INSTITUTE OF ENGINEERING & MANAGEMENT
Good Education, Good Jobs

UNIVERSITY OF ENGINEERING & MANAGEMENT
Good Education, Good Jobs

AGROGENIE: A SMART APPROACH TO AGRICULTURE USING MACHINE LEARNING

Authors: Abhiroop Sarkar, Kinshuk Ganguly, Snehal Ghosh, Debayan Ghosh, Subhajit Saha, Piyali Datta, and Deepsubhra Guha Roy
Institute of Engineering & Management, Kolkata
Presentation by Abhiroop Sarkar| Paper ID-092

3rd International Conference on Intelligent Systems and Applications ICISA | 2024

OUTLINE

OF THE PRESENTATION

- Introduction
- Literature Survey
- Methodology - Initial and Improved Approach
- Results and Analysis
- Future Scope
- Conclusion

INTRODUCTION

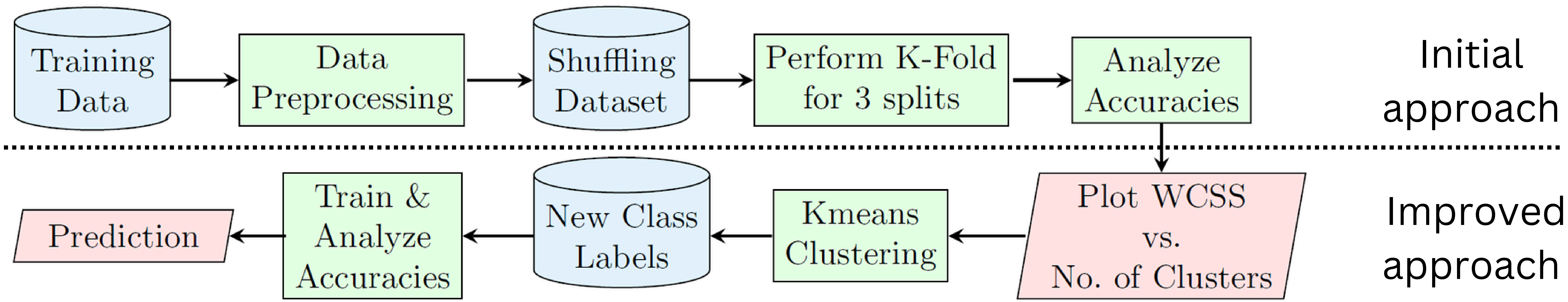
1. Machine Learning applications
    - Health, defense, education, and urbanization
2. Importance of Agriculture in India
    - Agriculture utilizes **60% of the land resources** in India, it is a significant industry that provides food for the nation’s 1.4 billion inhabitants.
3. Machine Learning in Agriculture
    - Typical practice to **forecast and analyze crop development** using machine learning approaches.
    - **IoT and ML technologies** are integrated with smart farming to maximize output and reduce resource use.
    - ML models built on large-scale farm data **enhance crop growth** and **output analysis and prediction** with high accuracy.
4. Our system would benefit **Farmers, Residents of drought-prone areas, Environmentalists,** and **Impoverished People.**

LITERATURE SURVEY

- **Kumar et al. 2019:** Utilized an SVM model on a dataset from 15 different crops, achieving 89.66% accuracy.
- **Pudumalar et al. 2017:** Employed Naive Bayes and KNN for crop yield prediction, with additional use of Artificial Neural Networks and Multi-layer perceptron for food production and import predictions.
- **Ruchirawya T. 2020:** Implemented Arduino microcontrollers for data collection and various ML techniques including NLP for sentiment analysis, achieving over 92% accuracy, potentially improving to 95%.
- **Udutalapally et al. 2020:** Developed a CNN model for plant disease prediction and a real-time health maintenance system, achieving 99% accuracy and demonstrating resilience to rust and weather variations in a three-month trial.
- **Madhuri Shripathi Rao et al. 2022:** Analyzed KNN, Decision Tree, and Random Forest on the same dataset, finding Random Forest to have the highest accuracy of 99.32%.

