TerraByte

What does a farmer need to know to deal with confidence, and how should it reach them?

Project: Individual BDes pre-thesis project, Creative and Applied Computations, Srishti Manipal Institute of Art, Design and Technology
When: August–December 2024 (semester); revised September 2026
Author: Annie Vaithara
Guide: Vivek Garady, Dean of Creative and Applied Computations
Role: Fieldwork, data, modelling and interaction design
Methods: Informal fieldwork, time-series forecasting, rolling backtest, service and message design
Tools: Python (pandas, Prophet), Figma, InDesign
Status: The first stage of a framework. One module (market prices) is designed as a concept; none of the framework has been tested with farmers beyond small smoke tests in 2024.

Overview

TerraByte is the start of a framework that could grow into a knowledge hub for farmers in rural regions. It began as prediction. It became a project about confidence.

Smallholder farmers in Palghar district, Maharashtra, often have a single buyer for their harvest. They see only the first step of the chain that carries it to the city, and many have little hope of getting help from government schemes.

I began in 2024 with prediction: map the weather and forecast crop prices for farmers. Two things I heard in the field changed that. Farmers who spoke Hindi, whom I could talk to directly, told me that a daily weather feed only made them anxious. And as my father relayed his conversations with Marathi-speaking farmers, not knowing market prices turned out to be only part of the problem. The larger part was how little confidence farmers had, in selling to a buyer they did not know and in appealing for government schemes at all.

That gave the project its direction: a hub that gives farmers information they can trust and act on, starting with market prices. My hypothesis is that a price a farmer can check and point to gives them something to bargain with and more confidence in their own trade. It is a hypothesis, not a result. Later modules would cover government schemes and larger-scale environmental disturbances, and whatever else farmers say they need. Those are direction, not design.

In 2026 I rebuilt the original forecasting pipeline and tested it against the simplest alternative: repeating last month's price. On the only data I had, the forecast was no better. That points the same way as what farmers told me: current prices first, forecasts second. The revised design starts from the latest actual price, stated so that a farmer can check it and cite it. Smoke tests in 2024 with a small group of farmers had also suggested that an SMS felt calmer and more dependable than a pop-up notification. The next step is to test the price message with farmers, including whether it changes how confident they feel when they sell.

When	What
Aug–Dec 2024 (semester)	Fieldwork around Malwada; starting idea of predicting weather and prices; crop-recommendation experiments, dropped; monthly dataset and first forecasts; SMS and Panchayat concepts; smoke tests with a small group of farmers; documentation and exhibition

When	What
Sept 2026 (revision)	Data source identified; forecasting pipeline rebuilt and backtested; project reframed as the first stage of a knowledge hub; Panchayat display prototype; revised price-message concept; testing protocol; this case study

---

1. Context

Palghar district, a few hours north of Mumbai.

Palghar sits in the North Konkan region of Maharashtra, with coastal plains in the west, hilly forested talukas such as Jawhar and Mokhada in the east, and a large Adivasi population. Paddy is the main monsoon crop. Mango is grown in talukas such as Dahanu and Wada, and Dahanu's chikoo (sapota) is the district's One District One Product.

Most farmers here are smallholders. Their produce passes from village buyers and collectors to wholesale mandis, commission agents and, eventually, city shops, and each step takes a margin. A farmer selling at the farm gate sees only the first step.

The premise from the start was that farmers' knowledge of their land and crops is not the gap. What they cannot see from the village is information from beyond it, and market prices above all. The fieldwork added a second gap I had not planned for: confidence in acting on what they do know.

My family has farmed near Malwada for about forty years. That gave me a way in, and a position I had to account for.

---

2. Fieldwork

Language decided who I could hear, and how.

The fieldwork took place during the 2024 semester, in the villages around my family's farm in Malwada. It consisted of informal conversations with farmers and other residents, and observation, arranged through my parents. There was no fixed interview guide or sampling plan, but I asked my father to raise two questions: how confident farmers felt about selling to a buyer they had not sold to before, and how aware they were of government schemes for farmers.

Language shaped the method. Most of the conversations were in Marathi, which I am not fluent in. Farmers were not comfortable talking at length with someone who was not, and opened up once my father, a fluent Marathi speaker, led the conversations and I observed. Some farmers spoke Hindi, and with them I could communicate directly and clearly, so I spoke with them myself. What follows therefore reaches me by two routes: my father's interpretation of the Marathi conversations, and my own conversations in Hindi.

What that means for the findings. My family owns land in the area and my father manages the farm. The people we spoke to knew my family and were talking to a landowner, which may have shaped what they chose to say. These are observations from a small, connected circle, not a representative survey. I did not speak to traders, extension workers or Panchayat officials, and I cannot separate the views of women farmers, or of Adivasi farmers whose first language is neither Marathi nor Hindi.

Ethics. The conversations were informal, so there was no written consent process. No participant is named in the project materials.

3. What I Heard

Missing information was only part of the problem. The crux was confidence.

1. Prices over weather and soil. Farmers cared more about what their harvest would sell for than about weather or soil information. What they were mainly looking for was direct, current market prices.

2. Daily weather made farmers anxious. Weather was part of my starting idea: I was mapping it alongside prices. But mapping weather day by day, and passing it on, did nothing for the farmers except make them anxious. I heard this directly, in Hindi, from farmers I could talk to without an interpreter. My reading is that a farmer cannot change tomorrow's weather, so a daily feed gives them something to worry about and nothing to do. These conversations say nothing about how farmers would feel about rarer, larger-scale information, such as a major disturbance to the season.

3. One buyer. Buyers who come to the village can move on to the next seller if their terms are not met, but most farmers have only one buyer willing to take their harvest. Farmers expressed frustration and said their options were limited, because there are no cooperatives or other channels to sell through.

