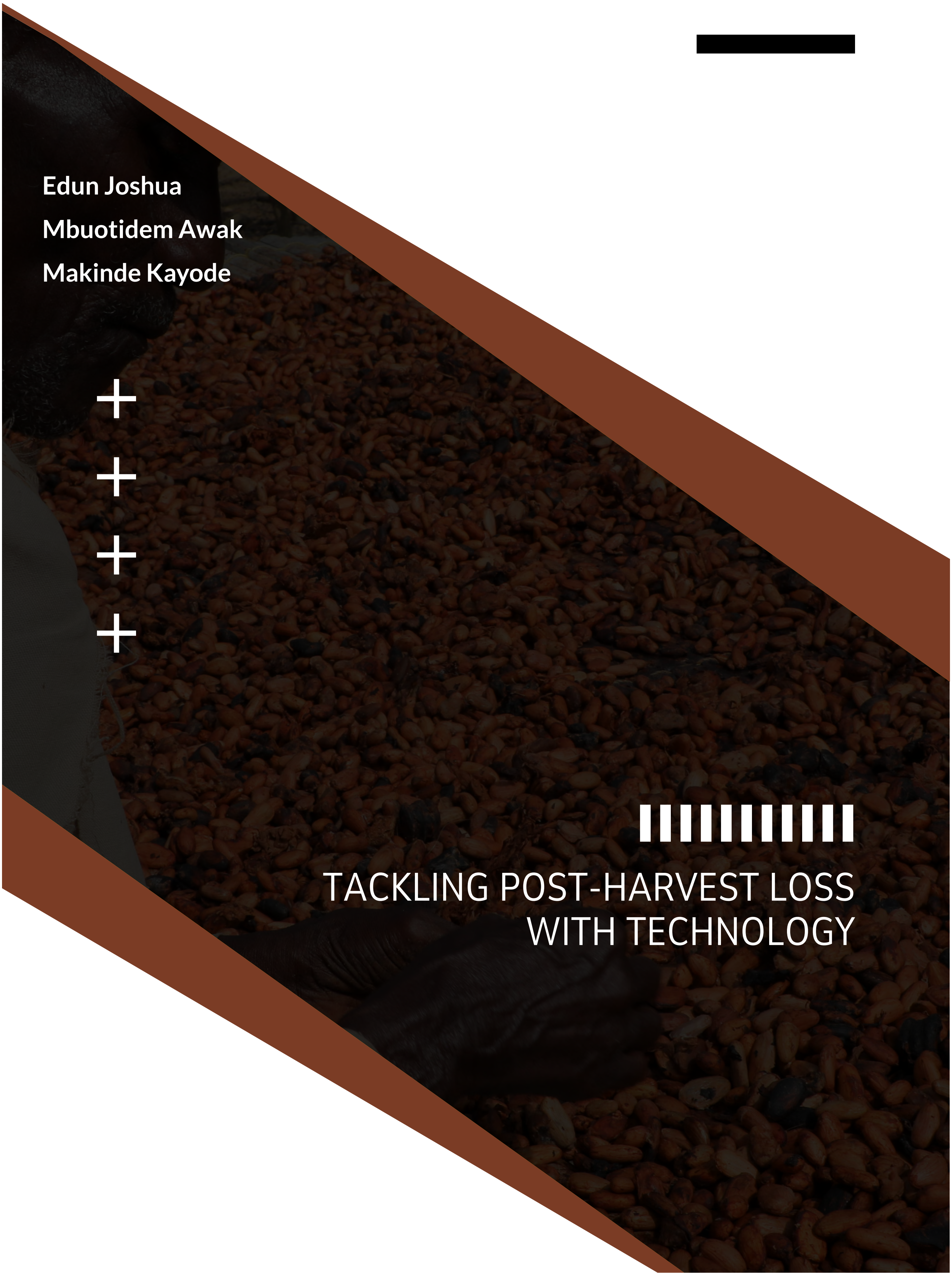


OKRA

Edun Joshua

Mbuotidem Awak

Makinde Kayode

TACKLING POST-HARVEST LOSS WITH TECHNOLOGY

Post-Harvest Losses Problem: Reducing Post-Harvest Losses to Build Youth-Led Agri-Businesses

Investigating Strategies to Mitigate PHL

Edun Joshua

Mbuotidem Awak

Makinde Kayode

Organizers: Data Community Africa, AgriConnect, Dicey Tech, REDtech Africa

Academic report

Lagos, November 2025

This page intentionally left blank.