PLAN

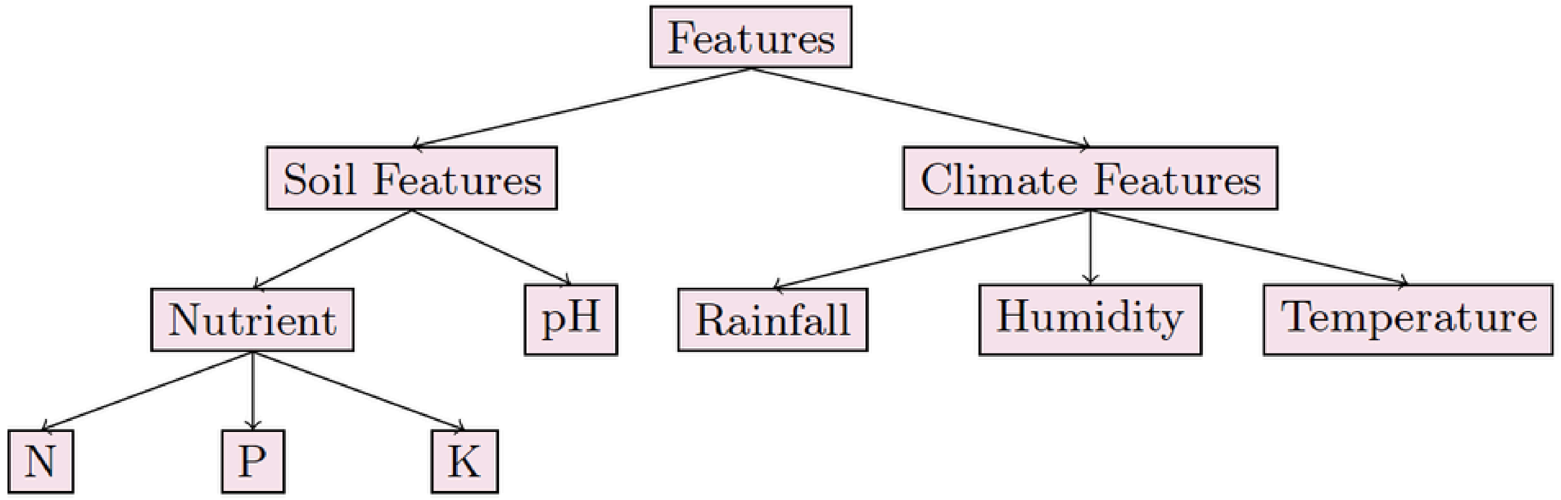
Flowchart



INITIAL APPROACH

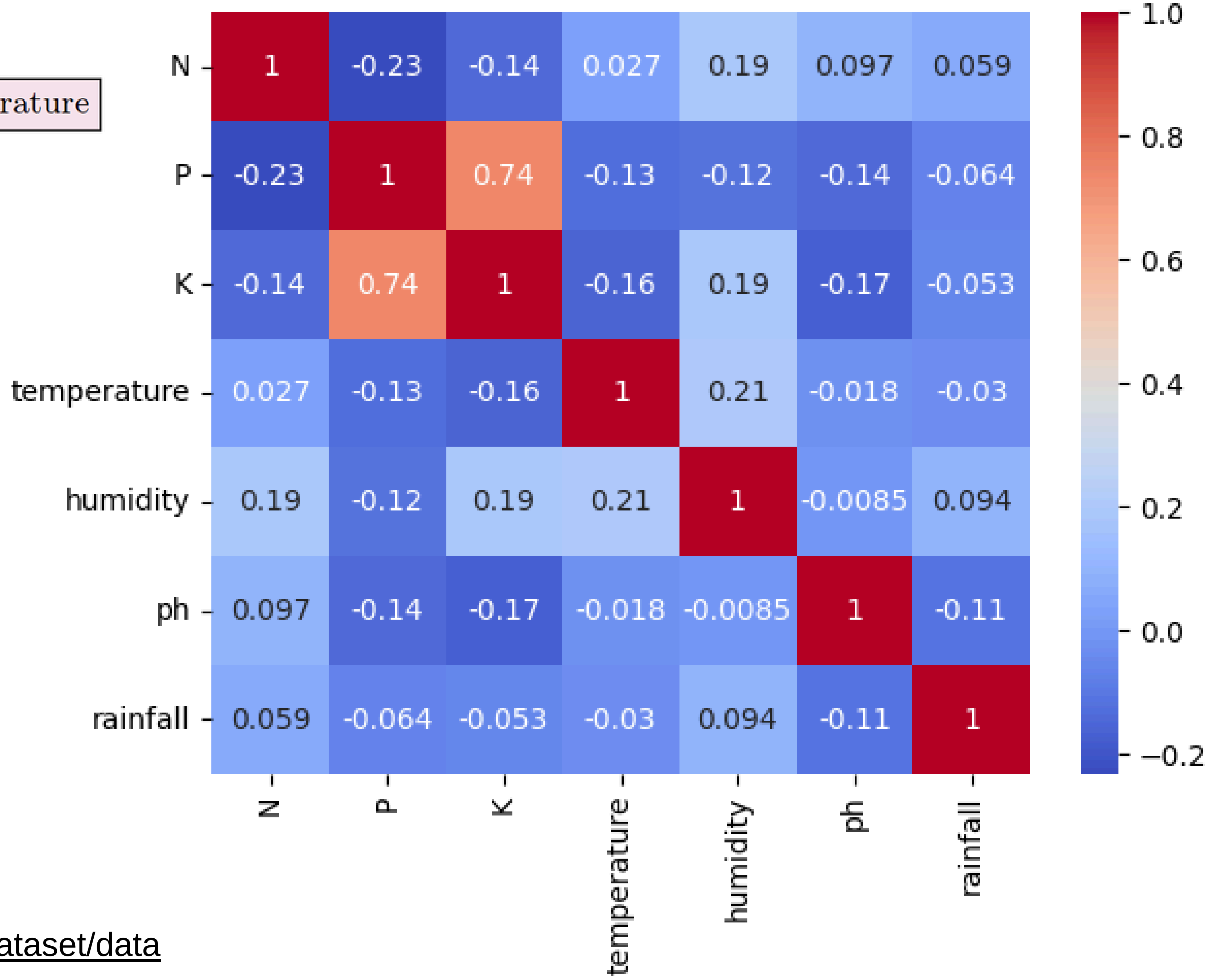
- The dataset comprises **2200 instances** with **7 features** and **22 crop labels.**
- **Null values** and **duplicate instances** are absent.
- **The correlation matrix** and the **corresponding heatmap** reveal a **negligible correlation** between features except N, P, and K values of soil.
- **The dataset was shuffled** to prevent any bias in training due to the sequential arrangement of the dataset.
- **Min-max scaling** and **standard scaling** have been performed to decrease biases arising from abnormal data values in the dataset.

