Nano World Journal

https://doi.org/10.17756/nwj.2023-s4-074

Research Article

Open Access

Harvest Prediction by KNN Classification

J. Yeswanth Sai Mahesh*, A. Mourya Bharadwaj, Shravan Nikhil and V. Balaji

Department of Computer Science and Engineering (AIML), Vardhaman College of Engineering, Hyderabad, Telangana, India

*Correspondence to:
J. Yeswanth Sai Mahesh
Department of Computer Science and Engineering (AIML),
Vardhaman College of Engineering,
Hyderabad, Telangana, India.
E-mail: yeswanthsaimahesh@gmail.com

Received: September 19, 2023
Accepted: November 29, 2023
Published: December 04, 2023

Citation: Mahesh JYS, Bharadwaj AM, Nikhil S, Balaji V. 2023. Harvest Prediction by KNN Classification. *Nano World J* 9(S4): S440-S442.

Published by United Scientific Group

Abstract

Agriculture is an important field all over the world where there are many challenges facing by the farmers. This has become a problem for developing countries. Using latest technologies many companies are using Internet of Things based services to reduce manual work. These methods are mostly useful in the case on reducing human work but not in prediction process. In this project crop yield prediction using latest machine learning technology and KNN (K-Nearest Neighbor) classification algorithm is used for prediction crop yield based on soil, temperature, and weather factors. We also prepared dataset with various soil conditions as features and labels for predicting type of each label is related to certain crop. In prediction process farmers can give input their soil features and result will be type of crop suitable for specific conditions and application also helps in suggesting best crops.

Keywords

KNN algorithm, Soil temperature, Reduce manual work

Introduction

Nanotechnology is a promising technology that has the potential to increase agricultural output by using nanofertilizers, effective herbicides, and insecticides, soil feature regulation, wastewater management, and disease detection. It is also beneficial for industrial food processing since it boosts food production with high market value, improves nutritional and sensory characteristics, promotes safety, and provides great antibacterial protection. By extending shelf life with nanoparticles, nanotechnology can help reduce post-harvest losses [1].

Nanotechnology is utilized to encapsulate nanopesticides and to create active nanomaterials. Polymer-based nanosystems can be used as encapsulation materials, whilst nanosized nanomaterials can be used as active components in nanopesticides. Different processes are used to create nanocapsules, and the active chemicals are then placed into the polymeric capsule. Nanospheres are similar to nanocapsules in construction; however, the material polymerizes with the active component and traps nanopesticides during production. Micelles are specifically employed to supply insoluble water [2].

The existing model will help the farmers to know only about their crop situation at the present time by based on their previous data information's it is helpful for them up to only the crop prediction what there are getting the output in the form of profit. The main thing is the to make farmer to sell their crop by their independent way.

According to the results, temperature is best predicted by the ARIMA model, and the accuracy of predictions made for rainfall by ARMA model is also good. Rainfall, which is an important factor for the prediction of crop yield is difficult to estimate precisely.

Climate factors may change due to other remaining variables which may influence the prediction of rainfall. Also, the proposed work makes use of fuzzy logic to estimate crop yield which works on a set range rather than discrete values, therefore, the error in predicted rainfall data does not cause problems as long as the difference between actual and estimated values is not drastic.

Experimentation

The research of our project is developed in the system which can help the farmers. In this module, rainfall water, temperature data set, groundwater level data set, soil types of data sets is taken for Indian data for past ten years. The data is converted into data frame and pre-processed such that zero values in all column's records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then the model is prepared for the given data set and predicted for upcoming years. Using ts. plot () the upcoming years values are plotted. We use the 'ARIMA' (it is a model used for forecasting or future predicting based on a historical times) function, the model is prepared for the given data set and predicted for upcoming years (Figure 1 and figure 2).

In the context of using KNN classification for harvest prediction, control factors and levels refer to the variables and their respective values that you can manipulate or study to understand their impact on the prediction accuracy and performance of the KNN model. Here are some control factors and levels to consider.

K-Value (Number of neighbors)

- Control levels: Vary K from 1 to a reasonable upper limit (e.g., 1, 3, 5, 7, 10, 15).
- Impact: Determine the optimal number of neighbors for KNN by comparing classification accuracy at different K values.

