International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF 2023: 8.176 ISSN: 2582-3930

Soil Monitoring and Crop Recommendation System Using Machine Learning and Internet of Things

Dr. P. B. Ambhore[1], Sadiya Abdul Salim[2], Sajeed Khan[3], Yash Ingle[4], Abhay Patle[5]

[1]*Guide of Department of Information Technology, Government college of Engineering Amravati 444604.*
[2]*Student of Department of Information Technology, Government college of Engineering Amravati 444604.*
[3]*Student of Department of Information Technology, Government college of Engineering Amravati 444604.*
[4]*Student of Department of Information Technology, Government college of Engineering Amravati 444604.*
[5]*Student of Department of Information Technology, Government college of Engineering Amravati 444604.*

---***---

Abstract - Agriculture is the pillar of the Indian economy, with over 50% of India's population dependent on it. Variations in weather, climate, and environmental conditions pose significant risks to agriculture. Machine learning (ML) serves as a crucial decision support tool for Crop Yield Prediction (CYP), aiding decisions on crop selection and management during the growing season. CYP involves predicting yields using historical data, meteorological parameters, and past yield records. ML-based crop yield prediction offers accurate forecasts by leveraging diverse datasets, including meteorological conditions, soil characteristics, and historical crop performance. This technology helps farmers make informed decisions, optimize resource allocation, and mitigate risks from environmental unpredictability. Integrating ML into CYP promotes sustainable farming practices, ensuring food security and economic stability amid a dynamic agricultural landscape. The Random Forest regression, a supervised learning model, has demonstrated high performance, achieving an accuracy of 92.3%. This method supports optimal yield forecasting, aiding farmers and policymakers in better planning for crop production and management.

Key Words: agriculture, crop yield prediction, machine learning, meteorological data, Random Forest regression, sustainable farming.

1.INTRODUCTION

Agriculture is pivotal for food security, sustainable development, and poverty elimination in India. However, it faces challenges due to governmental and economic structures, with support mainly for rice and wheat through mechanisms like the minimum support price. Agriculture is the broadest economic sector, contributing significantly to GDP and employing over half the population. To improve crop productivity, researchers use machine learning techniques. Among these, Random Forest classification has proven effective for enhancing yield rates and crop production. Precision agriculture leverages real-time data for smarter decision-making, making predictive models crucial for optimizing crop production in an increasingly digital world.

2. OBJECTIVE

- **Enhancing Agricultural Efficiency**: Predicting crop yields accurately allows farmers to make informed decisions on planting schedules, irrigation, fertilization, and pest control, leading to efficient resource utilization, reduced wastage, and improved productivity.
- **Optimizing Resource Allocation**: Machine learning models analyze factors affecting crop yield to allocate resources like water, fertilizers, and pesticides efficiently, reducing costs and environmental impact. For example, adjusting fertilizer usage based on soil quality predictions prevents over-application and minimizes waste.
- **Mitigating Food Insecurity**: Accurate crop yield predictions improve food production planning and management, helping mitigate food insecurity. Identifying regions at risk of low yields enables policymakers to implement targeted interventions such as agricultural subsidies, infrastructure improvements, and access to credit.
- **Improving Yield Predictions**: Machine learning algorithms use historical data and real-time inputs like satellite imagery, weather forecasts, and IoT sensor data to develop and refine predictive models, providing precise and timely crop yield predictions for optimal decision-making.

3. EXISTING SYSTEM

In the existing system [Fig. 1], there is no computerized system to recommend crops to farmers, leading to confusion about which crop to plant in which season. Farmers often suffer losses due to unpredictable rainfall or weather conditions. Natural disasters can devastate entire farms, collapsing the farmer's annual budget and sometimes leading to extreme distress. Without guidance on appropriate crops for specific lands, farmers may overproduce on unsuitable land, reducing profitability.

The proposed system aims to provide crop prediction solutions. By collecting and training area-wise, season-wise, and condition-specific data using machine learning algorithms, the system can predict suitable crops for different regions. The system considers soil, weather, region, season, and past production data to suggest the most profitable crops under current environmental conditions. This helps farmers make informed decisions about which crops to cultivate.