Post-Harvest Losses Problem: Reducing Post-Harvest Losses to Build Youth-Led Agri-Businesses

This page intentionally left blank.

Acknowledgements

Our deepest thanks go to the Data Community Africa community. We attribute significant professional growth to the foundation and resources they consistently provide.

We are also immensely grateful to REDTech and DiceyTech DiceyTech for sponsoring valuable hackathons like this. They empower data enthusiasts by offering crucial opportunities to challenge themselves and showcase their abilities.

This page intentionally left blank.

Abstract

In Nigeria, millions of tons of food go to waste every year; not because of poor harvests, but because they never make it from farms to markets. To tackle these losses, we present Okra — a data-driven digital marketplace and logistics platform built to seamlessly connect farmers, buyers, and logistics providers in real time. Its stand-out feature is an AI-powered tool for predicting freshness and quantity of produce using image recognition, and helping buyers make faster, more informed decisions. Okra also uses data such as rainfall forecasts and harvest schedules to predict high-risk post-harvest loss periods, enabling early interventions like dispatching logistics or alerting farmers.

Keywords: Post Harvest Loss, Digital Marketplace, AI.

This page intentionally left blank.

Contents

This page intentionally left blank.

List of Figures

List of Tables

This page intentionally left blank.

1
Introduction

Authors: Edun Joshua, Mbuotidem Awak, Makinde Kayode

AdventureWorks operates as a global manufacturing entity, distributing products through a dual-channel strategy comprising an online direct-to-consumer interface and a robust reseller network. This report presents a comprehensive analysis of business performance, customer behavior, and product trends based on the provided dataset. The primary objective is to identify high-quality insights with actionable business relevance, supported by quantitative evidence.

> ***Tip***
>
> *AdventureWorks has generated a total revenue of $109.85M with a net profit of $9.37M. The business relies heavily on its Reseller channel, which accounts for approximately 73% of total revenue.*

1.1 Business Performance Overview

The analysis of the dataset reveals a business with a substantial market footprint. The aggregate financial and operational metrics indicate a mature operation with a significant customer base. As detailed in Table 1.1, the company has processed sales amounting to nearly $110 million, serving over 19,000 customers with a high Average Order Value (AOV).

Table 1.1: *Key Performance Indicators (KPIs)*

Metric	Value
Total Revenue	$109.85 M
Total Profit	$9.37 M
Total Customers	19.12 K
Average Order Value (AOV)	$3.49 K

1.2 Channel Distribution Analysis

A critical component of the AdventureWorks business model is its channel diversification. The revenue stream is bifurcated into Reseller and Online channels. The data demonstrates a clear dominance of the Reseller channel, which contributes the vast majority of gross revenue.

As presented in Table 1.2, the Reseller channel generated approximately $80.5 million, while the Online channel contributed $29.4 million. This distribution suggests a B2B-centric business model where bulk purchasing or wholesale agreements drive the core financial performance.

Table 1.2: *Total Revenue by Sales Channel*

Channel	**Total Revenue ($)**	**Share of Total (%)**
Reseller	80,487,704.18	73.27%
Online	29,358,677.22	26.73%
Total	**109,846,381.40**	**100.00%**

1.3 Temporal Revenue Trends

The temporal analysis of revenue from 2011 through 2014 highlights distinct growth trajectories for both channels.

1.3.1 Growth Phase (2011–2013)

The fiscal period beginning in Q2 2011 marked the initial intake of revenue, with the Reseller channel quickly outpacing Online sales.

- **2011:** The Reseller channel demonstrated aggressive growth, expanding from $489k in Q2 to $4.74M by Q4. Online sales showed steady but more modest growth, ending the year at $1.9M in Q4.
- **2012:** Both channels stabilized, with Reseller revenues consistently ranging between $5.8M and $7.7M per quarter.
- **2013:** This year represented the peak performance period. Reseller revenue surged, hitting a high of $9.82M in Q3 2013. Online revenue also saw its highest values, reaching $4.3M in Q4 2013.