4. Little confidence, little hope. Apart from a few very popular government schemes, which farmers felt were being misused to some extent, they had little hope of actually getting the funding, support or help they needed. Across both questions my father asked, a lack of information was only part of the problem. The main one, as my father relayed these conversations to me, was a lack of confidence: in selling to someone new, and in getting help from government schemes. My reading is that without a current price to set against a buyer's offer, and without confidence that help was within reach, farmers had little to bargain with.

5. Data runs out. A common practice was to let a 30-day mobile data pack run its course even when the data ran out after 15 days, because renewing early was not affordable. Power cuts lasting several hours were frequent.

6. Outsiders are distrusted. Villagers were proud of their ways and sceptical of outside interventions, which they felt came from people who did not understand their lives. Needing my father to lead the Marathi conversations was a small example of this.

7. Land lies idle. Much agricultural land in the area is legally protected for Scheduled Caste and Scheduled Tribe communities. With large families and little land, ownership disputes are common, and fertile land can lie unused while families farm poorer plots. This is outside TerraByte's scope but is part of the picture.

> "Real time market price and price forecast, in that order of importance."
> — My conclusion, 2024 project notes

That note says what to send first. Taken together, the findings say more about how. Farmers wanted prices (1). More information was not simply better: a daily weather feed made them anxious (2). And what stood between them and a better deal was confidence as much as information (3, 4). My premise in Section 1 was that farmers lack information from beyond the village. The conversations suggested that was only half of it. Information matters to the extent that farmers feel able to act on it, and it has to reach them in a way that reduces worry instead of adding to it. A price they can trust and point to, in front of their buyer or a new one, might give them something to bargain with and more confidence in their own trade. Whether it does is the question the rest of this project has to answer.

4. 2024: What I Built

I started by predicting. Farmers pointed to prices, and I only partly followed.

The starting idea. I set out to predict things for farmers: the weather and crop prices, mapped and sent to them. I compiled monthly weather data alongside the prices, and a daily weather feed was part of the idea. The models never used the weather data (Appendix A.1), and after what farmers told me in Hindi (Section 3, finding 2), the revised concept leaves daily weather out.

Crop recommendation, dropped. My first models, in October 2024, were classifiers that recommended which crop to grow. Farmers cared more about what their crop would sell for, so I set them aside and concentrated on prices.

Price forecasts. In November 2024 I compiled a monthly dataset for rice, mango and onion, 2015–2023, and fitted a forecasting model (Prophet) with default settings.

Delivery. An SMS notification in Marathi, and a Panchayat interface where officials post notices that farmers can read.

The SMS concept read, in Marathi: "141 — mango price, four months from now. For more information, visit your nearest Gram Panchayat office." A number with no unit, four months out, came first. That was the opposite of what I had heard, and it gave a farmer nothing to check or point to in front of a buyer. I did not see either problem until the 2026 revision.

Smoke tests with farmers. I ran smoke tests with a small group of farmers: quick, informal checks that the idea made sense to the people it was for. I explained the basic idea of the end output aloud and showed a mid-fidelity prototype so that they could follow it. Shown as a pop-up notification, the message felt scary to them, like something they had to rush to act on. As an SMS it felt calmer, and they felt they could rely on it more. These were smoke tests, not alpha tests, so I treat the result as a pointer, not proof. The five-month semester left no time to go back and run proper alpha tests.

The pop-up and the daily weather feed are separate observations from different conversations, but they point the same way: information that arrives as an alert to react to produced worry. That is my reading, and I treat it as a pattern to test, not a finding.

---

5. 2026: The Data, Looked At Properly

The "prices" weren't prices.

Rebuilding the pipeline in 2026, I found every series dropping sharply in 2020: rice and mango to less than half their earlier level, onion by about a third. Shifts of that size in a staple crop are not a market crash.

My 2024 files named the source: the Labour Bureau of India, which "provided valuable monthly retail price indices", with weather from World Weather Online. They did not record the exact table.

In September 2020 the Labour Bureau moved its Consumer Price Index for Industrial Workers (CPI-IW) from base year 2001 = 100 to 2016 = 100. The break in these data falls in the same month. Rice, onion and mango are all items in the CPI-IW basket, and mango appearing only from February to August matches how a seasonal item is priced.

Rebasing predicts that the later series equals the earlier one divided by its 2016 average, times 100. Rescaling the rice index after the break by the earlier series' own 2016 average (252.2) puts January 2021 at 282.7 on the old scale, against 280.0 in January 2020 and 281.1 in August 2020, the last month before the gap. For onion and mango the same check gives year-on-year changes of −23% and −7%, which are plausible for such volatile prices but weaker evidence.

I have not located the exact table, so this is a strong inference rather than a confirmed fact, and whether the indices are all-India or for a single CPI-IW centre is unknown. If it holds, two things follow:

- **The numbers are index points, not rupees.** They show how an item's retail price compares with its average in the base year. In this case study, "price" means that published index value.
- **They describe what industrial workers pay in shops, not what a Palghar farmer is paid** at the farm gate or in a mandi. The two often move in the same direction, but the gap between them includes every margin along the chain in Section 1.

One more constraint came from the model rather than the data. The forecasting model expects daily data, so each monthly value was spread into synthetic days with random noise. That adds no information, and it makes the model's uncertainty ranges partly a product of the chosen noise level (Appendix A.2).

