launch a field test until later in the year during the soyabean season of 2022. This was because the field officers were occupied with other crops during this time, and the testing too would not have been realistic without being able to get live feedback from the cooperatives to be able to compare their actions with the recommendations made by the application. The field testing was only partially successful though. Since the prices during 2022–2023 remained flat, we did not get to see any interesting sales recommendations that would have led to a “wow” factor or, conversely, any starkly incorrect recommendations that would have helped us observe changes in the trust placed by various stakeholders in the technology or the risks of using the technology. However, this unique price pattern helped us recognize the importance of long-term price forecasting, a feature we are now incorporating in the Android application. To determine whether the application is safe to deploy for the cooperatives and the federation, it seems, however, that we still do not have a conclusive answer, and we may have to wait for another year to repeat the study or to incorporate forecasting and recommendations for other crops so that we can conduct year-round tests.

Second, we highlight that the long time-line of the project and an absence of clearly demonstrated tangible benefits affected the motivation of various partners to stay deeply engaged through the project. In fact, the IVR communication system played a crucial role in sustaining the relationships, because its benefit was clearly visible to all participants from the outset. We therefore believe that innovation projects done in close collaboration with field partners should not constrain themselves to only addressing one main problem that might be the core focus but also support field partners to address other problems that they might have. This helps build trust in the relationship and also contributes toward more immediate impact while longer-term research and development remains in progress.

Third, continuing from above, this project was funded under an *AI for Social Good* grant, and therefore it was imperative to identify a problem statement that would use AI in some manner. During our field visits, we identified many other problems, too, but did not work further on them to restrict the immediate scope of the project to the use of AI. For example, digitization of sales and procurement transactions of the cooperatives was an important need to move from a paper-based accounting system to a digital system. Similarly, the use of the IVR communication system could also be expanded significantly to providing advisory on cropping and other related aspects such as welfare schemes and nutritional security for agricultural households, which were also concerns that were brought up by the field partners. We did not proceed with these ideas, because they did not involve any specific use of AI, but arguably they could have been more impactful, at least so far in the project timeline. Such gaps that arise with having a technology going in search of problems has been discussed at length to emphasize that having flexibility in the choice of solutions to address felt needs of the community are a better means of producing meaningful innovations [37, 40].

Finally, we had hoped that this study will also help us investigate models of how users place trust in AI-based technologies and incorporate their outputs in decision making processes. Factors such as fairness, explainability, auditability, and safety have been identified as crucial to build a

trustworthy relationship and to reinforce these throughout the lifecycle of a project from its design to development and deployment [39]. Other factors such as reliability and immediacy have also been proposed, especially to handle uncertainty in AI-based outputs [16]. We have tried to follow these principles, but our study is yet to advance to a stage where we can directly observe how various participants react to the technology’s successes and mistakes and factor its recommendations in their decision making processes.

As our study progresses, we hope to uncover many such nuances and contribute toward impactful incorporation of technology in the operation of farmer cooperatives.

APPENDIX

Table A.1. Indexes Computed from Satellite Data

Data	Description	Dataset	Temporal Resolution	Spatial Resolution
NDVI	Normalized Difference Vegetation Index	MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m	16 Day	250 m
EVI	Enhanced Vegetation Index			
LST_Day_1km	Day land surface temperature	MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1 km SIN Grid	8 Day	1000 m
LST_Night_1km	Night land surface temperature			
AvgSurfT_inst	Average surface skin temperature			
SoilTMP0_10cm_inst	Soil Temperature at different depths			
Tair_f_inst	Air Temperature	GLDAS Noah Land Surface Model L4 3 hourly 0.25 × 0.25 degree V2.1 (GLDAS_NOAH025_3H)	3 hours	27830 m
Evap_tavg	Evapotranspiration			
Tveg_tavg	Transpiration			
Wind_f_inst	Wind Speed			
CanopInt_inst	Canopy water evaporation			
ESoil_tavg	Direct evaporation from bare soil			
PotEvap_tavg	Potential evaporation rate			
Psurf_f_inst	Pressure			
Qair_f_inst	Specific humidity			
Rainf_tavg	Rain precipitation rate			
SoilMoi0_10cm_inst	Soil Moisture			
RootMoist_inst	Root zone soil moisture			

Based on experiments, we found that the NDVI and average precipitation time series gave the best performance.