1.3.2 Consolidation and Decline (2014)

The first half of 2014 shows a shift in momentum. While Q1 2014 maintained strong figures (Reseller: $8.27M; Online: $4.58M), Q2 2014 indicates a significant contraction, particularly in the Reseller market, which dropped to $3.42M. Data for Q3 and Q4 2014 is currently absent, suggesting either a reporting lag or a cessation of the analyzed period.

This trend analysis suggests that while the Reseller channel is the primary revenue engine, it is also subject to higher volatility compared to the relatively stable Online channel.

2

Channel Performance Analysis

2.1 Overview of Sales Channels

AdventureWorks distributes its products through two distinct avenues: the **Online** channel (Direct-to-Consumer) and the **Reseller** channel (Business-to-Business). A comparative analysis of these channels reveals a significant divergence between revenue generation and profitability. While the Reseller channel is the primary driver of volume, the Online channel serves as the engine for profit stability.

2.2 Financial Performance and The Profitability Paradox

The most critical insight derived from the longitudinal data (2011–2014) is the inverse relationship between revenue volume and profit margin across channels.
As illustrated in Table 2.1, the Reseller channel consistently generates the majority of revenue, peaking at $32.89 million in 2013. However, this volume appears to come at a cost. The Reseller channel has operated at a negative profit margin since 2012, recording a loss of -2.85% in 2013. Conversely, the Online channel, while smaller in scale, maintains a robust profit margin consistently hovering around 40%.

Table 2.1: *Yearly Revenue and Profit Margin Comparison*

Year	Channel	Total Revenue ($)	Total Profit ($)	Margin (%)
2011	Online	3.86 M	1.54 M	39.91%
	Reseller	8.78 M	0.08 M	0.97%
2012	Online	6.39 M	2.38 M	37.28%
	Reseller	27.13 M	(1.43 M)	-5.29%
2013	Online	10.73 M	4.29 M	40.00%
	Reseller	32.89 M	(0.94 M)	-2.85%
2014*	Online	8.37 M	3.47 M	41.45%
	Reseller	11.69 M	(0.03 M)	-0.24%

**2014 data represents a partial fiscal year.*

This discrepancy suggests that the pricing strategy for Resellers—likely involving deep volume discounts—may be eroding the bottom line, whereas the Online channel benefits from premium pricing power.

2.3 Operational Metrics: Volume vs. Value

The fundamental operational differences between the channels are highlighted by the Average Order Value (AOV) and items per order. The data confirms that the Reseller channel functions as a bulk-wholesale mechanism, while the Online channel serves individual consumer needs.

- **Reseller Channel:** Characterized by a high AOV of **$21,147.58** and an average of **16 items per order**. This indicates complex, large-scale logistical requirements.
- **Online Channel:** Characterized by a lower AOV of **$1,061.45** and an average of **2.18 items per order**, reflecting a transactional, high-frequency retail model.

2.4 Geographic & Personnel Performance

2.4.1 Territorial Growth

The North American market remains the stronghold for AdventureWorks. Specifically, the **Southwest** territory is a dominant performer, generating over $7.19 million in 2013 alone. However, international markets are showing distinct trends. The **United Kingdom** and **France** showed positive year-over-year growth in 2013 (1.23% and 1.61% respectively), identifying them as key expansion targets. Conversely, established markets like Central North America have shown signs of saturation or decline (-0.68% growth in 2014).

2.4.2 Salesforce Efficiency

An analysis of the top salespeople within the Reseller channel reinforces the concern regarding profitability. Table 2.2 lists the top five performers by revenue. Notably, **all top five salespeople generated negative profit margins**.

Table 2.2: *Top 5 Salespeople by Revenue (Reseller Channel)*

Salesperson	Total Revenue	Total Profit	Profit Margin
Michael Blythe	$9,293,903	($281,662)	-3.03%
Jae Pak	$8,503,338	($142,034)	-1.67%
Tsvi Reiter	$7,171,012	($147,095)	-2.05%
Shu Ito	$6,427,005	($396,323)	-6.17%
Amy Alberts	$732,759	($24,279)	-3.31%

For example, Michael Blythe generated the highest revenue ($9.29M) but incurred a loss of over $281k. This systemic issue suggests that sales commissions or targets may be aligned solely with gross revenue rather than net profitability, incentivizing aggressive discounting to close deals.