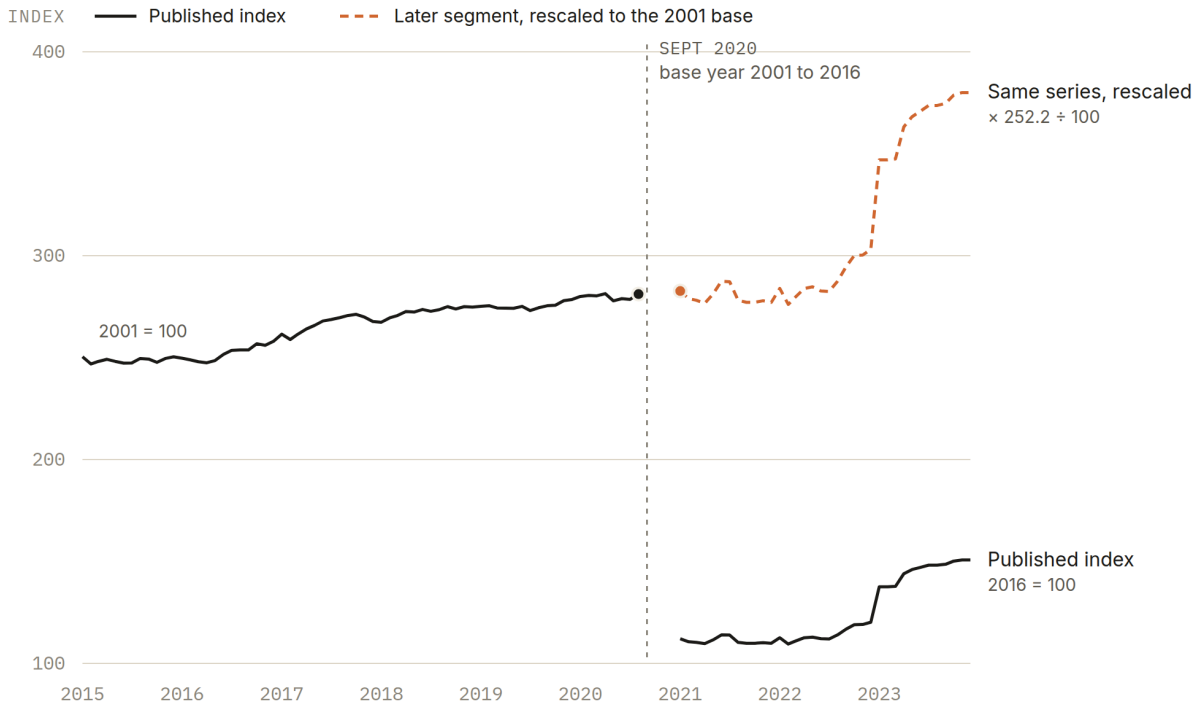

Figure 1: The rice index before and after September 2020. Multiplied by the earlier segment's 2016 average (252.2) and divided by 100, the later segment continues the earlier one.

---

6. The Test

On the only data I had, the forecast was no better than last month's price.

For every month from January 2018 with a published value, I retrained the model on everything before that month, forecast the month, and compared the result with the simplest alternative: repeating last month's published value.

One month ahead, 2018–2023	Rice	Onion	Mango
Months tested	67	54	27
Model: average error	2.0%	12.0%	32.5%
Last month's value: average error	0.9%	12.9%	21.9%
Model: median error	0.8%	7.1%	10.9%
Last month's value: median error	0.4%	9.5%	21.2%
Model's 80% range held the published value	85%	61%	26%

- **Rice:** last month's value had less than half the model's error, on average and in a typical month.
- **Onion:** roughly level on average; in a typical month the model did slightly better.
- **Mango:** better than last month's value in a typical month, but it missed badly at the start of some seasons, and its ranges held the published value only a quarter of the time.

What this does and does not show. It shows that on these data, a forecasting model added little over repeating the latest value, and that its uncertainty ranges were too narrow for onion and mango. It does not show that forecasting is useless for farmers. These are smooth monthly averages of shop prices, exactly the kind of series where "same as last month" does well (for rice, it was off by less than 1% on average). The prices farmers deal with are mandi prices, which move day to day. Whether a forecast beats "the latest price" for those is an open question (Section 9). The test is also not blind: I chose the model settings with all the data visible.

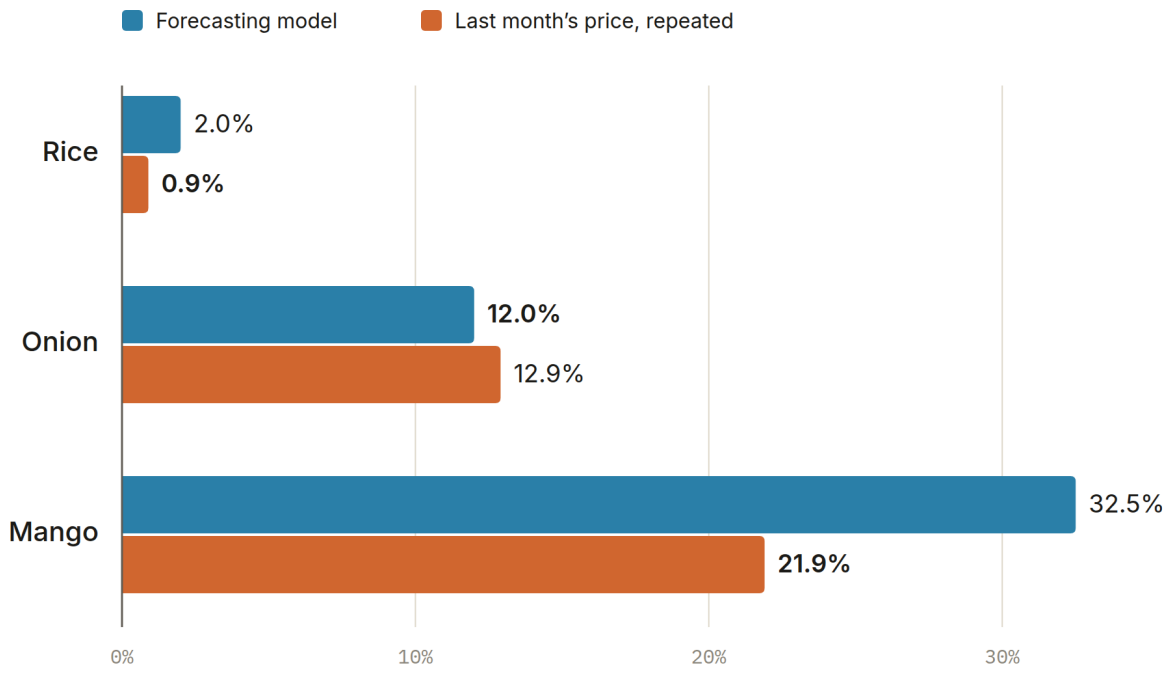

Figure 2: Average error one month ahead, 2018–2023: the forecasting model against last month's value, repeated.