Additionally, the system uses the Apriori algorithm, which takes real-time feedback from farmers regarding crop conditions, weather, and temperature. The algorithm identifies frequently occurring item sets, calculates the minimum support count, and merges conditions from various regions. It then calculates the support percentage for each crop group and recommends the highest percentage crop to the farmer, enhancing crop selection and potentially introducing farmers to new, profitable crops.

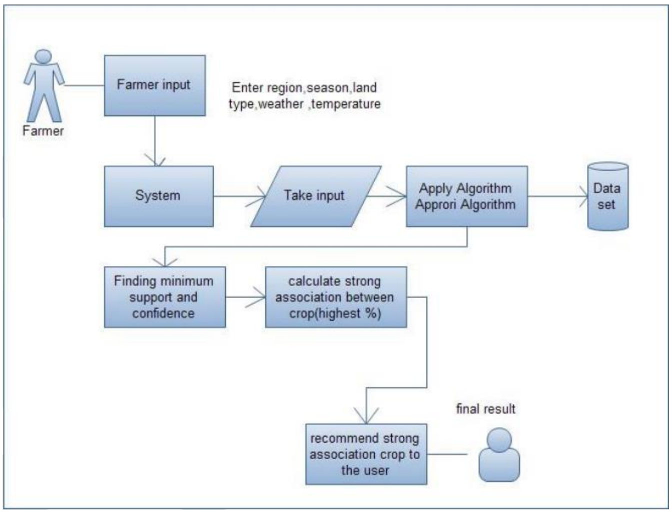


Fig. 1: Existing System

4. METHODOLOGY

4.1 Hardware Requirements

4.1.1 NPK Sensor

An NPK sensor [Fig. 2] measures essential soil nutrients: nitrogen (N), phosphorus (P), and potassium (K). It optimizes fertilizer application for better crop yields using advanced spectroscopy or electrochemical methods to provide accurate, real-time data.

4.1.2 pH Sensor

A pH sensor [Fig. 3] monitors soil acidity, influencing plant growth and nutrient uptake. It measures pH levels to maintain optimal conditions, enhancing nutrient availability and soil health.

4.1.3 Humidity/Temperature Sensor

Humidity sensors [Fig. 4] measure air moisture to optimize growing conditions, prevent fungal diseases, and ensure proper transpiration rates. They enable precise irrigation scheduling to conserve water.

4.1.4 Moisture Sensor

Moisture sensors [Fig. 5] accurately measure soil moisture levels, informing irrigation decisions to prevent overwatering or underwatering, thus promoting healthier crop growth.

4.1.5 Arduino Board

The Arduino Uno [Fig. 6] is a microcontroller with 32KB of flash memory, 2KB SRAM, and 1KB EEPROM. It has 14 digital I/O pins, 6 analog inputs, and operates at 16 MHz programmable via USB or external supply, it supports expansion through shields for added functionality.

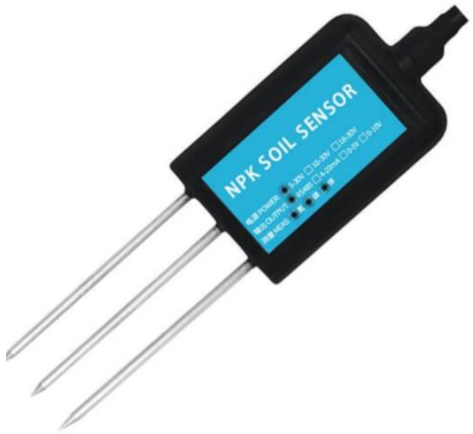


Fig. 2: NPK Sensor

Fig. 3: PH Sensor

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF 2023: 8.176 ISSN: 2582-3930