3

Okra: Our Solution Proposal

3.1 Core Concept and Functionality

Okra is a mobile-first, web-enabled platform designed as a comprehensive ecosystem to mitigate PHL. (View Okra here https://okra-ai.vercel.app/) The platform serves three primary user groups:

Farmers/Producers: Can create profiles, list their produce (crop type, quantity, harvest date, location), upload images for AI assessment, set indicative prices, and view market demand.

Buyers Can search for specific produce based on type, quantity, quality (informed by AI score), and location. They can connect with farmers, negotiate, and arrange purchases. Buyers include wholesalers, retailers, processors, and direct consumers.

Logistics Providers Can register their services (vehicle type – including refrigerated options, capacity, operational routes, pricing). The platform facilitates matching them with farmers/buyers needing transport.

3.1.1 Marketplace Networking

Farmers and producer groups register on the app to list available produce (crop type, quantity, harvest date). Buyers – including wholesalers, processors, and retailers – can search and place orders by region and crop. By centralizing listings, the app ensures farmers find customers quickly, and buyers can discover sources of fresh produce. This reduces unsold inventory and match supply to demand in real time.

3.1.2 Logistics Integration

The app includes a dashboard for logistics providers (truckers, couriers, cold-chain operators). When a sale is made, the system can automatically offer transport jobs to verified drivers. Users can see available vehicles, rates, and track shipments. For example, a farmer in Kano could schedule a refrigerated truck to deliver tomatoes to Lagos. This ensures reliable transport capacity and route planning, cutting delays that lead to spoilage.

3.1.3 AI-Powered Insights

A novel AI feature uses computer vision to analyze produce quality. Farmers or co-operatives upload smartphone images of their harvest batches. The app's AI model assesses freshness (e.g. spotting bruises, mold, color) and estimates volume or weight from the images. This accomplishes two goals: (1) It provides an objective quality grade for buyers to see, increasing trust. (2) It forecasts how long the produce will remain saleable. The model improves over time as more labeled images are fed back, refining its predictions.

3.1.4 Youth-led Agribusiness Empowerment

The platform encourages youth entrepreneurship. For example, tech-skilled youth can be trained to help digitize farms (taking images, using the app), to become last-mile delivery drivers, or to manage aggregation hubs. The platform could partner with agricultural colleges or startup incubators to recruit graduates. By framing agriculture as a tech-enabled business, the app draws young people into value chains. In effect, it transforms farming from subsistence into a connected, data-driven marketplace, which is more attractive to the next generation.

> ***Note***
>
> *Because of the time span of the hackathon, development of the Okra web app is not complete. Part of the frontend and most of the backend are still in development. The following sections show the current state of the project, including the mockups and the data strategy.*