Distance metric

- Control levels: Experiment with different distance metrics such as Euclidean distance, Manhattan distance, or custom-defined metrics.
- Impact: Assess how the choice of distance metric affects the accuracy of harvest prediction.

Feature selection/extraction

- Control levels: Vary the set of features used for prediction. Include/exclude relevant features.
- Impact: Evaluate which features have the most significant impact on prediction accuracy and the model's robustness.

Data pre-processing techniques

- Control levels: Apply various preprocessing techniques like normalization, standardization, or feature scaling.

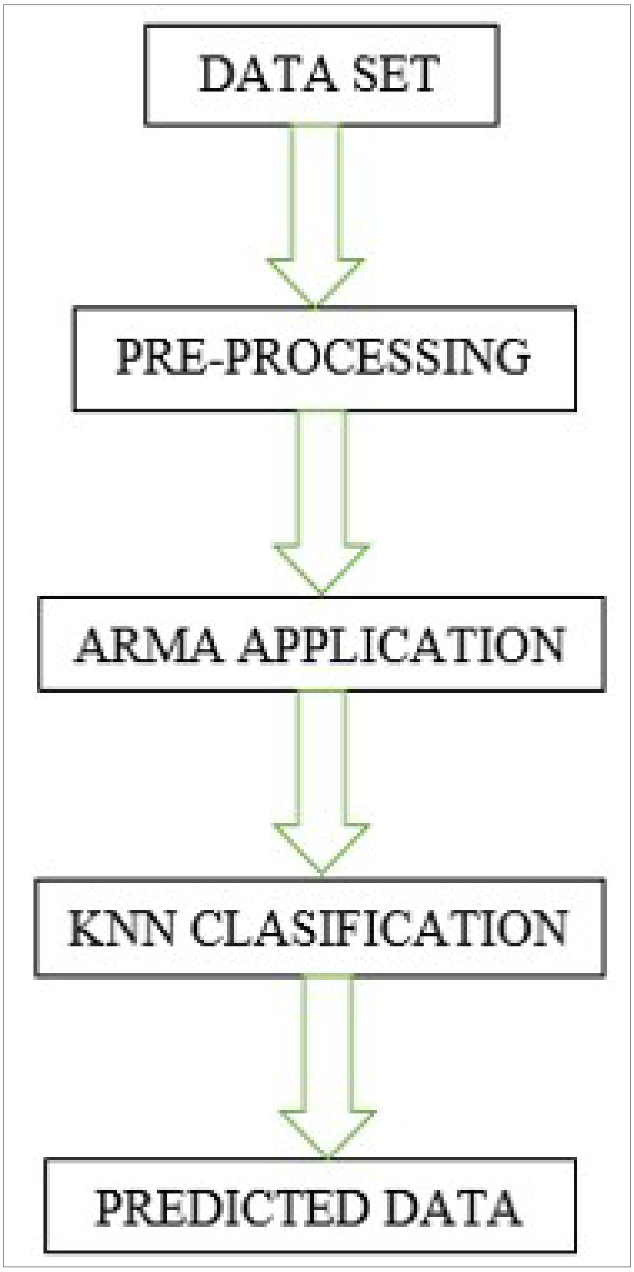

Figure 2: Methodology.

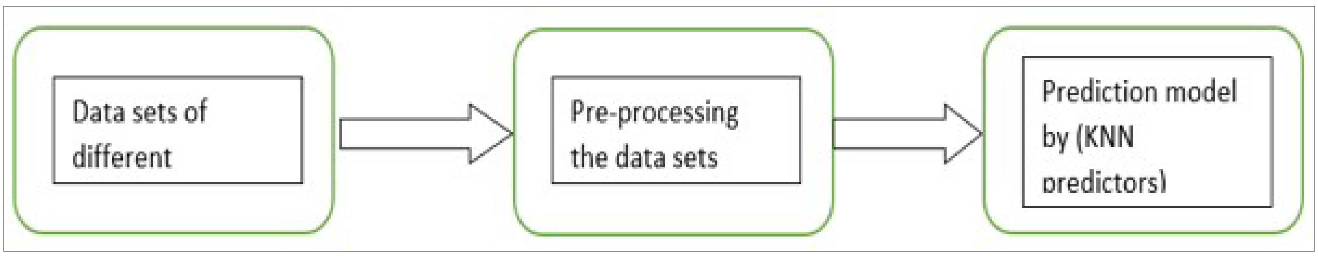

Figure 1: Steps.

