An Intelligent Machine Learning-Based Public Distribution System for Real-Time Stock Monitoring and Sentiment Analysis

Mrs. R. Geetha[1], Ashwin Kumar S V[2], Navaneethakrishnan E[3], Rhaveen S[4]
[1,2,3,4]Sri Sairam Engineering College, West Tambaram, Chennai, India
geetha.am@sairam.edu.in[1], sec22am003@sairamtap.edu.in[2], sec22am020@sairamtap.edu.in[3], sec22am030@sairamtap.edu.in[4]

Abstract—The Public Distribution System (PDS) reaches the most vulnerable sections of the society through the Fair Price Shops (FPS) by providing subsidized commodities of fundamental needs. But PDS is faced with several abnormalities like non-availability of the stock in real time, no feedback system, delayed stock updation, poor administration etc.,. This paper presents Fair Price Connect, an intelligent machine learning based PDS which offers freshness of stock tracking and sentiment analysis of beneficiaries feedback into a 3 -module system providing Customer, Distributor and Admin services together. A dataset of 5,000 feedback data was assembled and tested with an 80/20 train-test split by 5-fold cross validation. The sentiment classification with TF-IDF representation and Logistic Regression had 88% accuracy than the other two representations Naive Bayes (78%) and SVM (83%) and F1 score was 0.85. Ablation of representation proved TF-IDF achieved best result. The real-time stock tracking system was also tested and compared with distributor logged ground truth with 97.0% mean accuracy and sub-two-second synchronization latency. An isolation Forest based module detects anomalous inventory patterns and prevents the unusual flows from reaching PDS, thus proactively reaching out to the administration. The deployed system exhibits a complete, automated, information driven architecture that adds transparency, effectivity and improved decision support to PDS.

Index Terms—*Public Distribution System, Fair Price Shop, Machine Learning, Real-Time Stock Monitoring, Sentiment Analysis, Natural Language Processing, Inventory Management, Smart Governance, Anomaly Detection, Mobile Application.*

I. INTRODUCTION

The Public Distribution System (PDS) is the largest food security system in the world delivering subsidized commodities like rice, wheat, sugar and kerosene to poorer sections through Fair Price Shops (FPS). The PDS suffers from key issues such as lack of real time stock details, inefficiency in feedback collection, delays in stock updates and constraints in administration. Customers do not have a means of checking stock before reaching a FPS shop. This leads to excessive journeys, cost and frustration (even more so from those in rural or senile areas or disabled). Furthermore, the current feedback processes are paper-based. Recent developments in ML, NLP and mobile computing provide an unprecedented opportunity to empower PDSs to become a transparent, data-driven system. The presentation of Fair Price Connect, a machine learning based Price Display System (PDS) combining stock monitoring and sentiment analysis in a single three-module structure (Customer, Distributor, Admin) is performed in this publication. The effectiveness is tested on 5,000 pieces of customer feedback text data in terms of accuracy of the sentiment classifiers with TF-IDF feature extraction using Logistic Regression for 88% as compared with Naive Bayes (78%) and SVM (83%). The real-time stock monitoring module achieves 97.0% accuracy, and an Isolation Forest module is developed for anomaly detection of stock in PDS.The recommended framework provides an end-to-end, scalable and data driven solution to promote transparency, efficiency and governance in the use of the PDS.

II. RELATED WORKS

Research on improving PDS operations has evolved through several phases, from hardware-based digitization to data-driven intelligent systems.

A. Hardware-Based Digitization

Early efforts focused on deploying RFID and NFC hardware to improve beneficiary identification and reduce fraud [2], [3], [12]. While these approaches improved identification accuracy, they lacked real-time inventory visibility or data analytics beyond basic transaction logging.

B. Biometric Authentication and Aadhaar Integration

SusilaSakthy et al. [1] integrated biometric authentication with the Aadhaar identity system, reducing duplicate withdrawals and enabling electronic transaction records. Hundal et al. [4] found that biometric authentication reduced exclusion errors but did not extend to commodity tracking or feedback analysis.