---

7. What It Means for the Design

The evidence and the farmers pointed the same way.

Farmers told me what they wanted, not whether forecasts would work, and my backtest was on shop prices, not mandi prices. Neither proves that forecasts are useless. Together they are

enough to change where the design starts: with the latest actual price, not a forecast, and with calm delivery, not a feed.

What I heard	What the evidence says	What the design does now
Current prices first, forecasts second	On average, last month's value was at least as accurate as the model for rice and mango	Lead with the latest actual price; show an outlook only if testing shows farmers can use it
Daily weather made farmers anxious (conversations in Hindi)	Not tested. The 2024 pop-up, which felt rushed, hints at the same thing	No daily weather feed. If environmental information is added later, limit it to larger-scale disturbances, word it calmly and test it first
Most farmers have one buyer	An SMS price service in Maharashtra had no significant average effect on prices farmers received (Fafchamps and Minten, 2012)	Promise no better price; test whether knowing the going rate changes anything at all
Confidence, more than information, was the crux	Not yet tested	Make the price something a farmer can check and cite: the mandi and date on every message, and the same figure posted publicly at the Panchayat. Test whether it changes how confident farmers feel dealing with a buyer
Little hope of getting help from government schemes	Not tested	A schemes module is direction only. First find out what stops farmers applying: lack of information, paperwork, distrust or something else
Mobile data runs out mid-month	SMS and ordinary calls need no data pack	SMS and an automated voice call, plus a notice at the Panchayat office
A pop-up felt scary, like something to rush; an SMS felt calmer and more dependable (2024 smoke tests)	A small group and informal checks, with no numbers to report; not repeated since	A plain SMS worded without urgency, and no pop-ups or app notifications. Check that the voice call does not feel rushed either
Outsiders are distrusted	Not yet tested	Deliver through someone local, and test whether the Panchayat is trusted as a source of prices
—	The model's data are retail consumer indices	Use rupee prices from the mandis farmers sell to (for example Agmarknet's daily prices); compute any track record on those same prices

Related work. Giving farmers market information by phone has been tried and studied for two decades, with mixed results. Jensen (2007) found that when mobile phones reached fishers in Kerala, price differences between markets narrowed and waste fell; Srinivasan and Burrell (2013) later revisited the same fishers and found a more complicated picture. Closer to Palghar,

Fafchamps and Minten (2012) ran a randomised trial of Reuters Market Light, an SMS service that sent crop prices and weather to farmers in Maharashtra, and found no statistically significant average effect on the prices farmers received. On the interaction side, Avaaj Otalo (Patel et al., 2010) showed small farmers in Gujarat using a voice forum, reachable from any phone, to ask and answer agricultural questions, and Medhi, Sagar and Toyama (2006) showed that interfaces built from images and voice can serve people who cannot read text. Kay et al. (2016) found that people judged the chances of a bus arriving more accurately from a quantile dotplot, which shows a small number of equally likely outcomes, than from a continuous density plot.

The lesson I take from this work is that price information on its own rarely changes what farmers are paid. That fits what farmers told me: information was only part of the problem. It is also a warning for the hub as a whole. Information services have a mixed record, so each module has to earn its place by what farmers can do with it, and be judged on that, not on how much it shows.

---

8. Revised Concept: A Knowledge Hub, Starting with Price

One trusted place for what a farmer needs to know. The first thing it says is the price.

Concept. Only the first module is designed. The revised version has not been tested with farmers; the 2024 version was smoke-tested with a small group.

The idea. TerraByte is the start of a framework for a knowledge hub for farmers in rural regions: a calm, local, trusted source of the information they need to deal from a position of confidence. It is meant as public, non-commercial infrastructure delivered through local institutions and basic phones, not an app. It does not start as a forecaster. Forecasting was where I began, and Sections 3 and 6 are why it is not where the hub starts.

Why price comes first. Farmers asked for current prices. A price they can check gives them something to set against a buyer's offer, which my hypothesis says could give them bargaining power and more confidence in their own trade. And it is the information I have the most evidence about. The evidence is mixed: a randomised trial of an SMS service in Maharashtra found no significant average effect on prices received (Section 7). So the hub promises no better price, and the test in Section 9 is built to check the hypothesis instead of assuming it.

Module	What it would give a farmer	Why	Status
1. Market prices	The latest actual price from the mandi or procurement centre they sell to, named and dated	Farmers asked for prices; most have one buyer and nothing to set against the offer	Concept designed (below), untested. The pipeline so far runs on retail indices, not mandi prices
2. Government schemes	Which schemes exist, who qualifies and how to apply, in plain local language	Little hope of getting help, and low confidence to appeal for it	Direction only, not designed. First find out what stops farmers applying
3. Larger-scale environmental disturbances	Notice of major disturbances that affect a season or its prices, not daily weather	Daily weather made farmers anxious; big events may still matter for decisions	Direction only, not designed. What counts as worth telling, and how, is untested
4. Whatever farmers ask for next	Decided by what they say they need, not by what is easy to build	The hub should grow from their questions	Not defined

Principles across the hub, drawn from what I heard and from the 2024 smoke tests:

1. **Something to act on or point to, not raw data.** Each message should tell a farmer what it means for them.
2. **Calm.** No alerts, no pop-ups, no daily feeds. Worded without urgency.
3. **Checkable.** Every figure names its source and date.
4. **Reachable without a data pack or smartphone.** SMS, an automated voice call, and a notice at the Panchayat office.
5. **Public and local.** Delivered through local institutions, so a notice at the Panchayat also tells a buyer that farmers have seen the same figure.
6. **In the farmer's language, written and checked by native speakers.**
7. **Earns its place.** A module is added, and an outlook is shown, only if testing shows farmers can use it.