Table A.2. Short-term and Long-term Event Types Seen in the News

Type of event	Event Description	Occurrences
Short-term events	Low imports of soyabean in India	7
	High imports of soyabean in India	8
	Low exports of soyabean from India	12
	High exports of soyabean from India	13
	Hoarding of soyabean	2
Long-term events	Price fluctuations in soyabean last year	34
	Rainfall in Maharashtra	33
	Rainfall in Madhya Pradesh	21
	Rainfall in Rajasthan	11
	Poor quality of soyabean seeds	3

Based on experiments, we found that the time series for rainfall events and those related to poor seed quality gave the best performance.

Table A.3. Comparison of RMSE Score with Exogenous Time Series Drawn from Satellite and News Data (2014–2020)

Target Mandi	Surrogate Mandi Price Data		News Data		Satellite Data	
	Exogenous Variable(s)	RMSE	Exogenous Variable(s)	RMSE	Exogenous Variable(s)	RMSE
Kota	Mahidpur	201.73	Rainfall Quality + Seed Quality	253.40	NDVI	216.47
Bhawani	Ganjbasoda	196.66		231.87	+	214.13
Raltam	Sonkatch	205.97		259.86	Average Rainfall	218.00

Cells in bold indicates the models with the lowest Root Mean Squared Error (RMSE) scores among all of the multivariate models.

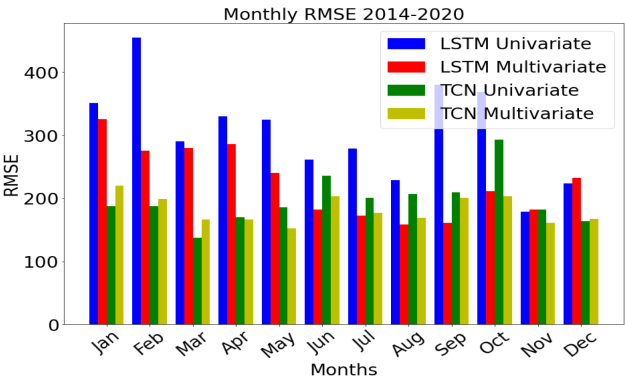


Fig. A.1. Monthwise RMSE scores (computed over 2014 to 2020).

Table A.4. Seasonal (September–January) Recommendation Metrics for Bhawani and Rajasthan 2014–2020

Years	Recommendation Metric	With Pertubation		Without Pertubation	
		Prospect Theory	Max Mean Forecast	Max Mean Forecast	Prospect Theory
2014	PAP	89.54	66.01	64.05	84.31
	NG	4.71	−5.05	−9.55	21.21
	NG Oracle	118.9			
2015	PAP	84.31	49.02	54.9	58.82
	NG	16.95	46	41.53	40.14
	NG Oracle	156.5			
2016	PAP	89.54	47.71	65.36	47.06
	NG	−10.22	−17.83	−14.29	−17.83
	NG Oracle	54.97			
2017	PAP	99.35	81.05	82.35	96.08
	NG	115.9	143.7	141.41	139.3
	NG Oracle	221.64			
2018	PAP	86.27	69.28	70.59	83.66
	NG	17.18	52.31	43.53	20.92
	NG Oracle	174.16			
2019	PAP	91.5	59.48	60.13	68.63
	NG	16.11	−35.29	−13.75	−15.92
	NG Oracle	253.29			
2020	PAP	100	81.7	79.08	93.46
	NG	15.33	84.57	97.62	87.33
	NG Oracle	342.64			

Cells in bold indicates the model with the top score for the specified metric within the particular year.