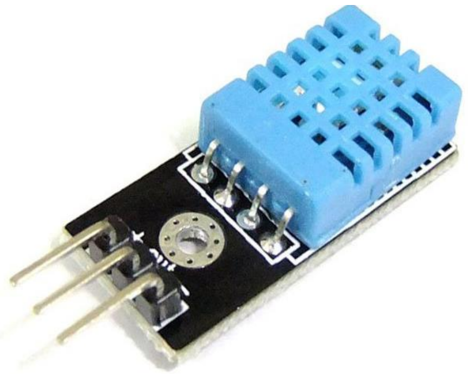


Fig 4: DHT 11 [Humidity + Temperature] Sensor

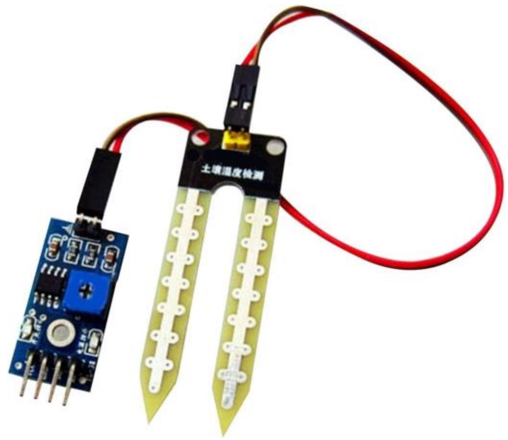


Fig 5: Moisture Sensor

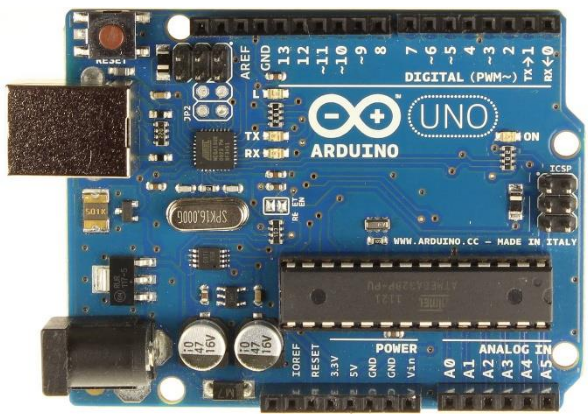


Fig 6: Arduino Uno Board

4.2 Software Requirements

Software requirements encompass various tools and languages such as ReactJS for UI development, HTML/CSS for web page structure and styling, JavaScript for dynamic features, Python/Flask for backend logic, and C++ for complex systems. Additionally, efficient communication is ensured through REST API design. These technologies collectively enable scalable, interactive, and modular web applications, facilitating rapid development and extensive functionality.

4.3 Algorithm Used

4.3.1 Random Forest Algorithm

The Random Forest algorithm plays a crucial role in crop yield prediction, utilizing an ensemble of decision trees trained on subsets of data to enhance accuracy and robustness. It effectively captures complex relationships among features like weather and soil attributes, making it adept at generalizing to various agricultural contexts. Random Forest's strength lies in handling diverse datasets, non-linearity, and determining feature importance. This paper compares Random Forest with logistic regression and naive Bayes for crop yield prediction. Its ability to analyze crop growth amidst changing climatic conditions makes it a valuable tool for agricultural research, offering insights into biophysical changes and informing decision-making for sustainable farming practices. Through its bagging method, Random Forest combines multiple models to provide optimal solutions, highlighting its efficacy in addressing the complexities of crop yield prediction in machine learning (see Fig. 7).

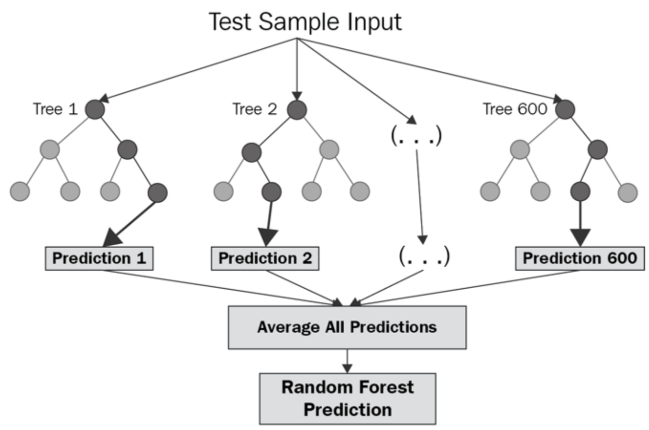


Fig. 7: Random Forest Algorithm Diagram

5. SYSTEM ANALYSIS

5.1 Data Flow Diagram

The data flow diagram (Fig. 8) outlines key interactions in the crop yield prediction system. A continuous feedback loop ensures the machine learning model evolves and becomes more accurate over time. Incoming data is preprocessed to handle missing values, outliers, and ensure uniform formatting, maintaining input quality for training and prediction.

