International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Weather Sense: Scraping and Deep Learning for Weather Analysis and Prediction

Devidas Bhat* [1], Balasubramani R.[2]

Submitted: 01/10/2023 **Revised**: 21/11/2023 **Accepted**: 01/12/2023

Abstract: Agriculture is essential in ensuring food security and development in the country. Maximizing scarce arable land is a pressing challenge in today's urbanization era. Agriculture can be made more efficient using technology and information science. This article presents an integrated approach to education in Indian agriculture that uses climate data to accurately analyze environmental factors such as temperature, soil, wind speed, and precipitation. The framework chooses the most accurate algorithm based on analysis and comparison. By providing accurate weather information, farmers can make informed decisions about planting, pest and disease management, and other factors affecting crop growth. The ultimate goal is to increase farmers' profits and promote sustainable agriculture. Capacity can be further developed by integrating features that help farmers use sustainable technologies in specific climate models

Keywords: *Deep learning, Agriculture, Web Scraping, Machine learning, Weather data analysis, Crop yield prediction.*

1. Introduction

The nation’s development has been significantly influenced by agriculture for decades, especially in ensuring food production. Maximizing the limited amount of arable land is a critical challenge in today's urbanized world. Since there is a shortage of usable land, it's crucial to research innovative farming techniques and maximize output using the already available resources. Fortunately, with the advancement of technology and scientific data, agriculture can become more efficient. Machine learning can improve the accuracy and sustainability of agriculture when farmers have good knowledge of planting and selecting crops. The biggest challenge among farmers in India is the lack of knowledge about good soil and proper development strategies. In contrast, agriculture in other countries relies on machinery and uses micromanagement techniques to measure soil, which improves crop yields. Combining the differences between Indian farmers' practices and soil and crop management techniques can increase productivity.

Machine learning and data mining has proven helpful in many areas to extract valuable and useful information for stakeholders. Machines are good at running the world in daily activities, forecasting the weather, predicting the coming of storms, and saving lives. Similar benefits can be achieved by applying artificial intelligence to the lives of farmers. Intelligent machines can help farmers understand the importance of crop yield predictions, improve their knowledge of soil quality, protect against certain weather constraints, and use fresh and soil-friendly technology to increase productivity. Although current methods often rely on expensive equipment and do not give accurate results, recent developments have focused on data analysis, developing and applying data mining techniques to predict crop yields. Some systems even recommend ranking crops by profit and market value. By leveraging these advances, agriculture can overcome limitations and increase efficiency and productivity.

This research examines the integration of weather and rainfall data in the context of the above trends and progress. We explore how climate data, advanced predictive models, and data mining techniques can support decision-making, sustainable agriculture, and improved crops. Using climate theory, farmers can improve resource allocation, use appropriate irrigation strategies, and adapt their farming strategies to weather conditions. This approach can lead to more robust and productive agriculture in changing grassland environments.

2. Related Work

An essential part of the Indian economy is the agricultural industry. However, there needs to be more technology adoption and more utilization of advancements in agriculture. This has led to a significant decrease in the contribution of farming to the Gross Value Added. Considering the importance of food production and farmers’ crucial role, this study aims to examine agriculture's current state and identify improvement areas. This research seeks to enhance the agricultural sector and its contribution to the overall economy by analyzing the existing challenges and exploring potential solutions.

[1] *Assistant Professor, Department of Information Science& Eng., Nitte University (Deemed to be),NMAM Institute of Technology, Nitte, India ORCID ID : 0000-0001-5636-5517*

[2] *Professor, Department of Information Science& Eng., Nitte University (Deemed to be),NMAM Institute of Technology, Nitte, India ORCID ID : 0000-0001-6394-6774*

** Corresponding Author Email:devidasbhat@nitte.edu.in*

In article [1], the authors focus on applying data mining methods in weather prediction. The authors explore using machine learning algorithms to analyze weather data and make accurate predictions. By leveraging data mining techniques, the study aims to improve the accuracy of weather forecasts, which can have significant implications for various sectors, including agriculture, transportation, and disaster management. The paper discusses the methodology employed and the algorithms utilized and presents the results and findings of the study. The research advances weather prediction techniques and highlights the potential benefits of data mining in weather forecasting systems.

Article [2] explores the use of machine learning algorithms in predicting crop yield based on weather data. The authors present a model that combines multiple machine-learning techniques and uses weather patterns as input. This study demonstrates the effectiveness of the model in accurately predicting crop yield. The results show the potential of machine learning to improve agriculture and increase productivity by giving farmers better insight into the growth patterns of air.