Table A.5. Seasonal (September–January) Recommendation Metrics for Ratlam and Madhya Pradesh 2014–2020

Years	Recommendation Metric	With Pertubation		Without Pertubation	
		Prospect Theory	Max Mean Forecast	Max Mean Forecast	Prospect Theory
2014	PAP	84.97	67.32	73.2	81.05
	NG	0.71	29.36	42.26	21.98
	NG Oracle	130.98			
2015	PAP	77.12	52.29	54.25	52.94
	NG	−12.06	−0.59	74.9	58.63
	NG Oracle	252.35			
2016	PAP	92.16	60.78	54.9	68.63
	NG	−5.27	−17.76	−23.59	−14.91
	NG Oracle	134.23			
2017	PAP	88.89	66.01	74.51	87.58
	NG	105.29	139.58	153.59	170.01
	NG Oracle	236.3			
2018	PAP	94.77	71.24	70.59	73.2
	NG	26.26	66.05	68.1	48.79
	NG Oracle	195.52			
2019	PAP	90.2	58.17	50.98	69.93
	NG	2.37	−44.17	−80.74	−73.56
	NG Oracle	374.88			
2020	PAP	100	75.82	71.24	84.97
	NG	3.2	133.2	107.95	107.16
	NG Oracle	287.16			

Cells in bold indicates the model with the top score for the specified metric within the particular year.

Table A.6. Evaluation for Long-term Forecasts Using LSTM and TCN for the Kota Mandi (Seasons 2014–2020)

Seasons	Exogenous Variables	LSTM Model			TCN Model			NG Oracle
		RMSE	PAP	NG	RMSE	PAP	NG	
2014	Target mandi	318.56	68.18	42.11	746.04	63.64	10.32	194.39
	Target mandi + Surrogate mandi	363.22	68.18	19.48	804.26	59.09	45.09	
	Target mandi + surrogate mandi + sine/cosine	395.00	72.73	52.99	624.01	72.73	24.48	
	Target mandi + surrogate mandi + satellite data	644.48	72.73	61.45	641.69	59.09	47.86	
	Target mandi + surrogate mandi + news data	446.61	77.27	57.34	491.17	77.27	42.35	
2015	Target mandi	302.72	54.55	77.63	191.50	40.91	19.74	179.27
	Target mandi + Surrogate mandi	235.50	50	17.04	201.39	40.91	−21.12	
	Target mandi + surrogate mandi + sine/cosine	309.01	45.45	44.08	643.22	36.36	−1.06	
	Target mandi + surrogate mandi + satellite data	456.99	54.55	37.28	593.28	27.27	−11.54	
	Target mandi + surrogate mandi + news data	236.94	68.18	65.05	574.71	40.91	−29.6	
2016	Target mandi	538.85	52.38	−44.5	701.85	33.33	−57.95	37.99
	Target mandi + Surrogate mandi	638.03	66.67	−51.47	738.12	14.29	−77.22	
	Target mandi + surrogate mandi + sine/cosine	531.15	38.1	−59.73	391.22	28.57	−73.32	
	Target mandi + surrogate mandi + satellite data	432.34	61.9	−47.86	420.75	14.29	−72.41	
	Target mandi + surrogate mandi + news data	552.89	47.62	−57.82	420.68	33.33	−38.76	
2017	Target mandi	286.02	95.45	269.32	311.13	95.45	270.82	457.16
	Target mandi + Surrogate mandi	305.52	90.91	241.81	285.66	90.91	267.01	
	Target mandi + surrogate mandi + sine/cosine	191.01	95.45	358.78	262.54	90.91	305.21	
	Target mandi + surrogate mandi + satellite data	464.95	95.45	213.47	264.36	86.36	356.68	
	Target mandi + surrogate mandi + news data	201.77	100	399.01	200.92	95.45	299.91	
2018	Target mandi	224.47	90.91	109.46	233.85	86.36	143.19	302.21
	Target mandi + Surrogate mandi	260.97	95.45	111.79	221.42	81.82	183.54	
	Target mandi + surrogate mandi + sine/cosine	198.19	86.36	208.17	231.44	81.82	179.8	
	Target mandi + surrogate mandi + satellite data	338.91	86.36	157.06	223.08	81.82	199.94	
	Target mandi + surrogate mandi + news data	209.58	90.91	171.78	219.44	90.91	239.69	
2019	Target mandi	655.04	72.73	17.48	348.18	50	−43.93	283.22
	Target mandi + Surrogate mandi	619.77	54.55	−110.19	338.03	45.45	−77.74	
	Target mandi + surrogate mandi + sine/cosine	745.50	68.18	29.37	568.42	54.55	9.51	
	Target mandi + surrogate mandi + satellite data	944.52	54.55	−53.06	629.20	59.09	−32.56	
	Target mandi + surrogate mandi + news data	780.40	45.45	−90.85	598.05	50	−71.59	
2020	Target mandi	664.01	100	313.39	472.38	100	399.81	710.36
	Target mandi + Surrogate mandi	672.35	95.45	439.21	474.07	100	549.34	
	Target mandi + surrogate mandi + sine/cosine	844.92	100	451.35	848.46	100	446.99	
	Target mandi + surrogate mandi + satellite data	1152.44	100	516.8	770.76	100	413.61	
	Target mandi + surrogate mandi + news data	787.12	100	503.4	777.92	100	375.73	