Historical data feeds into the machine learning model during training. The model periodically updates with new datasets to adapt to changing agricultural conditions. Preprocessing steps, including cleaning, transformation, and feature engineering, prepare the data for the model. During training, data flows into algorithms like regression or neural networks, followed by validation to assess accuracy. The diagram shows the output of predicted crop yields, completing the data flow cycle. Feedback loops incorporate new data to iteratively improve predictions.

The core system analyzes input data to generate future crop yield predictions based on learned patterns and relationships.

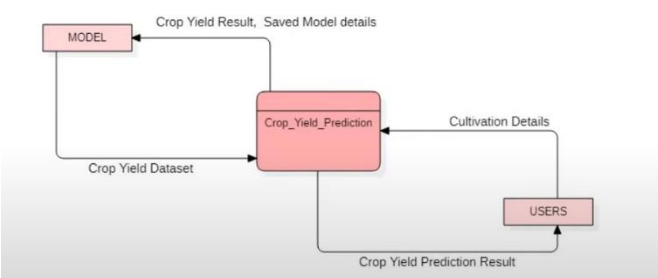


Fig. 8: Data Flow Diagram

5.2 Implementation

5.2.1 Data Collection

```
from __future__ import print_function
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import classification_report
from sklearn import metrics
from sklearn import tree
import warnings
warnings.filterwarnings('ignore')
```

```
PATH = 'Crop_recommendation.csv'
df = pd.read_csv(PATH)
```

```
df.head()
```

	N	P	K	temperature	humidity	ph	label
0	90	42	43	20.879744	82.002744	6.502985	rice
1	85	58	41	21.770462	80.319644	7.038096	rice
2	60	55	44	23.004459	82.320763	7.840207	rice
3	74	35	40	26.491096	80.158363	6.980401	rice
4	78	42	42	20.130175	81.604873	7.628473	rice

Fig. 9: Data Collection

Above [Fig. 5.3] outlines a data collection process for crops, capturing NPK value, temperature, humidity, and pH levels, facilitating informed agricultural decision-making.

5.2.2 Data preprocessing

```
[ ] df.size
15400

[ ] df.shape
(2200, 7)

[ ] df.columns
Index(['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'label'], dtype='object')

[ ] df['label'].unique()
array(['rice', 'maize', 'chickpea', 'kidneybeans', 'pigeonpeas',
       'mothbeans', 'mungbean', 'blackgram', 'lentil', 'pomegranate',
       'banana', 'mango', 'grapes', 'watermelon', 'muskmelon', 'apple',
       'orange', 'papaya', 'coconut', 'cotton', 'jute', 'coffee'],
      dtype=object)
```

Fig. 10: Data Preprocessing code

Above [Fig. 10] outlines a data processing involves analyzing collected information on crop NPK values, temperature, humidity, and pH levels to derive actionable insights. Techniques such as normalization, feature extraction, and statistical analysis streamline raw data into interpretable patterns, aiding in decision-making for crop management and optimization.

5.2.2 Feature Selection

Seperating features and target label

```
[11] features = df[['N', 'P','K','temperature', 'humidity', 'ph']]
     target = df['label']
     labels = df['label']

[12] # Initializing empty lists to append all model's name and corresponding name
     acc = []
     model = []

[13] # Splitting into train and test data

     from sklearn.model_selection import train_test_split
     Xtrain, Xtest, Ytrain, Ytest = train_test_split(features,target,test_size = 0.2,random_state =2)
```

Fig. 11: Feature Selection

Above [Fig. 11] identifies key variables—NPK values, temperature, humidity, pH levels—using correlation analysis and information gain to optimize predictive accuracy and efficiency.

5.2.3 Model Selection:

```
# Calculate average attribute values for each crop in the training data
average_attribute_values = train_data.groupby('label').mean()

# Display the average attribute values for the predicted crop
predicted_crop_average = X_test.mean()

print("\nAverage attribute values for the predicted crop:")
print(predicted_crop_average)

# Determine which crop should be grown based on the test data attributes
predicted_crop = RF.predict([predicted_crop_average.values])[0]
print("\nBased on the soil attributes, the recommended crop to grow is:", predicted_crop)
```

Above [Fig. 5.6] shows the code for model selection. Select a machine learning algorithm, such as a content-based recommendation algorithm, that is appropriate for the problem and data.