INITIAL APPROACH

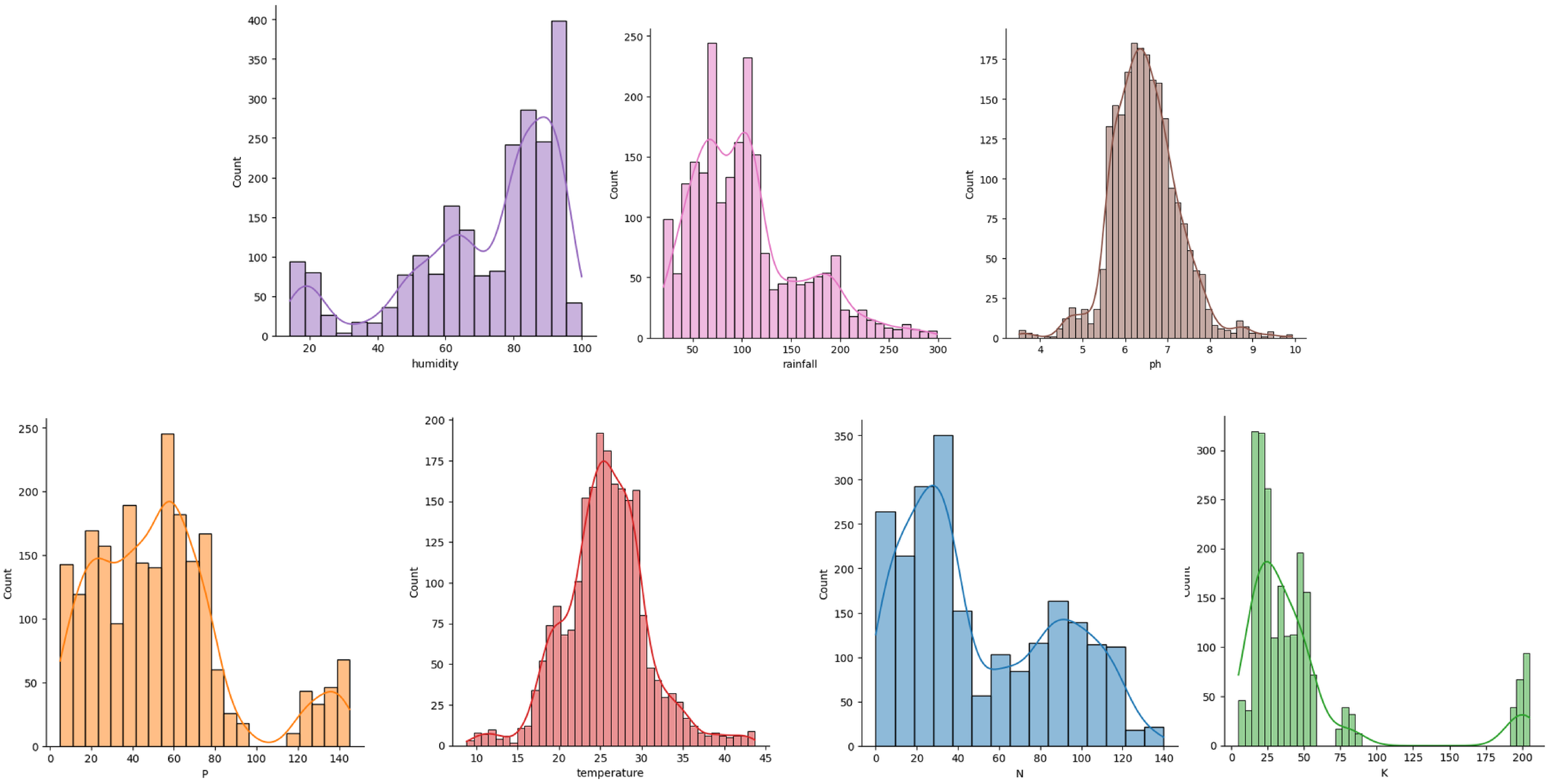


Feature Tree

Heatmap



https://www.kaggle.com/datasets/atharvaingle/crop-recommendationdataset/data



Displots for 7 features

IMPROVED APPROACH

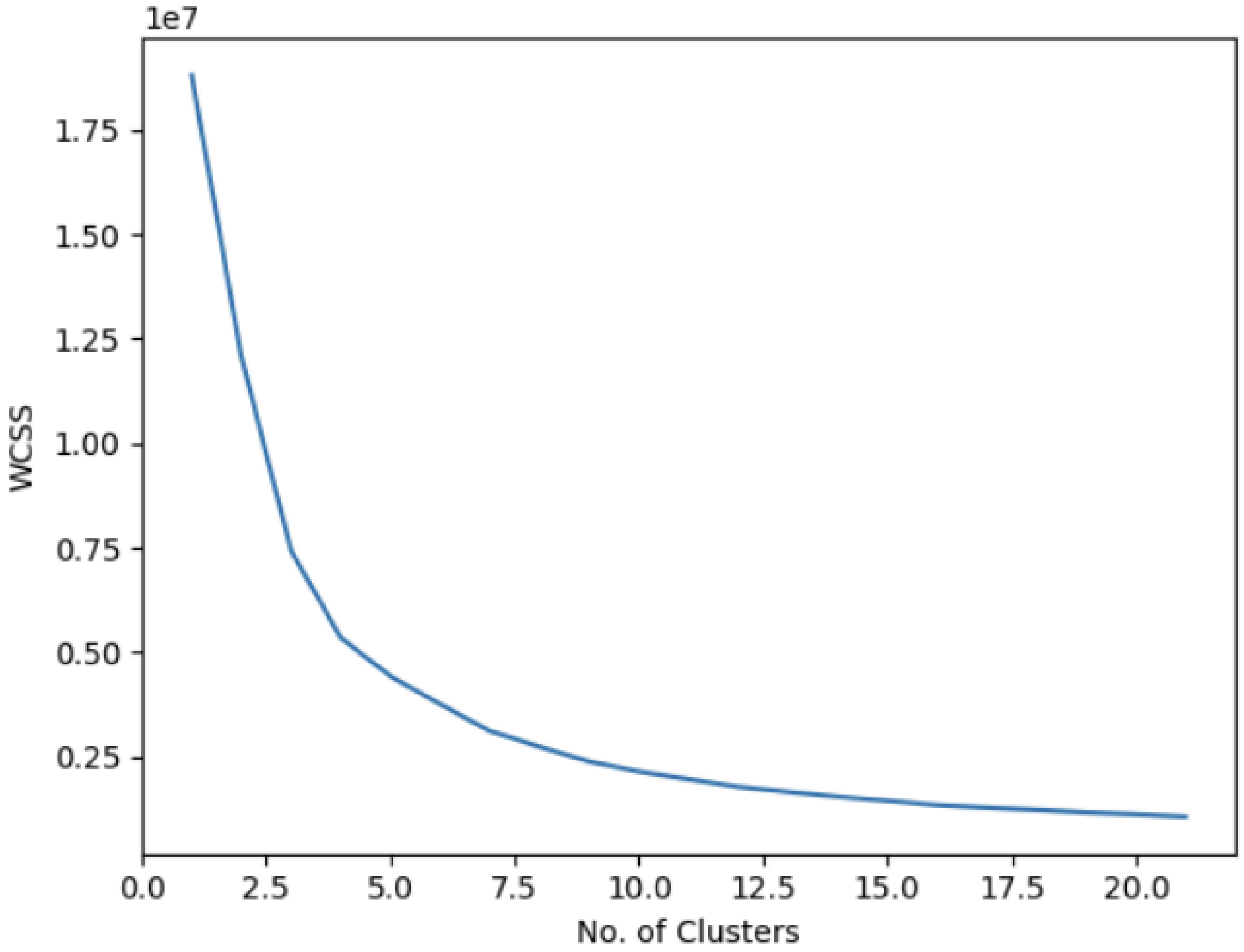
Motivation behind clustering:

We use **K-means clustering** to group crops with **similar features**, allowing farmers to select crops based on their resources without pressure to adopt new crops or change their existing practices. Thus, it is an example of **clustering followed by classification.**

Selecting the number of clusters:

An elbow curve plotting the **number of clusters** against **Within Cluster Sum of Squares (WCSS)** was constructed.

A distinct bend between **8 to 11 clusters**, as seen in Figure 'X' was observed



IMPROVED APPROACH

Training and selection of classifier model

- Identified **10 optimal clusters** via **elbow curve analysis** and trained **five models** using **3-fold K-fold cross-validation.**
- We selected the **Random Forest Classifier** for deeper analysis due to its superior performance.

Developing the final predictive model

- Formed a dictionary mapping **ten clusters** to **crop labels** and their **frequencies** from the original dataset.
- Only crop labels that appear **predominantly (over 50%)** in a cluster are considered to **minimize outlier influence.**
- The trained model **predicts the cluster** for a given input vector and displays **dominant crop labels** associated with the predicted cluster.

RESULT AND ANALYSIS

Accuracy of Different Models about the number of clusters formed

Models	Average Accuracy models as per number of clusters		
	9 clusters	10 clusters	11 clusters
Logistic regression	94.5000	94.500	94.5005
Support Vector Machine	92.7277	92.2276	92.7277
Random Forest Classifier	97.3636	97.9545	97.4546
K Nearest Neighbour	93.0455	92.2732	93.0455
Decision Tree	96.1816	96.7724	95.9545

Cluster Information and Predominant Crops in Bold

Cluster	Predominant Crops	Frequency
0	Lentil, Chickpea, Pigeonpeas, **Kidneybeans**	4,5,32,100
1	Papaya, Coffee, **Maize, Cotton, Banana**	8, 38, 99, 100, 100
2	Pigeonpeas, Papaya, **Coffee, Jute, Rice**	1, 18, 62, 100, 100
3	**Apple, Grapes**	100, 100
4	**Watermelon, Muskmelon**	100
5	Mothbeans, **Pigeonpeas, Mango**	42, 55, 89
6	Maize, Orange, Pigeonpeas, Mango, Papaya, **Mothbeans, Lentil, Blackgram, Mungkbean**	1, 2, 9, 11, 13, 58, 96, 100, 100
7	Pigeonpeas, Orange, **Papaya**	3, 18, 58
8	Papaya, **Orange, Coconut, Pomegranate**	3, 80, 100, 100
9	**Chickpea**	95