In article [3], the authors review the use of machine learning algorithms to predict crop yields using weather data. They analyze the relationship between weather variables and crop yields and develop machine-learning models that predict future crops based on historical data. The results demonstrate the effectiveness of the model in accurately predicting crop yield. This study highlights the importance of combining machine learning techniques with weather data to help farmers make informed decisions and improve their farming practices.

Article [4] presents a study on analyzing machine learning techniques that can be used to analyze soil data. It discusses the traditional data collection approach and proposes a model for using them for analysis. Work mainly focuses on historical data collected and doesn't suggest any methodologies to handle real-time data. The paper uses data fetched using a traditional approach and analyzes agrarian data to enhance agricultural decision-making processes. The authors discuss the architecture and components of the conventional data processing model, emphasizing its potential applications and benefits in agriculture.

The article [5] overviews web scraping and its diverse applications. They discuss web scraping techniques, including data extraction techniques and tools commonly used in the field. This article highlights the importance of web scraping in various areas, such as data analysis, market research, sentiment analysis, and content aggregation. The authors explore the ethical and legal issues associated with web scraping, emphasizing the need for responsible and ethical practices. This review is helpful for researchers and practitioners interested in harnessing the power of web scraping for information applications.

In article [6], the authors aim to identify the most common web scraping techniques used today and provide best practices for ethical and legal compliance in web scraping. This article compares the performance of various tools, such as requests library and Selenium, on modern websites. The findings also demonstrate why Selenium is the most popular web scraping tool, as most contemporary websites now generate material dynamically when this may not have been the case a few years ago. It also highlights the security level currently built into the foundation of contemporary websites.

Web scraping is collecting the data generated from the web page and converting it into a usable format. Web scraping is essential for training learning models because it allows the collection of large data sets, and the creation of learning models from different cultures and real-world data, improving their reality and performance.

Machine learning is a branch of computer science that includes algorithms that learn and improve from data without being specially designed. Machine learning, an essential aspect of artificial intelligence, enables computers to interpret and analyze input data to gain meaningful insights from pre-written instructions. Machine learning focuses on the development of self-learning algorithms, making it easier to learn from data. This learning process involves extracting patterns and experiences from data to identify facts and predict the future. With minimal human intervention, machine learning allows the computer to adapt and become autonomous as it learns. Machine learning is considered an essential part of computer science, has applications in many universities worldwide, and is considered the future of engineering and wisdom.

3. Proposed System

This work aims to develop and apply a rule-based model for reliable crop production prediction based on historical data collection. Real-time web scraping is used to gather data, which is then compared to data previously gathered to do this. Crop yields are forecasted by the system using various machine learning techniques, which also analyze local climate and weather data that can be used to predict crop yields. Figure 1 depicts the system's general architecture.

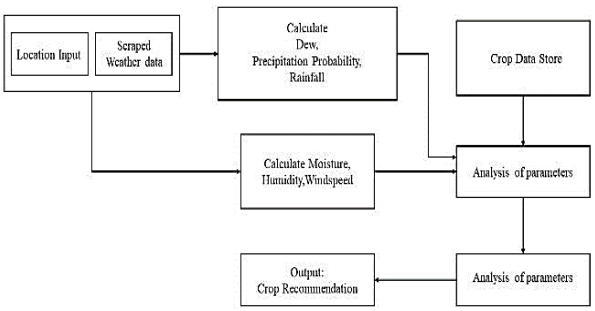

Fig 1. Figure depicting the model framework for the proposed system.

The proposed model framework contains five modules, as shown in Figure 1. In the context of this framework, web scraping is used to gather real-time weather patterns from multiple reliable sources. Extracting relevant information from various sources on the web is a crucial step. To extract data from numerous sources, the first step is to identify the websites that contain the desired information. This could include weather data providers, agricultural websites, or government portals offering relevant datasets. Once the target websites are determined, the web scraping process can be initiated.

The proposed model provides a holistic approach to leveraging weather data for maximizing crop yield. It empowers farmers with valuable insights and recommendations, enabling them to make informed decisions about crop selection and cultivation practices. By incorporating real-time weather data and historic trends, the model assists farmers in optimizing their agricultural practices and improving overall crop productivity.