5.2.4 Model Training:

Random Forest

```
7] from sklearn.ensemble import RandomForestClassifier

RF = RandomForestClassifier(n_estimators=20, random_state=0)
RF.fit(Xtrain,Ytrain)

predicted_values = RF.predict(Xtest)

x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('RF')
print("RF's Accuracy is: ", x)

print(classification_report(Ytest,predicted_values))
```

Fig. 12: Model Training

Above [Fig.5.7] shows the code for model training. Train the machine learning model on the preprocessed data to learn the relationships between the features and user preferences.

5.2.4 Model Testing

RF's Accuracy is: 0.9659090909090909

	precision	recall	f1-score	support
apple	1.00	1.00	1.00	13
banana	1.00	1.00	1.00	17
blackgram	0.78	0.88	0.82	16
chickpea	1.00	1.00	1.00	21
coconut	1.00	1.00	1.00	21
coffee	1.00	1.00	1.00	22
cotton	1.00	1.00	1.00	20
grapes	1.00	1.00	1.00	18

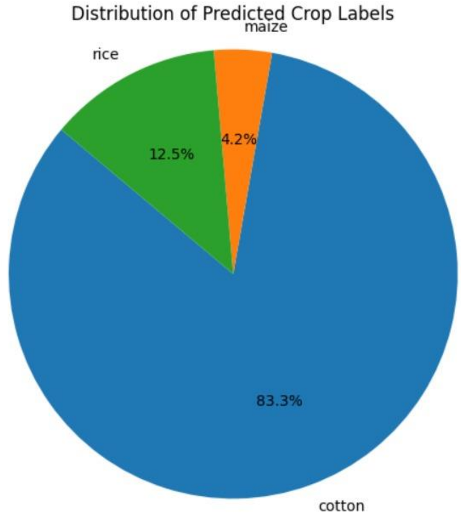
Fig. 13: Model Testing

[Fig. 13] Evaluate the performance of the model using metrics such as accuracy, precision.

Above identifies key variables—NPK values, temperature, humidity, and pH levels—using correlation analysis and information gain to optimize predictive accuracy and efficiency.

6. OUTPUT/RESULTS

6.1 Backend Output




```
Average attribute values for the predicted crop:
N              114.500000
P               47.791667
K               21.416667
temperature     23.971407
humidity        80.255968
ph               6.736649
dtype: float64

Based on the soil attributes, the recommended crop to grow is: cotton
```

Fig. 14: Predicted Crop Cotton

6.2 Frontend Output

Fig. 15: Home Page

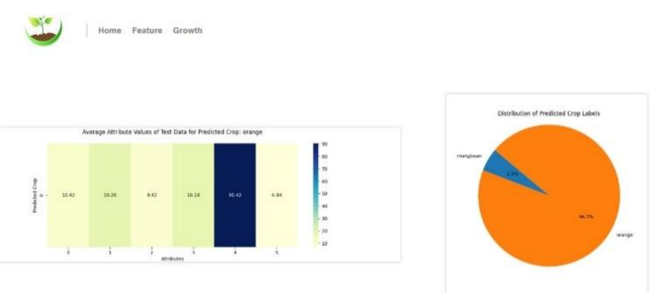


Fig. 16: Prediction Page

[Fig. 15 & 16] displays the prediction page of a website featuring forecasts for various crops based on sophisticated modeling techniques and comprehensive data analysis.

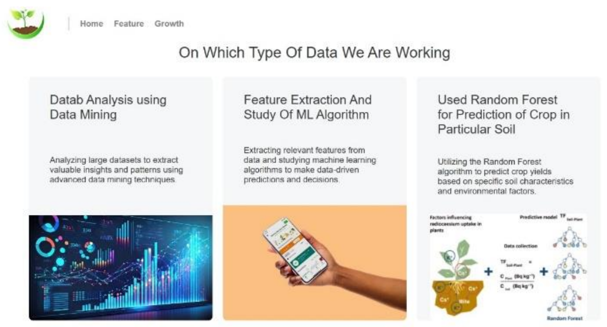


Fig. 17: Visualization Techniques

7. CONCLUSION

In conclusion, using machine learning for crop yield prediction holds great promise for revolutionizing agriculture. By harnessing the power of data and smart algorithms, we can make more informed decisions about when, where, and how to grow crops. This technology has the potential to increase efficiency, optimize resource use, and contribute to global food security. While challenges exist, ongoing research and advancements in machine learning techniques offer exciting possibilities for a more sustainable and productive future in agriculture.

