AG TECH

Helping Agri business stake holders

www.agtech.com

Group 02

- Srihari Kodam - 12310037
- Mark Yeddu - 12410075
- Rahul Sharma - 12410057
- Tarun Singla - 12410096

Food Prices Forecasting [impacting healthy diet of Citizens]

Business Problem – This project aims to develop a forecasting model for retail prices of wheat and rice in India, using historical price data, agricultural outputs, weather trends, and market indicators. The forecasts will support early warning systems for inflation, guide buffer stock decisions, and inform MSP policy recommendations and regulate prices of healthy diet to make within the reach of citizen.

Business Objectives -

- CPI Inflation Target (India): 4% ± 2%Food inflation/prices must be controlled to keep overall inflation within this band.
- Ensure cost+50% profit to farmers *(as per the Swaminathan Commission)*
- Price rise to keep under 10%

Success Criterion		
Business	Policy Support	Forecasts directly inform timely decisions (e.g., buffer stock releases, export bans, MSP setting).
	Lead Time	Forecasts are made available 2–3 months in advance to allow for effective planning.
Economic	Inflation Control	Food inflation/prices remains within Ministry of Consumer Affair's (MCA) 4% ± 2% band.
	Production planning	Farmers get early signals on future prices, helping optimize crop choices.
ML	Low Forecast Error	Metrics like MAE, RMSE, or MAPE are within acceptable thresholds. For food prices:→ MAPE < 10% is often considered strong.
	Generalizability	Model performs well on unseen seasons, regions, or policy scenarios.

Feasibility

ML feasibility [High]

Aspects to decide:

1. Availability for consumer food prices, crop production, weather etc.
2. ML is well-suited for this task due to complex, non-linear relationships (e.g., between weather and price). Time series (ARIMA, LSTM), tree-based models (XG-Boost), and hybrid models all apply.

Legal Feasibility [Moderate to High]

1. Most data used (prices, production, weather) is non-personal/public, so privacy laws
2. Must ensure compliance with terms of use for public data (e.g., from IMD, ISRO etc.); not all may be freely shareable for commercial use.
3. Using ML to influence food policy (MSP, subsidies) requires institutional trust and transparency. Interpretability is crucial.

Requirements on the Application [Moderate to High]

1. Application must be accessible to non-technical users in ministries and planning bodies — requires dashboards, charts, alerts.
2. Real-time not needed; monthly or seasonal forecasts are sufficient. Makes the system easier to maintain.

Data Collection

Data Source	Data Schema	Data Source
Price Data	Date, State, Commodity, Retail Price	https://fpma.fao.org/giews/fpmat4/#/dashboard/tool/domestic https://fcainfoweb.nic.in/reports/report_menu_web.aspx
Production Data	Season, State, Area, Production (Thousand Tonnes)	https://desagri.gov.in/statistics-type/five-year-estimates/
Commodity market Stock Data	Month, State, Commodity, FCI Stock Value	https://fci.gov.in/statistical-data/stock
Inflation Data	Year/Month, Commodity, Inflation	https://cpi.mospi.gov.in/AllIndia_Item_CombinedInflation_2012.aspx
Weather Data	Day, State, Rain Condition, Humidity	http://www.imdagrimet.gov.in/

Data Collection

Data Selection – we will try to use past 12 month’s data to capture near future forecasting.

- Standardize values as features are in different scales (e.g., prices in ₹, rainfall in mm).
- Normalize time trends if needed.

Feature Selection – We may perform some of below methods for feature selections

- Observing data correlations
  - May drop high collinearity data
- Lagged Features for Time Series - These help capture **temporal dynamics**
- Model-Based Feature Importance
  - Tree-based models (e.g., XGBoost, Random Forest) - Shows how important each feature is in reducing error.