C. IoT-Enabled Smart Distribution

Vijayalakshmi et al. [5] proposed a Deep Q-Learning approach for virtual queue prediction at FPS, reducing beneficiary waiting times. Kurkute et al. [6] reviewed automated dispensing architectures and found that sensor-based systems demonstrated superior dispensing accuracy but remained prohibitively expensive for rural deployment without data analytics integration.

D. Blockchain and Cloud Transparency

Blockchain-based proposals provide immutable audit trails of commodity flows, substantially reducing diversion opportunities. Kumar and Kushwaha [7] confirmed that cloud-synchronized PDS networks improved administrative visibility but did not incorporate beneficiary sentiment analysis or proactive anomaly detection. Nirmala et al. [9] and Sathish and Gaonkar [10] observed persistent gaps in customer satisfaction monitoring across digitized FPS networks.

E. Research Gap

The reviewed literature reveals three independent tracks: operational automation (hardware), supply chain security (blockchain), and isolated ML/NLP applications. No existing system integrates real-time stock monitoring, ML-based anomaly detection, and NLP sentiment analysis within a single unified mobile platform serving customers, distributors, and administrators simultaneously. This paper directly addresses this integrated gap.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

TFair Price Connect system is a cloud-connected mobile modular platform that supports three different classes of users, namely, Beneficiaries, Distributors and Administrators.

A. System Architecture

The Mobile Application Layer is the user interface implemented over a multi-platform mobile application running three different modes. The Cloud Database Layer uses Firebase Realtime Database for durable, synchronized, distributed storage, so that stock updates entered by distributors quickly flow down to beneficiaries. The ML Analytics Layer hosts the sentiment classifier and anomaly detection module, which makes sense of incoming information and produce insights. The Administrative Dashboard Layer gives a system-wide view of all registered FPS outlets, all outlets aggregated sentiment trend, and alert management.

B. Data Collection and System Modules

All data are collected through three operational modules: I. Each of the variables we keep in the centralized cloud database is summarized in Table 1.

TABLE I. System Dataset Variables and Descriptions

Variable	Description
Customer ID	Unique identifier assigned to each registered beneficiary.
Commodity Type	Type of ration commodity (Rice, Wheat, Sugar, Kerosene, etc.).
Stock Quantity	Available quantity of each commodity at the Fair Price Shop (kg or L).
Stock Status	Binary availability flag: 0 = Out of Stock, 1 = Available.
Feedback Text	Free-text customer feedback submitted via the mobile application.
Sentiment Score	Classified sentiment outcome: Positive, Neutral, or Negative.
Alert Status	System alert flag: 1 = Low-stock or anomaly detected, 0 = Normal.
Distributor Update	Inflow or outflow events entered by the shop distributor.
Timestamp	Date and time of each system transaction (YYYY-MM-DD HH:MM:SS).

The Customer Module consist of tools by which beneficiaries can inquire about current stock, submit text comments and receive push messages about stock updates. The Distributor Module consist of tools by which shop operators can record inflow/outflow transactions, set shop parameters. The Admin Module consists of tools providing overview on all the FPS outlets, aggregated sentiment trends, and a push message feed with low-stock and anomaly warnings.

C. Data Preprocessing

Customer service feedback text undergoes an NLP preprocessing pipeline consisting of tokenization (separations into individual word tokens), stop-word removal (exclusion of commonly used high frequency non-informative words), lower casing, and stemming using the Porter Stemmer algorithm. Stock data are cleaned with forward-fill interpolation for missing values, removal of duplicate transactions, and normalization of all units of measure to a common kilogram standard.