RESULT AND ANALYSIS

Merged Classification Reports for Folds 1, 2, and 3 for Random Forest

Folds	Fold 1				Fold 2				Fold 3			
Class	Prec	Rec	F1-S	Supp.	Prec	Rec	F1-S	Supp.	Prec	Rec	F1-S	Supp.
0	0.98	0.96	0.97	52	0.97	0.90	0.94	42	0.98	0.98	0.98	47
1	0.96	0.97	0.97	109	0.99	1.00	1.00	116	0.98	0.99	0.98	120
2	0.98	0.98	0.98	106	1.00	0.99	0.99	100	0.96	0.96	0.96	75
3	1.00	1.00	1.00	57	1.00	1.00	1.00	67	1.00	1.00	1.00	76
4	1.00	1.00	1.00	62	1.00	1.00	1.00	64	1.00	1.00	1.00	74
5	0.93	1.00	0.96	65	0.94	0.94	0.94	53	0.97	0.99	0.98	68
6	0.99	0.95	0.97	130	0.94	0.96	0.95	137	0.98	0.98	0.98	123
7	0.96	0.93	0.95	29	0.92	0.88	0.90	26	0.92	0.92	0.92	24
8	1.00	1.00	1.00	89	0.99	1.00	1.00	102	1.00	0.97	0.98	92
9	0.97	1.00	0.99	35	0.96	1.00	0.98	26	1.00	1.00	1.00	34
Accuracy			0.98	734			0.98	733			0.98	733
Macro Avg	0.98	0.98	0.98	734	0.97	0.97	0.97	733	0.98	0.98	0.98	733
Wt. Avg	0.98	0.98	0.98	734	0.98	0.98	0.98	733	0.98	0.98	0.98	733

RESULT AND ANALYSIS

Crops Prediction based on Input Values

N	P	K	Temperature	Humidity	pH	Rainfall	Prediction
61	37	40	27	92	5.5	148	Rice, Jute, Coffee
99	18	47	31	85	6.2	27	Muskmelon, Watermelon

Accuracy Scores for Random Forest Models

Model	Accuracy (%)
1	97.9564
2	97.6807
3	98.2265
Average	**97.9545**

FUTURE SCOPE

- We plan to implement an **IoT-based system** for **on-field data collection** for the required parameters for prediction and integrate it with the existing ML system.
- Implement the concept of Fuzzy logic by **Fuzzy C-means Clustering.**
- Limited by the **size of the dataset.**

CONCLUSION

- Utilized **K-means clustering** on a dataset of **2200** entries with **7 features** and **22 crop labels** to form new class labels comprising multiple crops under each label.
- Found the **optimal number of clusters** as **10** after plotting the elbow curve between WCSS and No. of clusters
- Employed **K-fold cross-validation** to evaluate performance, identifying the **Random Forest classifier** as the most accurate, achieving a **97.95% accuracy rate.**
- As precision agriculture increasingly relies on **data-driven decision making**, the outcomes show the practical application of **combined unsupervised and supervised learning** techniques in enhancing the efficiency of precision agriculture practices.

REFERENCES

- Li, L., Wang, B., et al.: Developing machine learning models with multi-source environmental data to predict wheat yield in china. Computers and Electronics in Agriculture 194, 106790 (2022)
- Van Klompenburg, T., Kassahun, A., Catal, C.: Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture 177, 105709 (2020)
- Kuradusenge, M., et al: Crop yield prediction using machine learning models: case of irish potato and maize. Agriculture 13(1), 225 (2023)
- Shaikh, T.A., et al..: Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. Computers and Electronics in Agriculture 198, 107119 (2022)
- Kumar, A., Sarkar, S., Pradhan, C.: Recommendation system for crop identification and pest control technique in agriculture. pp. 0185–0189 (04 2019).
- Pudumalar, S., et al.: Crop recommendation system for precision agriculture. In: 2016 Eighth International Conference on Advanced Computing (ICoAC). pp. 32–36 (2017).
- Ruchirawya, T.: Crop recommendation system. vol. 175 (10 2020)
- Udutalapally, V., et al.: Scrop: A novel device for sustainable automatic disease prediction, crop selection, and irrigation in internet-of-agro-things for smart agriculture. IEEE Sensors Journal PP (10 2020).
- Rao, M.S., et al..: Crop prediction using machine learning. Journal of Physics: Conference Series 2161(1), 012033 (jan 2022).
- Yaswanth Kumar Alapati, K.S.: Combining clustering with classification: A technique to improve classification accuracy. International Journal of Computer Science Engineering (IJCSE) 5(06) (11 2016)

THANK YOU

Presentation by Abhiroop Sarkar

QR for Repository