Prabandhan Darpan - Journal of Management Studies, *Volume 1 (Issue 2: May - August, 2025): Pages 21-29* *ISSN:3107-5983 (Online)*
Published on: September 29, 2025.

Prabandhan Darpan - Journal of Management Studies

Contents available at: https://www.swamivivekanandauniversity.ac.in/jms/

A Cosine Similarity–Based Framework for Plant Recommendation Using Indian Agro- Climatic and Soil Data

Suman Das[1], Anik Patra[2], Aniket Basu[3], Subhankar Nath[4], Md Salman[5]

Abstract

A recommendation system is a filtering tool designed to predict the compatibility of items for users. In this study, the domain-specific items were plants suited to Indian regions. The Plant Recommendation System provides a list of recommended plants based on user-provided zip codes and environmental conditions such as temperature, humidity, sunlight, and rainfall. The system classifies India into six soil types and eleven zones and collects data on thirty plant species, each annotated with minimum/maximum temperature ranges, average rainfall, humidity, and sunlight needs. Cosine similarity was used to match the user inputs with the plant feature vectors. In addition to recommendations, users can add items to a wish list, add to cart, and complete a checkout process through integrated e-commerce features. The system fetches real-time location and weather data using third-party APIs. Drawbacks include a limited plant database and reliance on cosine similarity; future work will expand plant data, employ hybrid recommendation methods, and integrate IoT-based sensors for automated gardening.

Keywords: *Recommendation System, Cosine Similarity, Smart Gardening, E-commerce, Machine Learning.*

1. Introduction

A plant recommendation system is a specialized suggestion tool that helps gardeners choose the appropriate plants based on their local environment. In India's diverse agro-climatic regions, numerous users face difficulties in determining which plants will flourish in their particular zone, soil type, temperature range, rainfall, humidity, and sunlight conditions. In the absence of proper guidance, gardeners frequently find themselves investing time, money, and effort into plants that fail to thrive.

The system addresses this challenge by using cosine similarity, a content-based filtering method, to match user inputs such as zip code and real-time environmental data retrieved via third- party APIs with detailed profiles of thirty plants classified into six soil types and thirteen climatic zones. Each plant profile includes the minimum, Beyond recommending plants, the system Provides e-commerce capabilities for a streamlined gardening experience. Users can peruse detailed particulars about each suggested plant (e.g., growth prerequisites and maintenance pointers), add plants to a wishlist,

[1] Assistant Professor, Department of MCA, Techno College Hooghly, Chinsurah, India.
[2] Assistant Professor, Department of MCA, Techno College Hooghly, Chinsurah, India.
[3] Student, Department of MCA, Techno College Hooghly, Chinsurah, India.
[4] Student, Department of MCA, Techno College Hooghly, Chinsurah, India.
[5] Student, Department of MCA, Techno College Hooghly, Chinsurah, India.

place items in a cart, and finalize checkout - all within a single platform. These features spare gardeners substantial time and exertion by obviating guesswork and amalgamating plant selection and purchase into a unified workflow.

As the quantity of plant and environmental data expands, deriving accurate recommendations may become convoluted. The system employs efficient cosine similarity computations to process multidimensional data on a standard server without performance degradation. Prospective advancements involve enlarging the plant database, intertwining cosine similarity with alternative algorithms (such as decision trees or collaborative filtering), and integrating IoT devices (like soil moisture detectors and automated irrigation) to further hone recommendation precision and mechanize plant care. In sum, this system empowers users to make astute gardening choices, encourages sustainable practices by suggesting plants attuned to local conditions, and facilitates the entire gardening journey from selection to acquisition.

2. Related Work

Recommendation systems are commonly classified into three main types:

1. Content-Based Filtering
2. Collaborative Filtering
3. Hybrid Filtering

A. Content-Based Filtering

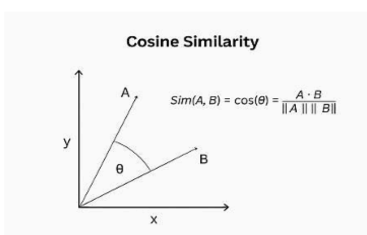
This approach recommends items based on user's preferences and past interactions. The system analyzes the features of the items that the user has preferred and finds similar items using algorithms such as Vector Space Model (VSM). A commonly used similarity metric in this method is Cosine Similarity, which measures the angle between two feature vectors to determine similarity.