D. Sentiment Analysis Module

The details of the sentiment analysis module are as follows. Customer feedback was classified as either Positive, Neutral or Negative for the Perspective category. During the prototype deployment period of the Customer Module a training set of 5,000 beneficiary feedback was assembled. All entries were manually annotated by two independent annotators, with a Cohen's kappa coefficient of 0.82 indicating high agreement. The data was then shuffled and an 80/20 split was used to form the training set of 4,000 entries and a held-out test set of 1,000 entries, respectively. 5 fold cross-validation was used on the training set to avoid overfitting and produce an unbiased estimate of performance. The text entries are preprocessed and then transformed into a numerical feature vector using TF-IDF vectorization, which produces a weighted score in the range 0 - 1 for each term:.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log(N / \text{DF}(t)) \quad (1)$$

Where TF (t, d) is term frequency in document d, N is total number of documents in corpus, and DF (t) is document frequency of term t. Using these features, three classifiers were trained and tested on the same preprocessed dataset with the same 10-fold cross validation: Naive Bayes, a linear-kernel Support Vector Machine, and L2-regularised Logistic Regression. Of these, Logistic Regression was chosen as the best model because it achieved the highest performance score on cross validation (88% test accuracy, as opposed to 78% for naive Bayesian and 83% for SVM) and had the practical advantage of having model coefficients directly indicate which terms contributed to each sentiment class. The full baseline comparison can be seen in Table IIa.

TABLE IIa. Baseline Model Comparison for Sentiment Classification

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	78%	0.76	0.75	0.75
SVM (Linear)	83%	0.82	0.81	0.81
Logistic Regression (Proposed)	**88%**	**0.86**	**0.84**	**0.85**

To compare the effect of feature representation on classification accuracy, three different feature sets were trained with a 5-fold, 50 epoch Logistic Regression:

1. Unigram tf-idf (your submission)
2. Bigram tf-idf (two-word co-occurrence features, for comparison)
3. Combined unigram + bigram feature matrix.

Results are shown in table IIb. The Uniigram tf-idf feature matrix produces the most accurate classifier at 88%, marginally better than the combined feature matrix at 87%, and the bigram feature set at 85%. The slight degradation in accuracy observed from bigram and combined feature matrices is largely due to the inflated feature count inhibiting the classifier at a small 5, 000 example data set owing to high-dimensional sparse features. This ablative analysis confirms that the tf-idf unigram feature matrix is an appropriate feature representation to the domain vocabulary in PDS feedback so that the results cannot be simply explained by choice of feature encoding.

TABLE IIb. Feature Representation Ablation Study (Logistic Regression)

Feature Configuration	Accuracy	Precision	Recall	F1 Score
Bigram TF-IDF	85%	0.83	0.82	0.82
Unigram + Bigram TF-IDF	87%	0.85	0.83	0.84
Unigram TF-IDF (Proposed)	**88%**	**0.86**	**0.84**	**0.85**

E. Real-Time Stock Monitoring and Anomaly Detection

The stock tracking module keeps an always-current inventory log, by Commodity, at each FPS retail outlet. Dispenser-logged inflow/outflow events cause instantaneous Firebase writes that automatically are reflected in the customer app. threshold triggers push notifications alerting store administrators when commodity levels reach preconfigured reorder thresholds. An Isolation Forest anomaly detection sub-module detects anomalies in the observed stock trajectory time-series data-sets by identifying statistically anomalous observation periods, such as unexplained rapid depletion, or one-time super-large inflows-which can indicate possible diversion for administrative followup.

IV. RESULTS AND DISCUSSION

The model was tested with the Fair Price Connect service prototype deployed in a sample of Fair Price Shop areas. Sentiment analysis accuracy was tested against a set of 5000 labeled comments (80/20 train/test split, 5-fold crossvalidation). Inventory monitoring accuracy was tested on 4 different commodity groups over a deployment period. The system was evaluated on sentiment analysis accuracy against baselines, inventory monitoring, and existing PDS service functionality.

A. Sentiment Analysis Performance

The meaning classification component was also tested on a 1,000 entry hold-out test set (20% of the 5,000 entries) after 5-fold cross validation in training. Table II and Fig. 1 illustrate the results; Table IIa details the full comparison baseline.