3.2 Screens

Figure 3.1: *Screenshot of the landing page*

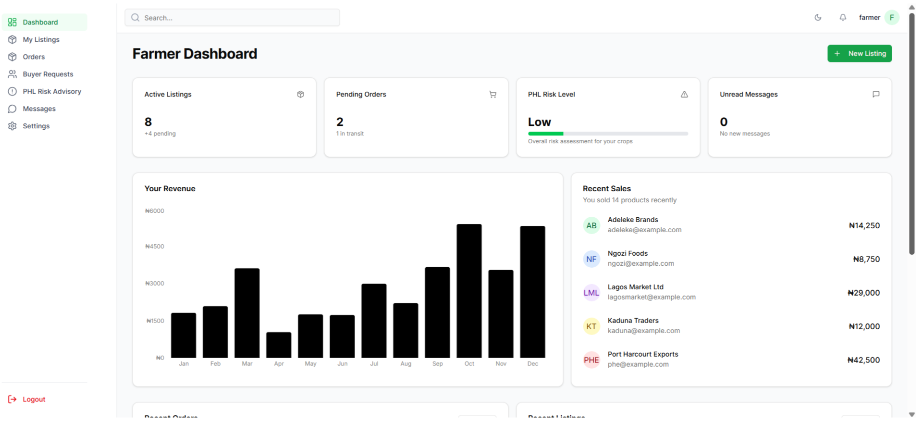


Figure 3.2: *Screenshot of the farmer's dashboard*

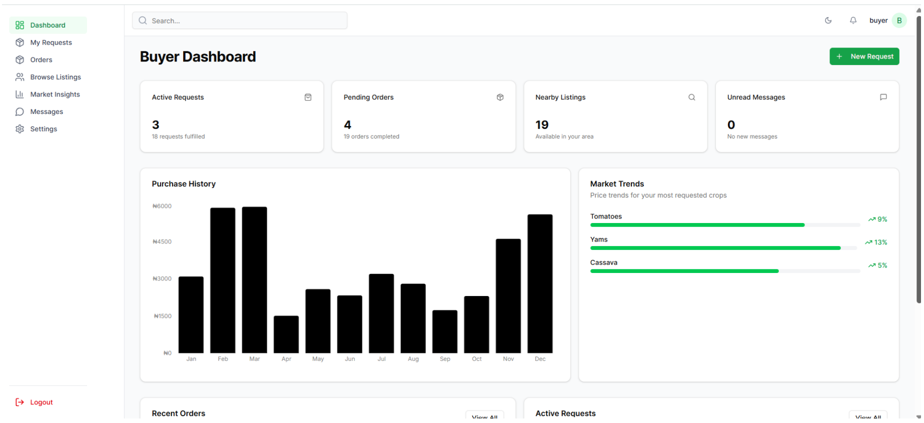


Figure 3.3: *Screenshot of the buyers's dashboard*

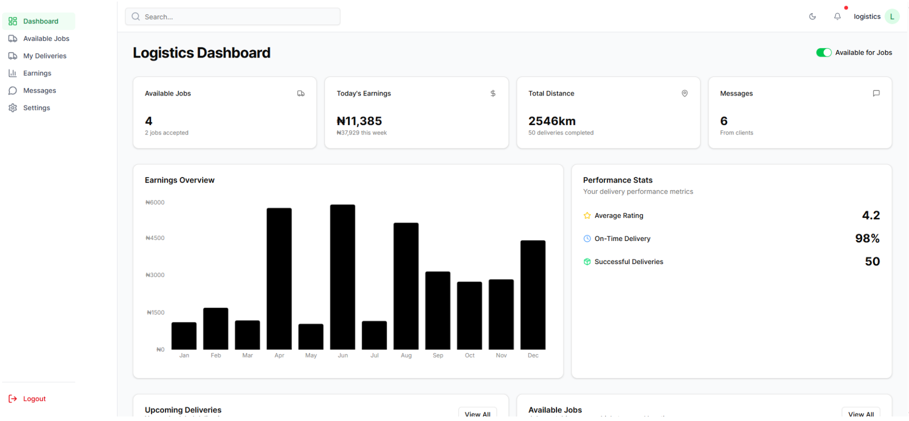


Figure 3.4: *Screenshot of the buyers's dashboard*

3.3 Data Strategy

3.3.1 Image Data Collection

We will collect large numbers of labeled images of harvested produce. This will be done in phases: initially, a pilot program in a few regions (e.g. Kano for tomatoes, Adamawa for maize) will train extension agents and youth volunteers to photograph produce at harvest and at sale. Each image will be tagged with metadata (crop type, variety, region, time, moisture, storage conditions, etc.). Over time, farmers using the app will also upload photos for every batch they list, creating a growing dataset.

3.3.2 Data Labeling and Training

Using expert agronomists and community training, images will be labeled for freshness (e.g. "Fresh", "Moderately aged", "Spoiled") and total count/weight. This labelled dataset feeds the computer vision model. We will employ transfer learning on convolutional neural networks so the AI can accurately assess new images it has not seen before. The initial model will be tested for accuracy and iteratively improved as more data arrives.

3.3.3 Feedback Loop (Active Learning)

Post-deployment, the app will track outcomes of predicted vs. actual. For example, if the AI predicts 10 days of shelf life but spoilage occurs in 7 days, this discrepancy is logged. Buyers and farmers can rate or report the accuracy of freshness ratings. This feedback is used to retrain the model periodically, increasing its accuracy over time. In effect, the app learns from real-world results.

3.4 Auxiliary data sources

In addition to user-generated data, the system will integrate public and commercial datasets for context. Possible sources include:

Crop statistics: Government or FAO data on regional production volumes and seasonality for major crops.

Geolocation Maps of farm locations, major roads, market centers.