The framework aggregates results using directed learning methods like Multiple Linear Regression and Back Propagation Networks from various machine learning methods to provide the most accurate results to end users. Additionally, the system offers informed recommendations regarding suitable nutrition for different crops. Comprehensive data on various crops throughout different seasons is collected. The framework analyzes weather data to assess crop yield potential and provides recommendations on suitable practices or interventions for optimizing crop yield if needed. Based on the generated report, the model recommends suitable crops for cultivation. The crop recommendation module utilizes the analyzed weather data and the crop preferences to suggest the most appropriate crops likely to thrive in weather conditions. The recommendations consider the specific temperature and rainfall requirements of each crop, ensuring better crop selection for farmers.

4. Result and Discussion

The training data used for our model is derived from data scraping techniques, particularly the collection of weather and crop data from various online sources. We collected location-based details to ensure the model's accuracy for a particular region. The government provided crop information used. The system was run on Dual Intel Xeon ES-2609V4 8C 2.0Ghz with 128GB of RAM. A hundred epochs and varying batch sizes were used during the training.

Using programming libraries like Beautiful Soup and Scrapy combined, the Python script can be written to visit each website, access the required web pages, and extract the desired data. Beautiful Soup assists in parsing the HTML structure of the web page and extracting specific HTML elements, such as tables, divisions, or classes, containing the relevant information. Scrapy, as a web scraping framework, provides additional functionality to handle crawling multiple pages, taking requests and responses, managing session states, and navigating website links. The combination of Beautiful Soup and Scrapy offers a powerful and flexible approach to web scraping for extracting weather data.

	Temp	Dew	Uvindex	Pressure	SolarEnergy	Precip	WindSpeed	WindGust	Humidity
N	7305	7305	4018	4148	4018	7305	7305	31	7305
Mean	27.1	22.5	7.56	1009	18.8	10.4	19.0	47.1	77.6
Median	27.0	23.3	8.00	1009	19.8	0	18.4	46.4	77.1
Standard Deviation	1.52	2.21	1.85	1.99	4.73	25.6	5.17	7.82	10.1
Minimum	23.1	12.2	2.00	997	4.70	0.00	0.000	37.1	47.5
Maximum	32.9	26.0	10.0	1017	27.7	389	98.5	66.6	99.2

Fig 2. Summary Statistics of Weather Parameters.

The management of pagination, login/authentication techniques, handling dynamic information loaded through JavaScript, and maintaining access from the website are only a few of the problems that need to be managed during scraping. Strategies like retaining cookies, employing proxies, or using delays in requests can be used to overcome these difficulties. Additionally, it's critical to ensure that online data scraping is private and compliant with the law by being aware of the ethical and legal ramifications.

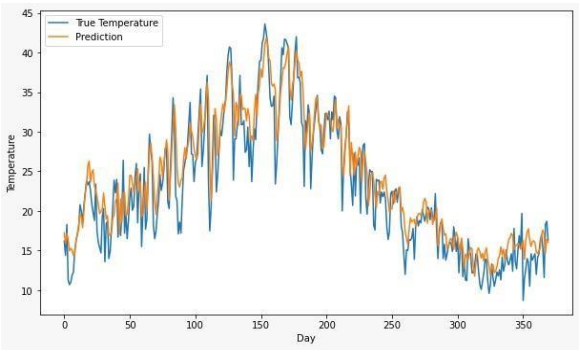

Fig 3. Graph with actual and predicted temperature Parameters.

After fetching the data through the web scraping technique, the mean, and the standard deviation is calculated to analyze the nature of the data, which is depicted in Figure 2. To improve the data, neural networks are used with different sets of neurons. Figure 3. Shows the prediction of actual and predicted values of temperature using the Bi-LSTM technique.

Model	MAE	MSE
Univariate 3-layer DNN 10+SGD	2.008685	6.347851
Univariate 3-layer DNN 16+ Adam	1.093827	1.482473
Univariate 3-layer DNN 16+ Adam+ Normalization	0.435637	0.341555
Univariate 3-layer DNN 16+ Adam+Scaling	1.050605	1.700835
Univariate Conv1d,2×LSTM,3-Layer DNN 16+Adam	1.801992	5.175402
Univariate Conv1d,2×LSTM,3 Layer DNN 16+SGD	0.590229	0.532987

Fig 4. MAE and MSE using DNN.