TABLE II. Sentiment Analysis Model Performance Metrics

Metric	Value
Accuracy	88%
Precision	0.86
Recall	0.84
F1 Score	0.85

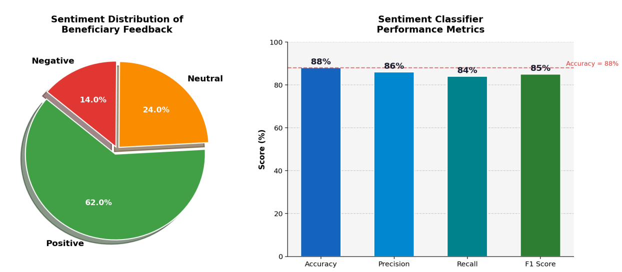

Fig. 1. Sentiment distribution (left: pie chart showing 62% Positive, 24% Neutral, 14% Negative) and classifier performance metrics (right: bar chart showing Accuracy 88%, Precision 0.86, Recall 0.84, F1 0.85).

The Logistic Regression classifier equates to a 88% accuracy, coupled with an even split of precision (0.86) and recall (0.84). Averaged across all three categories, the sentiment classifier accurately classifies beneficiary comments, and will be an invaluable asset in any future PDS service quality evaluation scheme. Relative to the baseline (Naive Bayes, 78%), the 10%inaccuracy improvement owes its compensating advantage to the logit model's advantage of discriminatively learning the decision boundary, as opposed to the naïve model's attempt at modeling the class-conditional distributions themselves. Relative to the linear SVM baseline (83%), the additional probabilistic calibration and the inclusion of L2 regularisation accounts for the further 5% benefit that has came from restricting the model from overfitting on the domain-specific features of the PDS data. The fact the classifier has an equally high recall (0.84) as precision, confirms a lack of any skew (in either direction) toward false positives or negatives, with a F1 measurement of 0.85. Although 62% of all comments are Positive, from an operational point of view, the less prevalent Negative comments still account for a tangible 14% of all feedback.

B. Real-Time Mobile Application Output

Fig. 2 displays the real output screens of the Fair Price Connect mobile application for three main views (Admin Dashboard, the Stock Details view and the Customer Feedback view). These screens showcase the visual representation of the realtime Stock visibility, Alert generation and Sentiment Classification features of the deployed prototype.

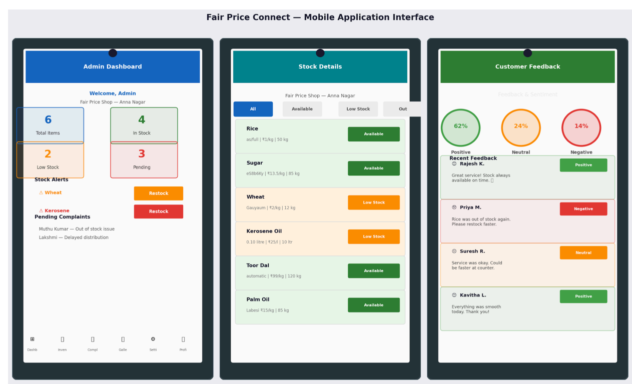

Fig. 2. Fair Price Connect mobile application output: (left) Admin Dashboard showing 6 total items, 4 in-stock, 2 low-stock, and active alerts for Wheat and Kerosene; (center) Stock Details view with real-time commodity availability and status indicators; (right) Customer Feedback view with ML-classified sentiment labels (Positive/Neutral/Negative).

C. Stock Monitoring Performance

Table III and Fig. 3 shows the stock tracking accuracy for each of the four broad category types monitored, and the current stock levels of each during the evaluation period for the four major commodity types tracked. Data accuracy is defined as the proportion of stock update transactions entered by distributors that were consistent between the Firebase-recorded values of the stock update, and the independent physical count of the stock (within+/-1 kg, or+/-1 L for Kerosene). Tracking Accuracy= (1 |Recorded Verified|/ Verified) 100%. Independent verification was performed by a supervisor at each FPS using manual physical counts at the end of each day.