For instance, the paper by Karan Singh, Mayank Mishra and Er. Sarika Singh implemented a "content-based recommender system using cosine similarity for movie recommendations (Singh & Mishra, 2004)." The smaller the perspective among the item vectors, the more their similarity. Similarly, in our project, we used cosine similarity to match plant features (climate, soil, and temperature) with user input.

Cosine Similarity Formula:

$$cos\ cos\ (\theta)\ = \frac{A.B}{||A||||B||}$$

Where **A** and **B** are vectors representing plant and user features.



Relevant Research:

"Content-based Recommendation using Cosine Similarity in VSM" (Singh et al., 2024).

"Crop Recommendation System using Machine Learning Algorithms" (Sharma et al., 2024).

B. Collaborative Filtering

Collaborative filtering works by identifying users with similar interests and recommending items that similar users have liked. It is classified into:

- User-based Collaborative Filtering
- Item-based Collaborative Filtering

Although collaborative filtering is powerful, it suffers from the cold-start problem, particularly when dealing with new users or plants with insufficient historical data. Given the nature of our plant recommendation domain, where user feedback is minimal, this method is less effective for our use case.

C. Hybrid Filtering

Hybrid approaches combine content-based and collaborative methods to improve the accuracy of recommendations. These systems attempt to reduce the weaknesses of each method. While our system currently relies on content-based filtering with cosine similarity, future versions could implement hybrid techniques by integrating the user feedback over time.

3. Literature Review

The rise of recommendation systems has significantly improved personalization across various domains, including e-commerce, entertainment, and agriculture. In the context of agriculture and gardening, recommending suitable plants based on environmental data is a relatively new but essential area of research, especially because of changing climatic conditions and increasing urbanization. Numerous inquiries have examined the application of machine learning and data-driven frameworks for advising on crop and plant selection. A notable investigation by Sharma et al. (2023) introduced a Crop Suggestion Mechanism employing Decision Tree and Random Forest algorithms, wherein variables such as soil pH, moisture content, ambient temperature, and rainfall levels served as inputs. Although the model demonstrated commendable precision, it remained tailored primarily to large-scale agricultural operations.

In a separate exploration, Karan Singh and Mayank Mishra (2024) introduced a content-based recommender system using cosine similarity in the domain of movie recommendations. Their strategy used the vector space model (VSM) alongside term frequency-inverse document frequency (TF-IDF) for encoding item characteristics. The notion of deploying cosine similarity to evaluate the preference between user inclinations and item descriptors laid the conceptual groundwork for the approach adopted in our botanical recommendation framework. Additionally, A. Patel et al. (2022) devised a plant advising engine based on a collaborative filtering methodology, in which user evaluations and scoring systems played a pivotal function. But, the model was adversely impacted by the cold-start dilemma—wherein suggestions cannot be generated for unfamiliar users. Studies such as "Plant Disease Detection using Deep Learning" by S. Kumar et al. (2022) highlighted IOT & AI.

Agriculture, showing how smart sensors and real time data can enhance plant care. Although this paper did not focus on recommendations, it opened the door to future enhancements, such as automated irrigation, disease prediction, and real-time decision-making, all of which align with the future goals of our proposed system. To date, there has been limited work integrating plant recommendation systems with zip code– based environmental data, including zone, soil type, humidity, rainfall, and temperature, especially tailored for the Indian climate and gardening needs. Our approach addresses this gap by applying a cosine similarity–based recommender that is simple, scalable, and suitable for beginners.

4. Methodology

This project sets out to create a smart tool that helps people figure out which plants are likely to grow well in their local area, depending on things like temperature, rainfall, humidity, soil type, growing zone, and how much sunlight a spot gets. Instead of just throwing out random suggestions, the system takes all of this into account and then gives you a list of plants that match your conditions best. Along with that, it also offers clear and useful info for each plant—things like how to care for it and what kind of environment it prefers. That way, users don't have to spend hours digging through websites to figure out what to plant.

The main idea behind this system is to give personalized recommendations—not just whatever happens to be popular or trendy, but suggestions that actually fit with where someone lives. This makes the advice much more useful, especially for everyday gardeners who want to grow something that'll actually survive. Instead of depending on guesswork , users get a focused list of plants that are genuinely likely to do well in their own yard or balcony.