A Univariate 3-layer Deep Neural Network (DNN) with 10+ neurons using Stochastic Gradient Descent (SGD) to find the Mean Absolute Error (MAE) and Mean Squared Error (MSE). In this model, a single input is considered at a time, having three hidden layers with at least ten neurons in each layer. SGD, an optimization algorithm, updates the model's parameters. Adam is an adaptive learning rate optimization algorithm. Here 16+ neurons are used in each hidden layer. Figure 4 shows that the MAE and MSE have reduced in Univariate three layer Deep Neural Network (DNN) with 16+ neurons using Adam's algorithm and Normalization.

The data that has been gathered must be processed and combined after being scraped from various sources utilizing web scraping techniques. To help with this data integration process, Python offers several modules and tools. An analysis phase comes next. The model includes an analysis module that processes the weather parameters to derive meaningful insights. This module calculates temperature variations, humidity, wind speed, precipitation probability, and rainfall patterns in two separate stages to assess their impact on crop growth.

The analysis of weather parameters is supplemented by additional input in the form of crop data. This auxiliary input helps understand different crops' specific requirements and preferences concerning temperature and rainfall. Combining the processed weather and crop data, the model generates a comprehensive report providing valuable insights into the optimal crop conditions.

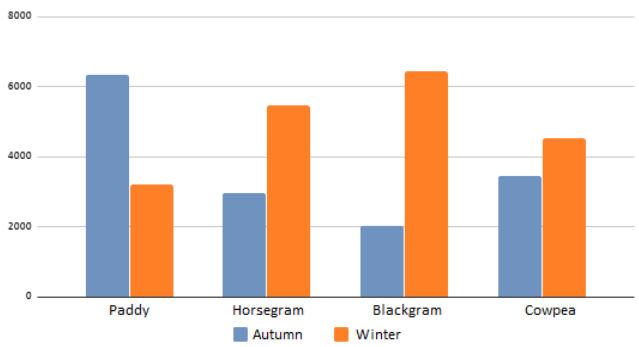

Fig 5. Figure showing the kinds of plants cultivated during two distinct seasons.

We leveraged machine learning techniques in the Python programming language. The back-end engine utilized was TensorFlow, with Keras as a high-level API for building and training neural networks. Additionally, we incorporated the Scikit-learn library to support various machine-learning tasks. The Scikit Learn library is supported by the back-end engines Keras and TensorFlow.

Our work analyzes machine learning techniques like SVM, Decision Trees, Random Forest, and Backpropagation Regression. Their accuracy is presented on a graph to show how accurate they are. The findings show that, when compared to the other approaches, the regression method produces excellent accuracy for the sample set.

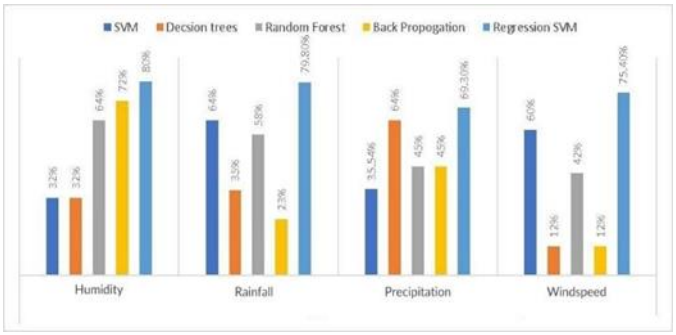

Fig 6. Figure showing various machine learning approaches' accuracy

Figure 5 demonstrates that Regressive SVM has a higher accuracy than other models, which leads us to this conclusion.

5. Conclusion

Farmers in India rely on traditional cultivation methods that hinder their ability to compete with modern knowledge-based agriculture. However, machine learning techniques have shown great potential in improving data analysis and prediction in many areas. This paper fetches real-time data

using a web scraping technique. It proposes integrating machine learning analysis into the Indian agricultural setup, specifically leveraging weather data to accurately forecast environmental conditions, including temperature, humidity, wind speed, dew, precipitation, etc. The framework selects the algorithm that produces the most accuracy based on careful analysis and comparison of various algorithms.

Furthermore, we acknowledge the significance of preserving accurate weather data and promoting sustainable farming practices based on climatic conditions. Farmers can make informed decisions about planting, planting time, pest and disease management, and other important factors affecting crop growth and development using precise weather data. By providing farmers with valuable and accurate information, our foundation strives to improve farmers' profits and health while promoting stability in agriculture in India.

We can further improvise our model by incorporating features that assist farmers in adopting sustainable techniques tailored to specific weather patterns.