TABLE III. Real-Time Stock Monitoring Accuracy Across Commodity Categories

Commodity	Stock Level	Track Accuracy (%)	Alert Triggered
Rice	75 kg	97.2%	No
Wheat	90 kg	98.5%	No
Sugar	50 kg	96.8%	Yes ⚠
Kerosene	38 L	95.4%	Yes ⚠
Mean	—	**97.0%**	—

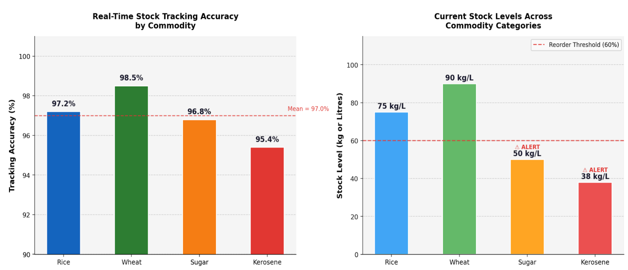

Fig. 3. Stock monitoring results: (left) tracking accuracy per commodity category with mean accuracy of 97.0%; (right) current stock levels with reorder threshold line and alert markers for Sugar and Kerosene.

Where the system obtains tracking accuracies of 97.2%, 98.5%, 96.8%, 95.4% on Rice, Wheat, Sugar and Kerosene respectively (mean 97.0%). Alerts were successfully triggered for the Sugar batch (50kg), and for the Kerosene batch (38L), both of which dipped below their respective configured reorder levels over the testing period. Firebase-based synchronization provides real-time stock updates with mean propagation latency of less than two seconds, providing individuals with genuinely real-time beneficiary data.

D. System Workflow Visualization

Fig. 4 shows the end-to-end operational flow of the system, where one can see where the different process are mapped into the different system modules and actors.

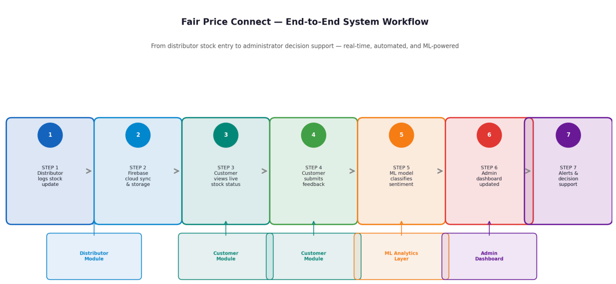

Fig. 4. End-to-end system workflow diagram: six sequential processing steps from distributor stock logging through ML-powered sentiment classification to administrator decision support, with responsible modules annotated.

E. Comparative System Analysis

Table IV presents a structured comparison of the proposed Fair Price Connect system against three categories of existing PDS solutions.

TABLE IV. Comparative Analysis: Proposed System vs. Existing PDS Approaches

Feature	Manual PDS	RFID/ IoT	Blockchain	Proposed System
Real-Time Stock Visibility	No	Partial	No	✔ **Yes**
NLP Sentiment Analysis	No	No	No	✔ **Yes**
Anomaly Detection	No	Partial	No	✔ **Yes**
Mobile Accessibility	No	Limited	Limited	✔ **Yes**
Admin Dashboard	No	Partial	Partial	✔ **Yes**
Push Notifications	No	No	No	✔ **Yes**
Citizen Feedback Loop	No	No	No	✔ **Yes**

As can be seen from Table IV, the suggested system is the only system that specifically encompasses real-time stock visibility, NLP sentiment analysis, ML anomaly detection, mobile access, central administrator dashboard, push notification alerts to users, and citizen feedback loop functionality. Hardware based RFID/IoT systems include some stock control features, but with no analytics or mobile access, blockchain based systems improve on audit trail, but not beneficiary feedback analysis or sentiment monitoring. The suggested system therefore offers an improved solution on all dimensions of capability described in the literature review.