- Impact: Determine the effect of data preprocessing on the model's performance.

Data splitting strategy (Training and testing sets)

- Control levels: Adjust the ratio of training data to testing data (e.g., 70-30, 80-20, 90-10).
- Impact: Investigate how different data splitting strategies affect model performance and generalization.

Feature engineering

- Control levels: Create new features, transform existing ones, or engineer domain-specific features.
- Impact: Assess the impact of feature engineering on the model's ability to predict harvest outcomes.

Handling imbalanced data

- Control levels: Explore different techniques for handling imbalanced classes, such as over-sampling, under-sampling, or synthetic data generation.
- Impact: Evaluate the model's performance when dealing with imbalanced datasets.

Algorithm variations

- Control levels: Experiment with variations of the KNN algorithm, such as weighted KNN or kernelized KNN.
- Impact: Compare the performance of different KNN variations in harvest prediction.

Hyperparameter tuning

- Control levels: Tune hyperparameters like leaf size, algorithm choice (e.g., brute force vs ball tree), or distance weighting.
- Impact: Optimize the KNN model's hyperparameters for better prediction accuracy.

Cross-validation techniques

- Control levels: Implement various cross-validation methods, such as k-fold cross-validation or leave-one-out cross-validation.
- Impact: Determine how different cross-validation techniques affect the model's robustness and generalization.

By systematically varying these control factors and levels, you can gain insights into how to configure and fine-tune your KNN classification model for accurate harvest prediction. This experimentation process can help you identify the most effective settings and strategies for your specific dataset and application.

Results and Discussion

The figure 3 is based on the farmers death of the farmers for not getting proper crops for the years and as per our survey many of the farmers had died due to the crop loss in the time and not getting the crop in time and crop failure due weather forecast and also the model can successfully predict ground water level for a given year when the previous years' value is known. By analysing figure 3, got to know that may of the farmers are died due to loss of crops and due to wrong weather they are losing. So, we had decided to build the system to save them from the predictions.

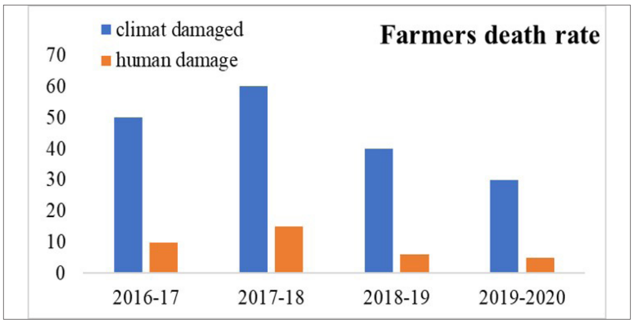

Figure 3: Data.

Conclusion

The proposed work makes use of fuzzy logic to estimate crop yield which works on a set range rather than discrete values, therefore, the error in predicted rainfall data does not cause problems as long as the difference between actual and estimated values is not drastic. The model can successfully predict crop yield for a given year when the rainfall and temperature values for the previous years is known. The model can successfully predict ground water level for a given year when the previous years' value is known. In addition, this project classifies the ground water level data set records using KNN to predict the model for future test record data sets. It will be helpful in analysing the ground water levels in the past and so as to predict the future levels. In future, logistic regression can be applied to further classify the data.

Acknowledgements

None.

Conflict of Interest

None.

References

1. Mali SC, Raj S, Trivedi R. 2020. Nanotechnology a novel approach to enhance crop productivity. *Biochem Biophys Rep* 24: 100821. https://doi.org/10.1016/j.bbrep.2020.100821
2. Neme K, Nafady A, Uddin S, Tola YB. 2021. Application of nanotechnology in agriculture, postharvest loss reduction and food processing: food security implication and challenges. *Heliyon* 7(12): e08539. https://doi.org/10.1016/j.heliyon.2021.e08539