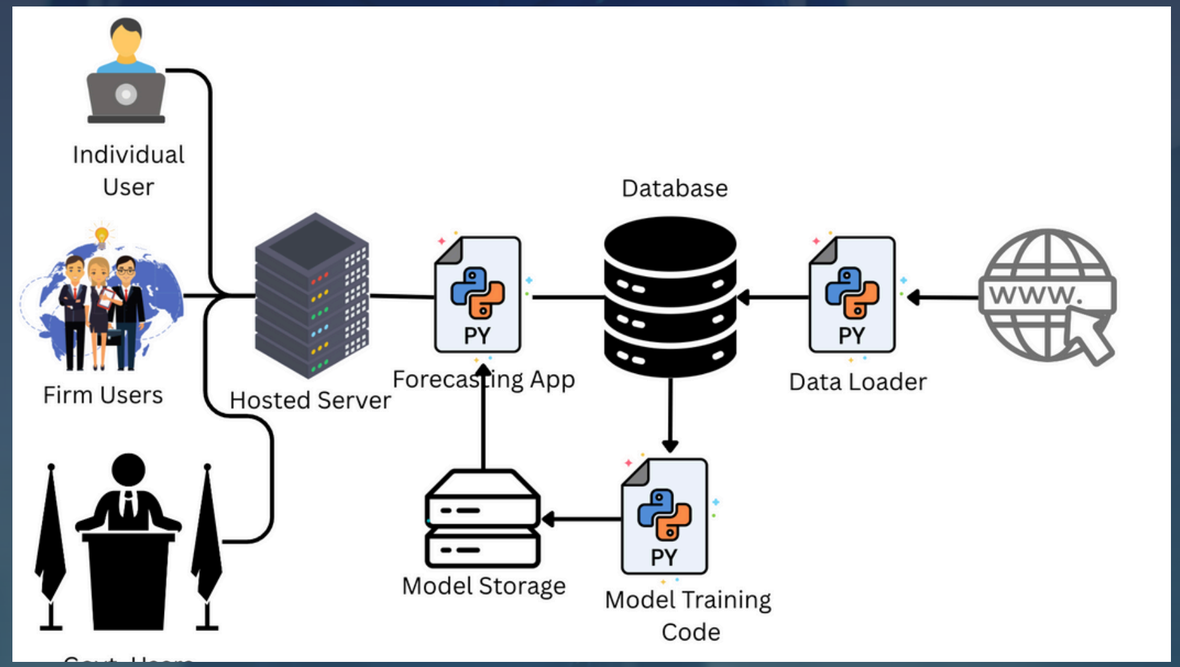
Unbalanced data:

- By assigning class weights
- Oversampling / Under-sampling

Feature Augmentation

- Price change%
- Rainfall deviation from normal
- Stock to consumption ratio
- Moving average of price or stock

Data Flow Diagram



Models and Expected evaluation

Model Type	Example Model	Performance	Robustness	Explainability	Resource demand
Baseline/Statistical	ARIMA, SARIMAX	Good	Sensitive to noise	Good	Light
Tree-based ML	Random Forest, XGBoost, LightGBM	Very Good	Handles missing data/irregularities	SHAP	Medium
Neural Networks	LSTM, GRU, Temporal CNN	Very Good	Needs clean, regular data	Hard	High
Hybrid/Advanced	Prophet, DeepAR, Transformer-based models	Good	Robust	Show seasonal trends	Medium

Model Evaluation

Metric	Description
RMSE	Penalize large errors more
R^2	Amount variation a model can cover of whole data
MAPE	% error – useful for comparing across scales

Models and Expected evaluation

These steps would include –

- Building features, handling missing values and standardize the data,
- Choosing models like XG Boost, ARIMA, SARIMAX, LSTM etc
- Training the model
- Evaluation (MAPE, RMSE, R^2 etc)

Pruning: Remove less important weights
Quantization: Convert weights to lower precision (e.g., INT8 for mobile/IoT deployment).
Distillation: Use a smaller model to learn from a complex one

We may have missing values for retail prices, may use a pre-trained ‘LSTM’ model as well to use self supervised learning

- Set Random Seeds for ML models
- Track Model Versions using Mlflow
- Log Experiments (datasets, hyperparameters)

Model Deployment

Deployment points	Details		
Define inference hardware	On premises - Government agency data centres	High-end CPU/GPU servers	
Model Evaluation under production condition	A/B Testing – Deploy in selected states (e.g., Punjab, West Bengal) and compare accuracy vs. baseline methods. Latency & Throughput – Measure how quickly predictions are generated in a real-time setup.		
Assure user acceptance and usability	Dashboards for Visualization Automated Reports for Decision Makers Use SHAP (SHapley Additive Explanations) to explain how inflation/weather impact price predictions.	Since the model serves government officials, policymakers, and farmers, it must be interpretable and accessible.	
Minimize risk of unforeseen errors	Detect anomalies in predictions vs. real prices. If price predictions are too extreme, flag for human review.	If the model predicts a 50%+ increase in wheat price in one month, an alert is triggered for manual review.	
Deployment strategy	Phased deployment – few states > Major states > Full scale > 3 month's hyper care Deploy API-based predictions that government portals & mobile apps can access.		

Model Monitoring

Since food prices depend on **inflation, weather, production, and government policies**, **data drift** is a major risk.

Key monitoring aspects for us:

- **Kolmogorov-Smirnov test / Jensen-Shannon divergence** to detect changes in distribution.
- Monitor Feature Importance (If inflation impact drops suddenly, the model may be outdated)
- Policy changes impact
- Global war/political changes impact

Detecting Hardware Degradation

- Track CPU, GPU, and RAM utilization

System Updates & Model Retraining

- Retrain the model every 3-6 months with fresh data
- Update MSP policies, FCI stock levels, and inflation rates in the feature set.
- software library updates

PEST Impact

Q&A TIME

Get the Answers You Seek