Sauda Haywood

DSC680: Milestone 2

Week 1-4: December 15, 2024

Project: Global Food Waste Analysis/Reducing Food Loss in Storage

Business Problem
Food loss during storage is a critical issue that affects food security, economic stability, and environmental sustainability. Policy makers and government agencies need data-driven strategies to reduce food losses and ensure efficient food supply chains.

Background/History
Globally, food loss occurs at various stages of the supply chain, with storage being prominent stage where significant loss happens due to poor infrastructure, inadequate storage methods, and pest infestations. Historical data indicates that many countries, especially in developing regions, face storage losses ranging from 10% to 20% annually, contributing to food scarcity and economic losses for farmers.

Data Explanation
Source: FAO's Food Loss and Waste Platform (https://www.fao.org/platform-food-loss-waste/flw-data/en/).

Data Explanation

Source: FAO's Food Loss and Waste Platform

Key Variables:

- m49_code: A numerical code representing a country or region.
- country: The name of the country where the data was collected.
- region: The geographical region of the country.
- cpc_code: A code representing the commodity classification.
- commodity: The type of food commodity (e.g., milled rice).
- year: The year when the data was collected.
- loss_percentage: The percentage of food lost during the supply stage.
- loss_percentage_original: The original loss percentage before any transformations.
- loss_quantity: The quantity of food lost (e.g., in kilograms).
- activity: Specific activities related to the data collection or food supply stage.
- food_supply_stage: The stage in the food supply chain being analyzed (e.g., storage).
- treatment: The storage intervention applied (e.g., trapping, duration).

- cause_of_loss: Factors contributing to the food loss (e.g., rodents, poor storage methods).
- sample_size: The size of the sample used for data collection.
- method_data_collection: The methodology employed to collect the data.
- reference: References for the data source or study.
- url: A URL linking to additional resources or data.
- notes: Additional notes or comments about the data.

Data Preparation:

To ensure reliable analysis, I cleaned the dataset by dropping columns with excessive missing values, such as region and treatment. Missing values in critical columns were handled using appropriate techniques, such as imputation or categorization. I also performed one-hot encoding to convert categorical variables into numerical format, making them suitable for machine learning models.

Findings and Recommendations

Commodity Findings:

- Commodities such as snails, fresh and chilled, exhibited the highest loss percentage (50%), followed by grapefruit juice (41.89%) and meat of pig with the bone (40.91%).
- Perishable goods were at greater risk of spoilage due to inadequate storage conditions.

Recommendations:

- Implement advanced refrigeration and packaging solutions for high-risk commodities.
- Train stakeholders on best practices for handling and preserving perishable items.

Activity Findings:

- Activities like Harvesting, Packaging, Sorting, Storage had the highest loss percentage (39%), followed by Farm, Marketing, Storage, Transportation (38%).
- Sorting and washing activities also contributed to significant losses (33.5%).

Recommendations:

- Optimize storage and transportation systems to reduce handling damage.
- Introduce efficient sorting and washing technologies to minimize losses.

Cause of Loss Findings:

- Rodent infestation was a leading cause, accounting for over 50% of losses in certain cases.
- Poor marketability due to over-ripeness, bruising, or rotting was another major factor.

Recommendations:

- Introduce pest-resistant storage facilities and cost-effective trapping solutions.
- Educate farmers on proper harvesting and handling practices to improve marketability.

Methods

1. Descriptive Analysis
2. Visualization
3. Random Forest Model and Linear Regression

Analysis

Key Findings:

- High Loss Variability: Loss percentages varied widely depending on treatment and cause of loss.
- Rodents as a Major Cause: Rodent infestations accounted for over 50% of losses in certain cases.
- Effective Treatments: Interventions such as trapping significantly reduced losses compared to untreated conditions.
- Duration Impact: Longer storage durations correlated with higher loss percentages, emphasizing the need for timely interventions.

Model Building and Recommendation

Linear Regression Model:

- The linear regression model performed poorly, with an RMSE of 16,396,338.78 and an R^2 value of -9.18, indicating it was not suitable for the data.

Random Forest Model:

- The random forest model significantly outperformed the linear regression model, achieving an RMSE of 2.997 and an R^2 value of 0.693, explaining 69.3% of the variance in loss percentages.

Recommendation:

- The **random forest model** is recommended for analyzing food loss data due to its ability to handle non-linear relationships and categorical variables effectively.

Conclusion

Food loss during storage represents a significant challenge to global food security and sustainability. Data analysis revealed that targeted interventions, such as improved storage methods and pest control, can effectively mitigate losses. These measures not only reduce food insecurity but also support economic growth and environmental conservation.

Assumptions

1. Storage conditions in the dataset accurately represent typical scenarios.
2. Loss percentages were reliably recorded based on controlled experiments.
3. External factors, such as weather conditions, remained constant across observations.

Limitations

1. Data may not generalize to all regions, especially those not represented in the dataset.
2. Missing values in key fields, such as region and sample sizes, could introduce biases.
3. The analysis focuses solely on storage losses, excluding upstream and downstream supply chain stages.

Challenges

1. Lack of modern storage facilities in developing regions.
2. Limited knowledge of effective storage practices among farmers.
3. Insufficient funding for large-scale implementation of recommended treatments.

Future Uses/Additional Applications

1. Expanding the analysis to include data from additional countries and regions.
2. Leveraging machine learning models to predict loss under varying conditions.

Implementation Plan

1. **Data Collection and Baseline Assessment:**
    - Conduct nationwide surveys to evaluate current storage practices and identify high-loss areas.
2. **Pilot Projects:**

    - Implement and monitor recommended interventions in regions with the highest losses.
3. **Scaling and Policy Integration:**
    - Expand successful interventions to other regions and incorporate findings into national and regional policies.
4. **Continuous Monitoring:**
    - Establish a centralized database to track food loss trends and assess intervention efficacy over time.

Ethical Assessment

1. Ensure equitable access to recommendations, particularly in low-income regions.
2. Include diverse regions, commodities, and storage practices in the analysis to avoid bias.
3. Prioritize long-term solutions that balance economic and environmental impacts.
4. Respect local practices and tailor interventions to cultural and logistical contexts.

Citations

FAO. "Food Loss and Waste Database." *Food and Agriculture Organization of the United Nations*, 2024. https://www.fao.org/platform-food-loss-waste/flw-data/en/.