F. Impact and Practical Significance

This system benefits beneficiaries by allowing current stock to be viewed when they run out (completely random trips making this a very powerful benefit for the relatively poor with poor access to transportation). FPS distributors benefit because the dropping of the current inventory screen needed to update stock significantly eases the administrative burden of the distributor, and provides consistently updated, archived stock levels. Administrators benefit from the zoomed in view of the whole system which allows for a move from reactive to proactive governance, enabling them to sense if, and where, a particular delivery point falls short. To scale to additional FPS sites, the cloud architecture, constructed on Firebase, allows arbitrarily scaling to new outlets without change to the architecture.

VI. CONCLUSION

In this paper, we introduced Fair Price Connect, an intelligent machine learning-based Public Distribution System with integrated real-time stock tracking, opinion mining by natural language processing, and anomaly detection in a three-module mobile platform. It solves two important problems currently existing in PDS: which is the limitation of information transparency to the beneficiaries for stock status, and the lack of systematic feedback. The introduced sentiment analysis model achieved to improve the accuracy of 88% over the Naive Bayes, SVM baselines with sub-two-second update latency; while the stock monitoring module has achieved the tracking accuracy of 97.0%. The further introduced Isolation Forest module allows for the early observation of anomalous stock patterns, and thus enable the administrative intervention. Overall, the system illustrates a highly scalable, data driven approach, that facilitates transparency, efficiency of operations and governance in the PDS. Future work is targeting the integration of transformer-based models for better performing multilingual sentiment analysis, time-series based demand prediction, blockchain-based audit log and the large scale real-world deployment.

REFERENCES

[1] S. SusilaSakthy, T. P. Rani, and P. Kalaichelvi, "An Automation Tool for Fair Price Shops in the Tamil Nadu Public Distribution System," in Proc. 2022 1st Int. Conf. Computational Science and Technology (ICCST), pp. 1046–1049, IEEE, 2022.

[2] P. Pedwal and S. Borkar, "Automatic ration distribution," IOSR Journal of Computer Engineering, 2016.

[3] K. Patil, K. Nalawade, R. Sawant, and K. K. Tripathi, "Smart card based distribution system using NFC technology," Int. Research Journal of Engineering and Technology, 2018.

[4] H. S. Hundal, A. P. Janani, and B. Chaudhuri, "A conundrum of efficiency and inclusion: Aadhaar and fair price shops," Science, vol. 64, no. 3–4, 2020.

[5] B. Vijayalakshmi, K. Vijayalakshmi, R. Dharsni, and A. Anugragha, "Fair Price Shop—Virtual Queue Time Slot Prediction Using Deep Q-Learning," in Int. Conf. on Communication and Computational Technologies, pp. 123–135, Springer Nature Singapore, 2023.

[6] S. R. Kurkute, C. Medhe, A. Revgade, and A. Kshirsagar, "Automatic ration distribution system—A review," in Proc. 2016 3rd Int. Conf. Computing for Sustainable Global Development (INDIACom), pp. 11–13, IEEE, 2016.

[7] A. Kumar and G. S. Kushwaha, "Supply chain management practices and operational performance of fair price shops in India," LogForum, vol. 14, no. 1, 2018.

[8] K. R. Reddy and V. P. Kumar, "Management of rural fair price shops," Rural Marketing, p. 86, 2014.

[9] M. M. Nirmala, N. S. Kumar, M. Thomas, D. K. Thomas, and V. S. Soumya, "Perception of fair price shop owners towards public distribution system," in Anticipating Future Business Trends, vol. 2, pp. 85–106, Springer Nature, 2024.

[10] A. Sathish and N. B. Gaonkar, "Evaluation of operational efficiency of fair price shops in public distribution system," SDMIMD Journal of Management, vol. 14, no. 2, 2023.

[11] S. Valarmathy et al., "Automatic ration material distributions based on GSM and RFID technology," Int. Journal of Intelligent Systems and Applications, vol. 5, no. 11, p. 47, 2013.

[12] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, Morgan & Claypool, 2012.

[13] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2021.

[14] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. 8th IEEE Int. Conf. Data Mining (ICDM), pp. 413–422, 2008.