6. References and Footnotes

6.1. References

References need not be cited in text. When they are, they appear on the line, in square brackets, inside the punctuation. Multiple references are each numbered with separate brackets. When citing a section in a book, please give the relevant page numbers. In text, refer simply to the reference number. Do not use “Ref.” or “reference” except at the beginning of a sentence: “Reference [3] shows ... .” Please do not use automatic endnotes in Word, rather, type the reference list at the end of the paper using the “References” style.

Reference numbers are set flush left and form a column of their own, hanging out beyond the body of the reference. The reference numbers are on the line, enclosed in square brackets. In all references, the given name of the author or editor is abbreviated to the initial only and precedes the last name. Use them all; use *et al*. only if names are not given. Use commas around Jr., Sr., and III in names. Abbreviate conference titles. When citing IEEE transactions, provide the issue number, page range, volume number, year, and/or month if available. When referencing a patent, provide the day and the month of issue, or application. References may not include all information; please obtain and include relevant information. Do not combine references. There must be only one reference with each number. If there is a URL included with the print reference, it can be included at the end of the reference.

Other than books, capitalize only the first word in a paper title, except for proper nouns and element symbols. For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation See the end of this document for formats and examples of common references.

Declaration

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Devidas, Balasubramani, “Weather Prediction using Data Mining Techniques,” International Conference in Recent Innovations in Electrical, Electronics & Communication Engineering, (ICRIEECE) 2018.

[2] V. Gupta, P. Kumar, and M. Singh, “Crop Yield Prediction Using Machine Learning Algorithms and Weather Data,” IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018.

[3] A. Yadav, N. Kumar, and M. Sharma, “Crop Yield Prediction using Machine Learning Algorithms and Weather Data,” International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), 2019.

[4] Devidas, Balasubramani.R, (2020). “A predictive modeling framework for enhancing crop yield using an automated data-driven approach,” Volume 9, 6900-6904. doi:10.30534/ijatcse/2020/05952020.1007/978-981-15-0663-5_11,IJATCSE.

[5] V. Singrodia, A. Mitra, and S. Paul, “A Review on Web Scraping and its Applications,” International Conference on Computer Communication and Informatics (ICCCI), 2019.

[6] Ajay Sudhir Bale, S. Kamalesh, Naveen Ghorpade, Rohith R, Rohith S, and Rohan B S, "Web Scraping Approaches and their Performance on Modern Websites," in Proceedings of the IEEE International Conference on Computer Science and Engineering (IC-CSE), Bengaluru, India, Year (to be filled), pp. 956-959. doi: 10.1109/ICESC54411.2022.9885689

[7] P. Vinciya, Dr. A. Valarmathi, “Analysis for Next Generation High Tech Farming in Data Mining,” IJARCSSE, vol. 6, Issue 5, 2016.

[8] Santanu Koley, Shivnath Ghosh, “Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks,” IJRITCC, vol. 2, Issue 2,292-297, 2014.

[9] Zhihao Hong, Z. K. Iyer, Kalbarczyk, R “A Data-Driven Approach to Soil Moisture Collection and Prediction.”

[10] S. K. A. S, S. M. S, and A. Bhandary, “Lemon Maturity Estimator: An Approach Using Color Image Processing Techniques,” 2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimisation Techniques (ICEECCOT), Msyuru, India, 2018, pp. 1213-1218, doi: 10.1109/ICEECCOT43722.2018.9001657.

[11] Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). “Analysis of agriculture data using data mining techniques: application of big data. Journal of Big Data”, 4(1), 20.

[12] Gholizadeh, A., Carmon, N., Klement, A., Ben-Dor, E., & Borůvka, L. (2017). “Agricultural soil spectral response and properties assessment: effects of measurement protocol and data mining technique,” Remote Sensing, 9(10), 1078.

[13] Sandeep Kini M, Devidas, Smitha N Pai, Sucheta Kolekar, Vasudeva Pai, Balasubramani R. "Use of Machine Learning and Random OverSampling in Stroke Prediction", 2022 International Conference on Artificial Intelligence and Data Engineering (AIDE), 2022

[14] Kumbhkar, M., Shukla, P., Singh, Y., Sangia, R.A., Dhabliya, D. Dimensional Reduction Method based on Big Data Techniques for Large Scale Data (2023) 2023 IEEE International Conference on Integrated Circuits and Communication Systems, ICICACS 2023,