Cells in bold indicates the model with the top score for the specified metric within the particular year.

ACKNOWLEDGEMENTS

We also extend our thanks to the IIT Delhi High Performance Computing (HPC) Facility and Baadal team for their valuable infrastructure support. Additionally, we acknowledge the Center for Collective Development and the Gram Vaani field teams for their unwavering support and assistance in carrying out the work on the ground.

REFERENCES

- [1] Mohammed Abdellaoui, Han Bleichrodt, and Olivier l'Haridon. 2008. A tractable method to measure utility and loss aversion under prospect theory. *J. Risk Uncert.* 36, 3 (2008), 245–266.
- [2] Mohammed Abdellaoui, Frank Vossman, and Martin Weber. 2005. Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Manage. Sci.* 51, 9 (2005), 1384–1399.
- [3] T. S. Abdul-Jabbar, A. T. Ziboon, and M. M. Albayati. 2023. Crop yield estimation using different remote sensing data: Literature review. In *IOP Conference Series: Earth and Environmental Science*, Vol. 1129. IOP Publishing, 012004.
- [4] Trupti Agarwal. 2014. *Business Today, Year in Review: Why 2014 Was a Bad Year for Commodity Investors*. Retrieved June 23, 2023 from <https://www.businesstoday.in/in-depth/year-roundup-2014/story/2014-year-roundup-commodity-prices-fall-in-impact-on-investors-138467-2014-12-26#:~:text=%22Supply%20of%20South,analyst%2C%20Geofin%20Comtrade>
- [5] Nidhi Aggarwal, Sargam Jain, and Sudha Narayanan. 2017. The long road to transformation of agricultural markets in India: Lessons from Karnataka. *Econ. Politic. Weekly* 52, 41 (2017), 47–55.
- [6] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271. Retrieved from <https://arxiv.org/abs/1803.01271>
- [7] Sussy Bayona-Oré, Rino Cerna, and Eduardo Tirado Hinojoza. 2021. Machine learning for price prediction for agricultural products.
- [8] Parthasarathi Biswas. 2020. Reports of crop damage in MP, Maharashtra push up soyabean price. <https://indianexpress.com/article/india/reports-of-crop-damage-in-mp-maharashtra-push-up-soyabean-price-6903625/>
- [9] Harish Damodaran. 2016. The Indian Express, Agricultural marketing: For the Kisan, it's the Bania who still calls the shots. <https://goo.gl/JrFW3t>. Accessed June 23, 2023.
- [10] Nayan Dave. 2022. Worried Over Future Contracts, Farmers Hold on to Soyabean Stocks, Seek Better Prices. Retrieved Jun 23, 2023 from <https://www.financialexpress.com/market/commodities/worried-over-future-contracts-farmers-hold-on-to-soyabean-stocks-seek-better-prices/2895100/>
- [11] Shumin Deng, Ningyu Zhang, Wen Zhang, Jiaoyan Chen, Jeff Z. Pan, and Huajun Chen. 2019. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In *Proceedings of the World Wide Web Conference*. 678–685.
- [12] Cooperation & Farmers Welfare Ministry of Agriculture & Farmers Welfare Government of India Department of Agriculture. 2020–2021. Annual Report. Retrieved from <https://agricoop.nic.in/Documents/annual-report-2020-21.pdf>
- [13] Wenwen Dou, Xiaoyu Wang, Drew Skau, William Ribarsky, and Michelle X. Zhou. 2012. LeadLine: Interactive visual analysis of text data through event identification and exploration. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST'12)*. 93–102. <https://doi.org/10.1109/VAST.2012.6400485>
- [14] *The New Indian Express*. 2017. Vegetable Oil Imports Increase 34 Percent in July at 15.24 Lakh Tonnes. Retrieved Jun 23, 2023 from <https://www.newindianexpress.com/business/2017/aug/14/vegetable-oil-imports-increase-34-percent-in-july-at-1524-lakh-tonnes-1643153.html>
- [15] Sutanuku Ghosal. 2015. Soybean, Chana Prices Go Up; Maharashtra, MP, Rajasthan Farmers Relieved. Retrieved Jun 23, 2023 from <https://economictimes.indiatimes.com/news/economy/agriculture/soybean-chana-prices-go-up-maharashtra-mp-rajasthan-farmers-relieved/articleshow/47409641.cms>
- [16] Ella Glikson and Anita Williams Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. *Acad. Manage. Ann.* 14, 2 (2020), 627–660.
- [17] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* (2017). <https://doi.org/10.1016/j.rse.2017.06.031>
- [18] Nishant Gupta. 2021. *Price Forecasting and Anomaly Detection for Agricultural Commodities*. Master's Thesis. Amarnath and Shashi Khosla School of Information Technology.
- [19] Barbara Harriss-White et al. 1996. *A Political Economy of Agricultural Markets in South India: Masters of the Countryside*. Sage Publications India.
- [20] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.* 9, 8 (1997), 1735–1780.