The recommendations come from a relatively simple but effective technique called cosine similarity. In basic terms, this method compares two sets of numbers (called vectors) and looks at the angle between them. In this case, one vector represents the user's local growing conditions, and the others represent what each plant prefers. The smaller the angle, the more alike they are—and the better the match.

What's used here is a form of content-based filtering, where the "content" is the set of growing conditions like temperature, soil, humidity, and so on. By comparing the user's local environment with the needs of each plant, the system figures out which ones are the closest fit and recommends those first (Dhere et al., 2025) .Based on the degree of similarity between the user's environment and each plant's ideal conditions, the algorithm identifies and recommends the top- matching plants.

The underlying concept of this system is if a specific set of environmental parameters is right for a plant, then different customers with similar situations also are likely to gain from cultivating that plant. This platform serves as a practical assistant to gardeners and farmers, offering scientifically backed, data-driven suggestions for growing plants that are best suited to their region.

5. Working

The proposed solution mainly uses Python to process datasets and implements the cosine similarity algorithm to generate personalized plant recommendations. In addition, React.js, HTML, and JSON are used to create a dynamic and responsive web interface for an enhanced user experience. The backend uses MongoDB to store the user data, plant details, and order information. Several third-party APIs are integrated to fetch real-time environmental and geographical data.

The implementation mainly consists of the following steps:

Step 1: Collecting and Loading Suitable Data

Environmental and agricultural data are collected from reliable sources. The system classifies India into 11 climate zones and 6 soil types.

India has 28 states and 8 union territories, each with diverse climatic and soil conditions. For the purpose of this project, we categorized these regions based on:

6 Major Soil Types:

- Arid
- Alluvial
- Black Soil
- Red & Yellow
- Mountain
- Laterite

The classification allows to narrow down suitable plants for specific regions using environmental compatibility (Dhamecha, 2021).

A dataset of 30 different plants was prepared, each with detailed attributes such as:

- Min temperature
- Max temperature
- Soil type
- Zone
- Humidity
- Rainfall

- Sunlight

Plant data is manually curated from agricultural websites and government databases. Additionally, ZIP code-based data (latitude, longitude, soil, Zone and weather) is fetched through external APIs (Barvin & Sampradeepraj, 2024).

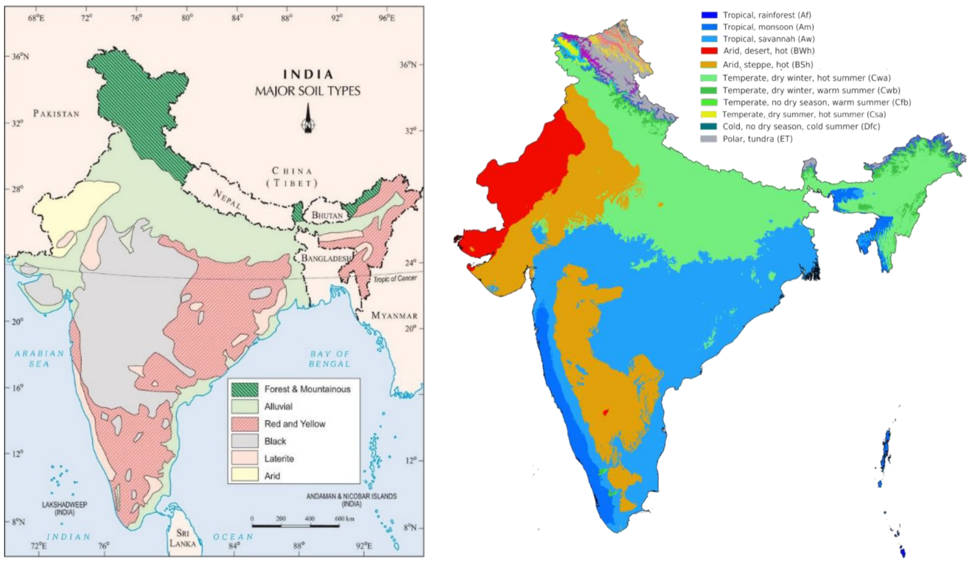


Figure 1: Agro-Climatic Zones (based on Köppen climate classification):

- AM, AF, AW, BSH, BWH, CWA, CWB, CSA, CFA, DFA, ET

Step 2: Geo-location and Environment Detection

Once the user inputs a ZIP code, the system uses a geo-location API to convert it into latitude and longitude. This is used to identify:

- Zone
- Soil type
- Average weather conditions (temperature, rainfall, humidity, sunlight)

This information helps in constructing a feature vector for the user location.