Weather/Climate Historical and current weather data per region, since humidity and temperature affect spoilage.

Market Prices Data from local commodity exchanges or market surveys, to show price trends.

Demographic Data Farm population and sizes, to profile areas served.This external data enriches the dashboard: for instance, linking rainfall patterns to losses in maize, or highlighting regions that produce a given crop. All data is tied to time and location, enabling spatio-temporal analysis.

3.4.1 Privacy and Governance

Farmers' personal data (names, exact addresses) will be protected. Only aggregate or anonymized data used on dashboards. Data use agreements will ensure that insights (e.g. "X tonnes of tomatoes sold from Kano") respect user privacy while guiding decisions.

3.5 OkraAI

Okra features a Fruit Ripeness Prediction tool, using computer vision (CV), a specialized branch of machine learning that enables computers to analyze and interpret visual data from images. This technology facilitates the detection and prediction of fruit and vegetable freshness.
For this object detection task, the YOLOv11 model was specifically chosen due to its:

Speed: YOLOv11 is known for its fast processing speed, making it suitable for real-time applications.

Accuracy: It achieves high accuracy in detecting and classifying objects within images.

Flexibility: The model can be trained on various datasets, allowing it to adapt to different types of produce.

3.5.1 Dataset

A hybrid dataset was constructed by combining multiple primary and secondary datasets. This approach aimed to enhance the model's generalization capabilities and ensure a broad representation of fruit ripeness stages. The dataset includes the following classes:

- Ripe
- Unripe
- Rotten

Data sources include:

1. **Ripe Orange Dataset**: Primary data collected and annotated by Team Okra.
2. **Tomato Checker Dataset**: Source: Roboflow (https://universe.roboflow.com/money-detection-xez0r/tomato-checker/dataset/).
3. **Banana Ripeness Dataset**: Source: Roboflow (https://universe.roboflow.com/arm-oeppz/banana-8qkur/dataset/2).
4. **Rot Detection**: Source: Roboflow (https://universe.roboflow.com/srmist-doq3j/rot_detection/dataset/13).
5. **Orange Dataset 2**: Source: Roboflow (https://universe.roboflow.com/mert6107/orange_detection-5f84p/dataset/).

This dataset includes prevalent perishable fruits and vegetables including bananas, oranges, eggplants, garden egg, spinach, grapes, tomatoes, and cucumbers.

3.6 Data Split

The combined dataset was divided into training and validation sets using a 70:30 split.

Table 3.1: *Class distribution before re-splitting*

No. of Images	Ripe	Rotten	Unripe
Train	3000	2879	2105
Train Split	2100	2015	1473
Validation Split	900	864	632

3.7 Model Training

The pretrained YOLOv11s model was fine-tuned using the custom dataset, running with Pytorch and Ultralytics within the free tier Google Colaboratory environment. The compute environment specifications were: 112GB ROM, 12.7GB RAM, and a T4 GPU with 15GB RAM. The hyperparameters were set as follows:

Model: `yolov11s` (selected for a balance between speed and accuracy)
Epochs: 20
Image Size: 640
Optimizer: AdamW (learning rate = 0.001429, momentum = 0.9)

The training process was completed in 0.797 hours. The model's performance on the validation set is summarized in the table below:

3.8 Deployment

The YOLOv11s model achieves a mean Average Precision (mAP50) of 0.847 and an mAP50-95 of 0.65, exceling at fruit discrimination with high precision (Box(P)) and

Table 3.2: *Model performance on the validation set.*

Class	Images	Instances	Box(P)	Box(R)	mAP50-95
all	2396	3938	0.822	0.802	0.65
ripe	900	1077	0.896	0.949	0.895
rotten	864	1810	0.703	0.530	0.332
unripe	632	1051	0.868	0.928	0.721

Speed: *0.2ms preprocess, 5.1ms inference, 0.0ms loss, 2.0ms postprocess per image.*

recall (Box(R)) scores, and consequently, high mAP50 and mAP50-95 values. This indicates that the model is both accurate in its detections and effectively identifies most instances of these classes.

However, the model shows a comparatively lower performance in detecting rotten fruits. The precision (0.703) and recall (0.53) are notably lower than those for ripe and unripe categories, resulting in a lower mAP50 (0.62) and a significantly lower mAP50-95 (0.332). This suggests that the model may be less accurate in identifying rotten fruits and might miss a significant portion of them.