Module 1 in detail.

1. **The latest actual price comes first**, in rupees per quintal, from the mandi or procurement centre the farmer sells to, named and dated so that it can be checked and cited. Farmers asked for it, and repeating the latest value matched or beat the model.
2. **No outlook by default.** Whether a line about next month helps, and in what form (a word, a range or a track record), is what the test in Section 9 decides.
3. **If an outlook is shown, its track record is stated in plain words and defined**, for example "the actual price ended up inside this range about N times in 10", computed on the same mandi prices the message shows. The retail-index figures in Section 6 cannot be reused for this.
4. **Plain SMS, and the same message as an automated voice call.** No data pack or smartphone is needed, and the call needs no reading. The wording carries no urgency. In the 2024 smoke tests an SMS felt calmer and more dependable than a pop-up, which felt like something to rush. A call arrives on someone else's schedule, so it could feel rushed too; Section 9 checks that.
5. **A weekly notice at the Panchayat office**, printed or on a simple display, for people who do not receive messages. A public notice also means a buyer knows that farmers have seen the price.
6. **Marathi wording written and checked by native speakers, and Hindi where farmers speak it.** I can hold a conversation in Hindi but not in Marathi, so the Marathi cannot be left to me.

An open question about paddy. Palghar has government procurement centres that buy paddy at the Minimum Support Price (₹2,300 per quintal for common paddy in 2024–25). If most paddy in the area is sold there, a market-price message may matter little for paddy. The test asks each farmer where, and to whom, they sell.

How the design got here.

- **Starting idea (2024):** predict weather and prices for farmers.
- **2024 SMS:** a forecast four months out, with no unit, came first (Section 4).
- **2024 smoke tests:** shown as a pop-up, the message felt rushed; as an SMS, calmer and more dependable. They support keeping SMS as the default channel (Section 4).
- **Daily weather:** made farmers anxious, as I heard directly in Hindi (Section 3). The revised concept leaves it out.
- **2026 Panchayat display prototype** (`assets/mockups/lcd_dashboard_prototype.html`): a monochrome screen built on the pipeline's real outputs. On review it had four problems. "151 > 165" reads as "151 is greater than 165". The main screen is English capitals, with Marathi only on a test screen. "80% range. In past tests it held: 61%" assumes statistical literacy that nobody has tested. And it shows index points as if they were prices. It stays in the repository as a record of the iteration, not as the proposal.
- **Revised concept:** a knowledge hub whose first module is the latest actual price, with the principles above, still to be tested.

---

9. Next: Testing with Farmers

Four versions of one message, in front of farmers. Then two open questions about the rest of the hub.

Protocol written, not yet run (`research/user_testing_protocol.md`).

Card	Version
A	Latest price only
B	Latest price, plus an outlook in words (“next month: likely a little lower”)
C	Latest price, plus a range (“most likely between ₹a and ₹b”)
D	Latest price, plus a range and its track record (“the price has ended up inside ranges like this about 6 times in 10”; an example value for testing understanding only)

A voice version and a Panchayat notice are tested as well. This is a structured test: a written script, cards that are scored, and a record of who took part and who did not. It goes beyond the 2024 smoke tests, which explained the idea aloud and showed a mid-fidelity prototype. For the voice version, people are also asked whether it feels calm or rushed, because in 2024 a pop-up felt rushed and an SMS felt calm. The cards use real mandi prices from the same month a year earlier, so no one can act on them.

Beyond the price message. The same sessions end with two open conversations that are not tests of a design. One asks what has stopped or would stop farmers from applying for government schemes, and what they would want to know. The other asks how often, if at all, they want to hear about weather or larger events, and what makes that feel calm or worrying. Their answers decide whether modules 2 and 3 of the hub should exist and in what form.

Who. Six to eight farmers who sell their harvest, including women, people without smartphones and at least one person whose first language is not Marathi, and one or two Panchayat staff.

How. My father facilitates in Marathi and reads every card word for word; I observe and take notes. With Hindi-speaking participants I can facilitate directly, reading the same script. Each card is scored on whether the person can say correctly what it tells them. For each card he also asks whether it would make them more or less confident to ask their buyer for a better price, or to approach another buyer. What people say they would do in one session is weak evidence, so this is a first check, not proof.

What would change my mind. If people misread the range or the track record, the outlook stays off the message. If the latest price alone changes nothing about how they sell, the problem is the buyer, not the information. If knowing the price makes no difference to how confident they feel, a price message does not address the problem they described, and TerraByte has to look beyond information. If what stops farmers applying for schemes is paperwork or distrust rather than information, a module that only lists schemes will not help. If farmers say they want regular weather after all, the daily-weather finding needs a second look.

A second open question. Does a forecast beat “the latest price” for the mandi prices farmers actually face? Answering it means rebuilding the pipeline on Agmarknet’s daily mandi prices and rerunning the same backtest.

---

10. Reflection

What I’d do differently.

Start from what people need, not from what a model can predict. I began by asking what I could forecast. What farmers told me was about what they could do with information: the worry a daily weather feed caused, the wish for a plain price, the little hope of getting help. I only partly followed it in 2024, and the framework only took shape in 2026.