- [21] Suresh P. Iyengar. 2022. The Hindu, SEBI Extends Trading Ban in 7 Agri Commodities. Retrieved June 23, 2023 from <https://www.thehindubusinessline.com/markets/commodities/sebi-extends-trading-ban-in-7-agri-commodities/article66288761.ece>
- [22] Radheshyam Jadhav. 2021. 14% of APMC Mandis, Farmers Have Joined e-NAM. Retrieved from <https://www.thehindubusinessline.com/data-stories/deep-dive/14-of-apmc-mandis-farmers-have-joined-e-nam/article35005589.ece>
- [23] Robert Jensen. 2007. The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *Quart. J. Econ.* 122, 3 (2007), 879–924.
- [24] Girish K. Jha and Kanchan Sinha. 2013. Agricultural price forecasting using neural network model: An innovative information delivery system. *Agric. Econ. Res. Rev.* 26, 347–2016-17087 (2013), 229–239.
- [25] Daniel Kahneman. 2003. A psychological perspective on economics. *Am. Econ. Rev.* 93, 2 (2003), 162–168.
- [26] David Kaleko. 2017. Feature Engineering Handling Cyclical Features. Retrieved from <https://blog.davidkaleko.com/feature-engineering-cyclical-features.html#The-Magic>
- [27] Virendra Kumar, K. G. Wankhede, and H. C. Gena. 2015. Role of cooperatives in improving livelihood of farmers on sustainable basis. *Am. J. Educ. Res.* 3, 10 (2015), 1258–1266.
- [28] Dan Li, Yijun Li, Chaoqun Wang, Min Chen, and Qi Wu. 2023. Forecasting carbon prices based on real-time decomposition and causal temporal convolutional networks. *Appl. Energy* 331 (2023), 120452. <https://doi.org/10.1016/j.apenergy.2022.120452>
- [29] Jianping Li, Guowen Li, Mingxi Liu, Xiaoqian Zhu, and Lu Wei. 2022. A novel text-based framework for forecasting agricultural futures using massive online news headlines. *Int. J. Forecast.* 38, 1 (2022), 35–50. <https://doi.org/10.1016/j.ijforecast.2020.02.002>
- [30] Lovish Madaan, Ankur Sharma, Praneet Khandelwal, Shivank Goel, Parag Singla, and Aaditeswar Seth. 2019. Price forecasting & anomaly detection for agricultural commodities in India. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*. 52–64.
- [31] Bharti Malukani. 2016. Export potential of soyabean from India: A trend analysis. Vol. 3, 43–53.
- [32] Ashima Mittal, Sahil Dahake, Tanmay Patel, Utsav Deep, Diwakar Prajapati, Akshay Gupta, Vivek Singh, Aadish Jain, and Aaditeswar Seth. 2022. Role of the mass media in monitoring and influencing the performance of social welfare schemes in India. In *Proceedings of the ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS'22)*. 78–102.