The inference speed of 5.1ms per image, after preprocessing and postprocessing, indicates that the model is efficient for real-time applications.

3.9 Confusion Matrix

The confusion matrix shows that the YOLOv11s model effectively classifies ripe (1015), rotten (1034), and unripe (984) fruits, reasonably well. However, there were significant misclassifications of background as rotten (766), and some ripe (119) and unripe (147) fruits were confused with background. Despite these errors, misclassifications between ripeness categories were low, showing reliable differentiation between maturity stages. While performing well, the model's robustness could be improved by more background data and stronger regularization to reduce false positives.

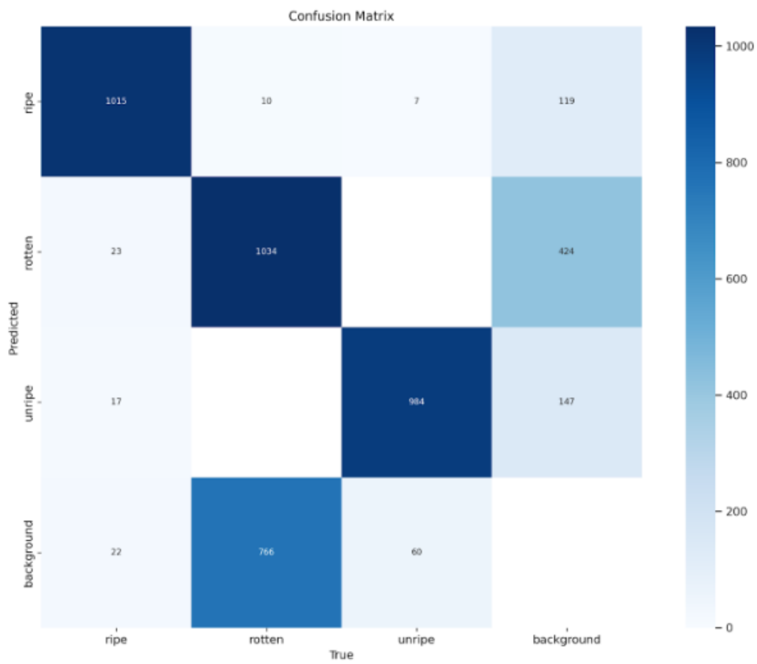


Figure 3.5: *Confusion matrix of the model*

3.10 Training Progression

The metric progression shows a consistent trend of improvement with increasing epochs, indicating effective learning and minimal overfitting. The proximate training and validation scores further underscore the model's generalizability to unseen images. Visual analysis reveals that at epoch 20, the terminal iteration, the model attains optimal performance, displaying the nadir of loss values and the apex of mean Average Precision (mAP), Precision, and Recall scores.

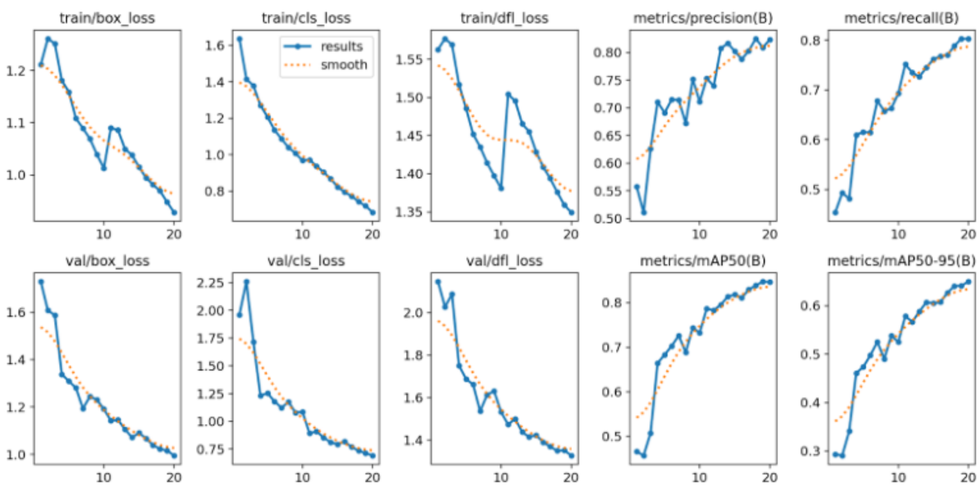


Figure 3.6: *Training progression of the model*

3.11 F1-Confidence Curve

The F1-confidence curve in Figure 3.7 illustrates the F1-score across various confidence thresholds. A higher F1-score indicates a more effective model in differentiating between the dataset's classes. Our model achieves a peak F1-score of 0.81 (81%) at a confidence threshold of 0.278. This suggests that for optimal deployment, the recommended confidence threshold should be approximately 0.3.