The primary objective of crop yield prediction with machine learning is to empower farmers with accurate and timely insights into the factors influencing crop growth. By leveraging diverse datasets encompassing meteorological conditions, soil attributes, and historical crop performance, machine learning algorithms can discern intricate patterns and relationships. The ability to analyze and learn from this wealth of information equips these models to make precise predictions regarding crop yields, aiding farmers in making informed decisions throughout the cultivation cycle.

One noteworthy advantage of this approach is the personalized nature of recommendations. Farmers can input specific conditions such as weather, soil type, temperature, and region, ensuring crop recommendations are tailored to their unique farm characteristics. This customization allows for optimized decision-making and a shift towards precision agriculture. Additionally, the system's scalability enhances prediction accuracy and accommodates diverse agricultural landscapes, while incorporating advanced technologies like IoT devices and remote sensing further refines predictive capabilities, enabling a more granular understanding of agricultural conditions.

ACKNOWLEDGEMENT

Dr. P. B. Ambhore, Department of Information Technology, Government College of Engineering, Amravati, has been a tremendous source of support and guidance to the authors throughout this project. Their moxie and experience have been inestimable in suiting this work. fortuitous to have had such a surprising

REFERENCES

1. Molden, D.; Vithanage, M.; de Fraiture, C.; Faures, J.M.; Gordon, L.; Molle, F.; Peden, D. Water Availability and Its Use in Agriculture. *Treatise Water Sci.* 2011, *4*, 707–732. [Google Scholar] [CrossRef]
2. Keen, B.A. Weather and crops. *Q. J. R. Meteorol. Soc.* 1940, *66*, 155–166. [Google Scholar] [CrossRef]
3. Javadinejad, S.; Eslamian, S.; Askari, K.O.A. The analysis of the most important climatic parameters affecting performance of crop variability in a changing climate. *Int. J. Hydrol. Sci. Technol.* 2021, *11*, 1–25. [Google Scholar] [CrossRef]
4. Beillouin, D.; Schauberger, B.; Bastos, A.; Ciais, P.; Makowski, D. Impact of extreme weather conditions on European crop production in 2018: Random forest—Yield anomalies. *Philos. Trans. R. Soc. B Biol. Sci.* 2020, *375*, 20190510. [Google Scholar] [CrossRef]
5. Uleberg, E.; Hanssen-Bauer, I.; van Oort, B.; Dalmannsdottir, S. Impact of climate change on agriculture in Northern Norway and potential strategies for adaptation. *Clim. Change* 2014, *122*, 27–39. [Google Scholar] [CrossRef]
6. Yadav, M.R.; Choudhary, M.; Singh, J.; Lal, M.K.; Jha, P.K.; Udawat, P.; Gupta, N.K.; Rajput, V.D.; Garg, N.K.; Maheshwari, C.; et al. Impacts, Tolerance, Adaptation, and Mitigation of Heat Stress on Wheat under Changing Climates. *Int. J. Mol. Sci.* 2022, *23*, 2838. [Google Scholar] [CrossRef]
7. Gallego, A.; Carrión, M.C.; Ruiz, D.P.; Medouri, A. The relationship between AR-modelling bispectral estimation and the theory of linear prediction. *Signal Process* 1994, *37*, 381–388. [Google Scholar] [CrossRef]
8. US AID. Climate Change Risk Profile: Philippines. 2017, pp. 1–4. Available online: https://www.climatelinks.org/sites/default/files/asset/document/2017_Climate_Change_Risk_Profile_Philippines.pdf (accessed on 2 February 2022).
9. EastAfrican. 3000 Rwandan Families Face Hunger due to Drought—Rwanda|ReliefWeb, (n.d.). Available online: https://reliefweb.int/report/rwanda/3000-rwandan-families-face-hunger-due-drought (accessed on 3 February 2022).