- [33] Kedar nath Mukherjee and others. 2011. *Impact of Futures Trading on Indian Agricultural Commodity Market*. University Library of Munich, Germany.
- [34] Yadong Pei, Chiou-Jye Huang, Yamin Shen, and Mingyue Wang. 2023. A novel model for spot price forecast of natural gas based on Temporal Convolutional Network. *Energies* 16, 5 (2023). <https://doi.org/10.3390/en16052321>
- [35] Tarun Satsangi. 2016. *The Economic Times, Soybean Prices to Remain Sluggish Despite Crop Loss Due to Unseasonal Rainfall*. Retrieved June 23, 2023 from <https://economictimes.indiatimes.com/markets/commodities/news/soybean-prices-to-remain-sluggish-despite-crop-loss-due-to-unseasonal-rainfall/articleshow/94894067.cms>
- [36] Anirban Sen, Debanjan Ghatak, Gurjeet Khanuja, Kumari Rekha, Mehak Gupta, Sanket Dhakate, Kartikeya Sharma, and Aaditeswar Seth. 2022. Analysis of media bias in policy discourse in India. In *Proceedings of the ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS'22)*. Association for Computing Machinery, New York, NY, 57–77. <https://doi.org/10.1145/3530190.3534798>
- [37] Aaditeswar Seth. 2022. *Technology and (Dis)Empowerment: A Call to Technologists*. Emerald Publishing, UK.
- [38] Aaditeswar Seth, Akshay Gupta, Aparna Moitra, Deepak Kumar, Dipanjan Chakraborty, Lamuel Enoch, Orlanda Ruthven, Paramita Panjal, Rafi Ahmad Siddiqi, Rohit Singh, et al. 2020. In *Proceedings of the International Conference on Information and Communication Technologies and Development*. 1–15.
- [39] Ehsan Toreini, Mhairi Aitken, Kovila Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya, and Aad Van Moorsel. 2020. The relationship between trust in AI and trustworthy machine learning technologies. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 272–283.
- [40] Kentaro Toyama. 2015. *Geek Heresy: Rescuing Social Change from the Cult of Technology*. Perseus Books Group.
- [41] Christopher Trepel, Craig R. Fox, and Russell A. Poldrack. 2005. Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Cogn. Brain Res.* 23, 1 (2005), 34–50.
- [42] T. V. Sivanandan. 2014. The Hindu, Moisture Stress in Soil Affects Red Gram. Retrieved June 23, 2023 from <https://www.thehindu.com/todays-paper/tp-national/tp-karnataka/moisture-stress-in-soil-affects-red-gram/article6333277.ece>
- [43] Ruijin Zhu, Wenlong Liao, and Yusen Wang. 2020. Short-term prediction for wind power based on temporal convolutional network. *Energy Rep.* 6 (2020), 424–429. <https://doi.org/10.1016/j.egy.2020.11.219>

Received 15 February 2023; accepted 24 February 2023