Put the rigour where the people are. I tested the model month by month, but the interface only with smoke tests in 2024; the five-month semester left no time for proper alpha tests. Next time, the message gets tested before the model gets tuned.

Keep track of where data comes from. The answer to my biggest data question had been in my own 2024 files all along.

Language decides who you can hear. The weather finding came from the conversations I could have directly in Hindi; much of the rest reached me through my father in Marathi. A study in Palghar needs Marathi from day one, and a plan for speakers of Adivasi languages. It also needs a plan for who you can hear without an interpreter.

Information isn't confidence, and more isn't better. Farmers said as much in 2024: not knowing was only part of the problem, and a daily weather feed added worry. Knowing the going rate doesn't create a second buyer, or the confidence to look for one. The 2024 design treated information as the answer; the revised one treats it as a means, and has to be judged on whether it changes confidence, not on how much it shows.

Marathi fieldwork conversations were led by my father; conversations in Hindi I held myself. Guide: Vivek Garady, Srishti Manipal. The code, data notes and testing protocol are in the project repository.

Appendix A: Technical Detail

A.1 Data

Component	Source	Resolution	Notes
Price indices (rice, mango, onion)	Labour Bureau of India, monthly retail price indices, compiled Nov 2024 (exact table not recorded)	Monthly	Index points, not rupees; base year changes in September 2020 (Section 5)
Weather (temperature, rain days, rainfall, sun hours)	World Weather Online, compiled with the price data	Monthly	Expanded to daily, but not used by the models

Rice covers January 2015 to December 2023; mango and onion run to March 2023. Onion has no data for January–August 2020 and rice none for September–December 2020. Mango has values only in its February–August season and none in 2020. Full details, including the linking check for all three items, are generated in `data/DATA_NOTES.md`.

A.2 Synthetic Daily Series

Prophet expects daily data, so each published monthly value P_m is expanded into daily values:

$$P_t = P_m + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \quad \text{clipped at zero}$$

There is no smoothing between months: the daily series is a step function with random noise. σ is 5 index points for rice and mango and 10 for onion, carried over from the 2024 script as illustrative choices. The average of each month's synthetic days stays within 1.8% (rice), 1.5% (mango) and 3.3% (onion) of the published value.

Two consequences follow. The daily series contains no information beyond the monthly values, so the model is in effect forecasting next month's average. And the width of the model's ranges depends on σ: refitting rice with σ from 2.5 to 20 (`check_noise_sensitivity.py`) moves its central January 2024 outlook only from 165.8 to 160.4, but widens its 80% range from ±4.1 to ±25.5. For mango, whose ranges are dominated by the model's own trend uncertainty, the range moves from ±24.8 to ±38.5.

A.3 Handling the September 2020 Break

- **No conversion factor is applied.** An earlier draft of the 2026 revision rescaled rice by factors of 6 and 15 so that it matched the paddy Minimum Support Price. That was a calibration to a government-set price for a different commodity, not a unit conversion, and it was removed (`reports/superseded_2026_draft/`).
- **The model estimates the shift itself.** Every data point carries a before/after indicator for September 2020, and the model learns the size of the shift from the data.
- **Baselines never compare across the break.** A month is only scored against a baseline whose reference value comes from the same side of the break; this removed one month per crop.
- **Linking is the standard alternative, and is not implemented.** Using each item's own 2016 average as a linking factor would be more principled than estimating the shift, and would change every number in Section 6.

A.4 Model

For each crop, Prophet is fitted to the logarithm of the index:

$$\log P(t) = g(t) + s(t) + \beta\, x(t) + \epsilon_t$$

where $g(t)$ is Prophet's piecewise-linear trend, $s(t)$ is yearly seasonality as a Fourier series, $x(t)$ is the September 2020 indicator and β the size of the shift. Working in logarithms makes seasonal swings and the shift act as percentages, and keeps back-transformed forecasts and ranges above zero. No holidays, weekly seasonality, weather or arrival volumes are used.

Setting	Rice	Mango	Onion
Training data	Jan 2015 – Dec 2023	Feb–Aug of each year, Mar 2015 – Mar 2023	Jan 2015 – Mar 2023
Fourier order (yearly)	5	8	10
Changepoint prior scale	0.05	0.15	0.20
Output	One-month outlook with 80% and 95% ranges	Same, within the Feb–Aug season	Same

The outlook covers the calendar month after the last published month (rice: January 2024; mango and onion: April 2023). The settings were chosen in 2026 with the full dataset visible.

A.5 Backtest Detail

The model was refitted on an expanding window for every month from January 2018, forecast every day of that month, and the average of those forecasts was compared with the published monthly value.

	Rice	Onion	Mango
Months tested	67	54	27
Months where the model was closer than last month's value	22	34	13
Model's 95% range held the published value	93%	72%	52%
Same month last year: average error*	6.3%	36.2%	24.1%

* On the months where both baselines are available (56 rice, 43 onion and 20 mango months). On that subset the model's average error is 2.1% (rice), 9.4% (onion) and 39.7% (mango). "Same month last year" is the weakest baseline for every crop; an earlier version of the validation compared the model only against it, which made the model look stronger than it is.

- **Rice.** The model's worst month was January 2023 (15% error), a sudden jump it could not anticipate. Its outlook of a 9% rise for January 2024 comes from having learned that single jump as a seasonal pattern.
- **Onion.** An earlier version of this backtest, without the break term, forecast values between 0 and 83 for September–December 2021 when the published values were 171–230 (preserved in `reports/superseded_2026_draft/`). The worst remaining month is December 2020 (75% error), just after the break.
- **Mango.** Most misses come at the start of a season, when the model must jump from the end of one season to the next; in March 2018 it forecast 1,269 against a published 289. These failures were left untuned, to avoid tuning against the test results.