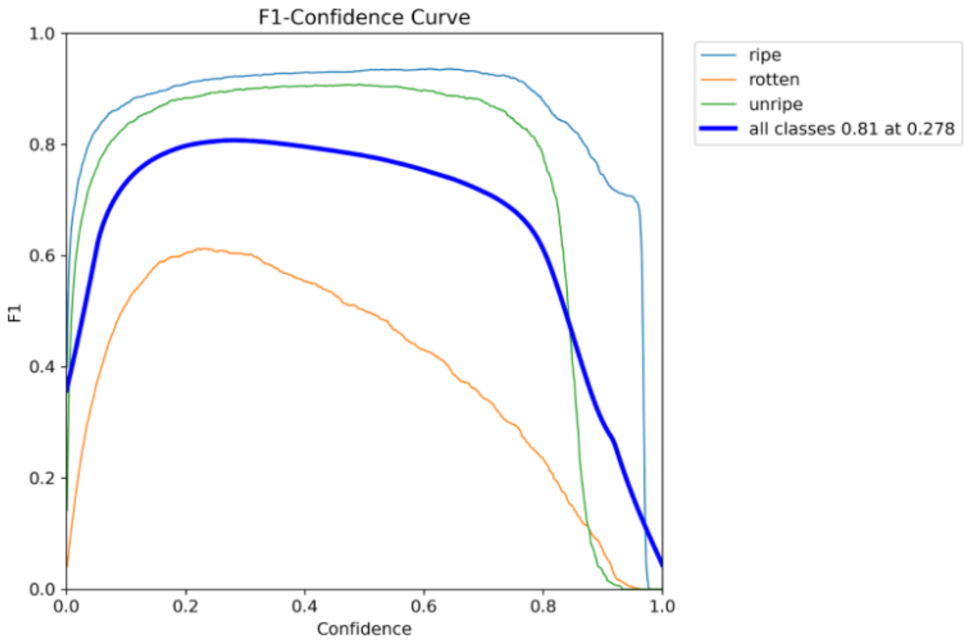
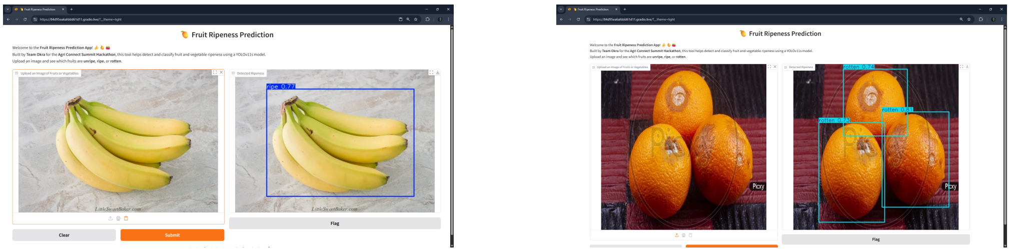


Figure 3.7: *F1-confidence curve of the model*

3.12 Deployment and Demo

For integration into the Okra app, the trained YOLOv11s model was exported to the ONNX (Open Neural Network Exchange) format (ONNX Community, 2019). ONNX is a widely supported format for representing neural network architectures, facilitating deployment across diverse platforms such as mobile devices, web applications, microcontrollers, and servers.

To showcase the model's capabilities, a demonstration web application was developed using the Gradio tool. Images illustrating the model in operation are provided below.



(**a**) *Gradio Demo 1* (**b**) *Gradio Demo 2*

Figure 3.8: *Gradio Demo of the model*

This page intentionally left blank.

Bibliography

ONNX Community (2019). *Open Neural Network Exchange (ONNX)*. https://onnx.ai/. Accessed: 2025-05-15.

This page intentionally left blank.