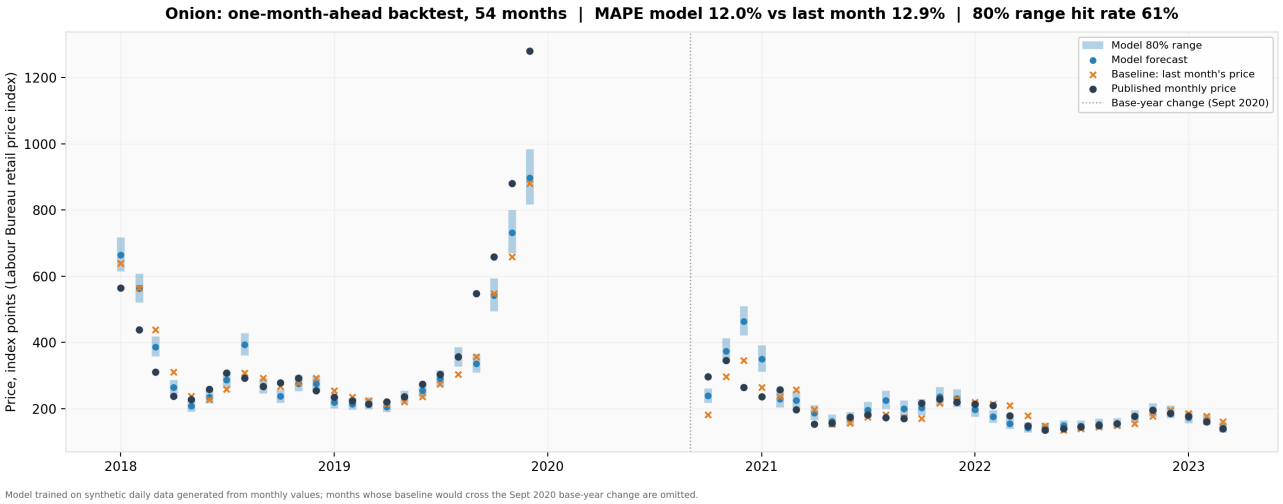

Figure 3: Onion backtest: published monthly values, model forecasts with 80% ranges, and the last-month baseline.

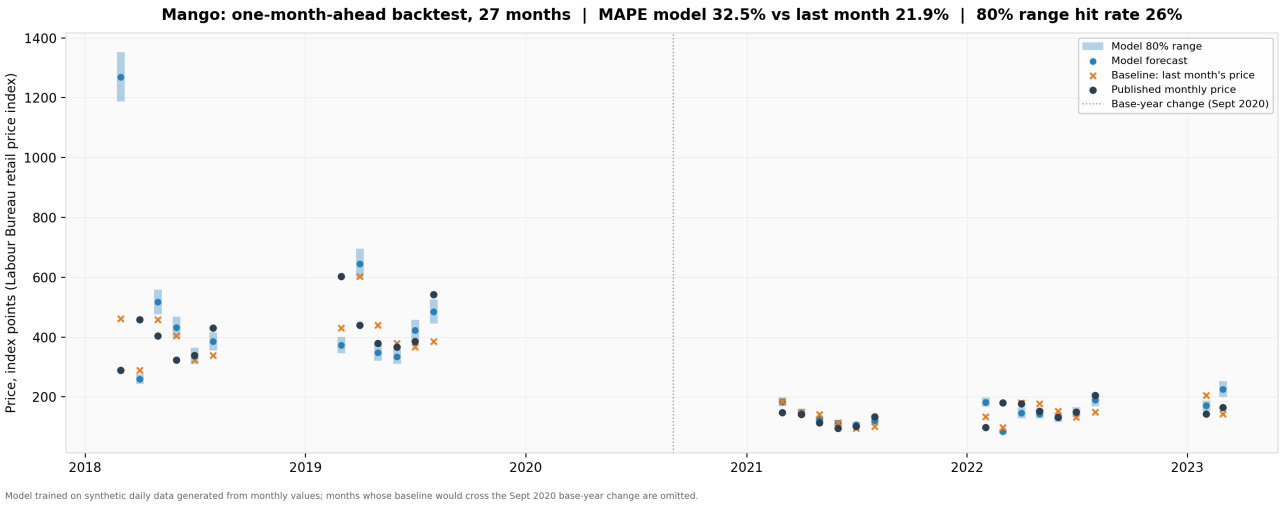

Figure 4: Mango backtest. Most misses occur at the start of a season.

A.6 Limitations

1. **Retail indices, not farm-gate prices.** The series are the Labour Bureau's retail price indices (very likely CPI-IW items); whether all-India or a single centre is not recorded, so nothing here is Palghar-specific.
2. **Synthetic daily data.** The daily series adds no information, and the model's ranges depend on the chosen noise level.
3. **Forecast accuracy.** The model does not beat repeating last month's value for rice or, on average, for mango, and is roughly level for onion; its ranges are too narrow for onion and mango.
4. **Test design.** Settings were chosen with all data visible; the mango test covers only 27 months.
5. **Shocks.** A model built only on past prices cannot anticipate export bans, tariff changes or fuel price spikes.
6. **Fieldwork.** Informal conversations within one family's network, mostly led in Marathi by my father rather than by me, plus my own conversations in Hindi. The confidence finding (Section 3) is my father's account of those conversations and my reading of it, not a measured result. The weather finding comes from my Hindi conversations and my reading of them; the number of farmers is not recorded here. Neither is a measured result.
7. **Smoke tests only.** In 2024 a small group of farmers were shown the idea and a mid-fidelity prototype (SMS against pop-up); there was no time for proper alpha tests. The revised message, the voice call and the 2026 Panchayat display have not been shown to farmers or Panchayat staff.
8. **Only the first module exists as a concept.** Schemes and environmental information are direction, not design, and nothing in them has been tested. The pipeline covers prices only.

A.7 Architecture and Reproducibility

The diagram shows the hub, not the code. The forecasting pipeline sits behind the diagram because it does not feed the default message; it runs in four steps: `normalize_data.py` (data preparation and the synthetic daily series), `validate_forecasts.py` (the rolling backtest), `generate_forecasts.py` (outlooks and plots, which also updates the display prototype) and, optionally, `check_noise_sensitivity.py`. All use fixed random seeds, and a clean re-run reproduces every data file and results table. Details are in `README.md`.

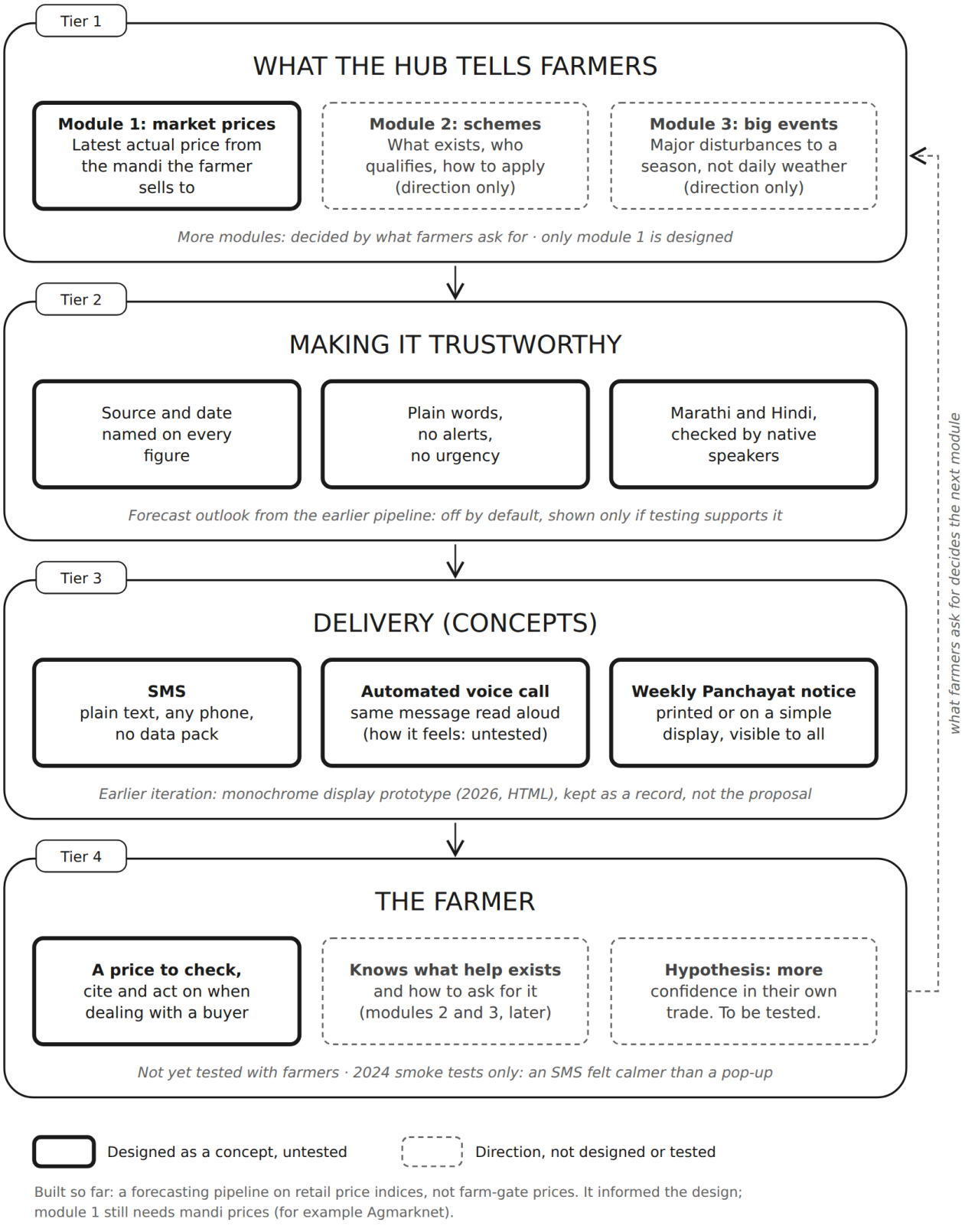

Figure 5: TerraByte system architecture. Tier 1: what the hub tells farmers (module 1, market prices, is a designed concept; modules 2 and 3 are direction only). Tier 2: making it trustworthy. Tier 3: delivery concepts. Tier 4: the farmer, with the confidence hypothesis still to be tested. A dashed line returns from the farmer to decide the next module. Nothing here has been tested with farmers beyond the 2024 smoke tests (Section 4).

References

- Fafchamps, M. and Minten, B. (2012). Impact of SMS-Based Agricultural Information on Indian Farmers. *The World Bank Economic Review*, 26(3), 383–414.
- Jensen, R. (2007). The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. *The Quarterly Journal of Economics*, 122(3), 879–924.
- Kay, M., Kola, T., Hullman, J. R. and Munson, S. A. (2016). When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. *Proceedings of CHI 2016*.
- Medhi, I., Sagar, A. and Toyama, K. (2006). Text-Free User Interfaces for Illiterate and Semi-Literate Users. *Proceedings of ICTD 2006*.
- Patel, N., Chittamuru, D., Jain, A., Dave, P. and Parikh, T. S. (2010). Avaaj Otalo: A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. *Proceedings of CHI 2010*.
- Press Information Bureau (22 October 2020). Revised Consumer Price Index for Industrial Workers (CPI-IW) based on new series 2016=100 from the existing 2001=100.
- Srinivasan, J. and Burrell, J. (2013). Revisiting the Fishers of Kerala, India. *Proceedings of ICTD 2013*.