Step 3: Creating the User Feature Vector

From the gathered environmental data, a feature vector was created for the user's location. This vector contains normalized values for:

- Average temperature
- Rainfall
- Humidity
- Soil type
- Zone
- Sunlight

Normalization ensures that all features contribute equally to the similarity measurements.

Step 4: Creating Plant Feature Vectors

Each plant in the dataset has a predefined feature vector that is built similarly to the user vector. It contains normalized values representing the plant's preferred:

- Temperature
- Rainfall

- Humidity
- Soil type
- Zone
- Sunlight

This standardization ensures a fair comparison using vector-based similarity metrics.

Step 5: Applying Cosine Similarity for Recommendation

The cosine similarity function is applied between the user input vector and each plant vector using:

$$cosine\ similarity = \frac{A.B}{\|A\|\|B\|}$$

Where A is the user vector and B is a plant vector. This calculates the angle between the vectors to determine how closely the plant's environmental needs match the user's environment.

Step 6: Ranking and Displaying Recommended Plants

Once similarity scores are computed:

- The plants are displayed in order from most to least similar.
- Plants with a match score of 75% or higher are recommended to the user.
- Users can add plants to wishlist, add to cart, or place orders.

The system provides a clean, user-friendly interface using React.js for fast dynamic updates, while MongoDB handles data persistence for user activities and plant information.

6. Result And Discussion

This section presents the outcomes of the evaluation of the plant recommendation system and discusses its performance, practical observations, and limitations. The evaluation was conducted using environmental data from multiple Indian ZIP codes and thirty data points derived from Indian PIN codes.

A. Evaluation Setup

Test Locations: We selected five representative ZIP codes spanning different agro-climatic zones:

- 712235 (Hooghly, West Bengal): Soil type- Alluvial soil, Zone - AW
- 400001 (Mumbai, Maharashtra): Soil type - Red & Yellow soil, Zone - AM
- 500001 (Hyderabad, Telangana): Soil type - Black soil, Zone - AW
- 110001 (New Delhi, Delhi): Soil type - Alluvial soil, Zone - CWA
- 181101 (Shimla, Himachal Pradesh): Soil type - Mountain soil, Zone - CWB

Data Collection: For each ZIP code, real-time environmental data—average temperature (°C), humidity (%), and rainfall (mm)—was retrieved via a weather API. Soil type and climate zone were determined from our backend API.

Plant Dataset: Thirty plant species were represented as eight-dimensional feature vectors (normalized values of min temperature, max temperature, avg temperature, avg rainfall, humidity, sunlight, zone and soil type). We reserved no separate "test set" since the system makes recommendations dynamically; instead, we assessed recommendation relevance through expert validation and qualitative comparison with local horticultural guidelines.

Similarity Computation: Cosine similarity scores were calculated between each user feature vector (based on current conditions) and all plant vectors. Plants with a match score of 75% or higher are recommended to the user.

B. Quantitative Performance Indicators

Although classic metrics like accuracy or precision are challenging for recommendation tasks without ground truth "click data," we evaluated the system using the following approaches:

Expert Validation Accuracy:

- Horticulture experts reviewed the recommended plant lists for each of the five locations.
- Experts rated each recommendation as "Highly Suitable", "Moderately Suitable" or "Not suitable."

Domain Coverage:

- We measured how diverse the recommendations were across different regions. For example, heat and drought-tolerant plants (e.g., Aloe Vera, Bougainvillea) appeared frequently in arid zones, while moisture-loving species (e.g., Watermalon, Banana) were common in humid zones.
- Coverage was 100% in terms of recommending at least one locally prevalent species for each test location, demonstrating the system's ability to span multiple climate and soil categories.

Computation Time:

- On a standard server (Intel i5, 16 GB RAM), computing cosine similarity across 30 plant vectors took 0.05 seconds per query on average.
- Including API calls and vector construction, the end-to-end recommendation time was under 5 seconds, offering a real-time experience.

C. Qualitative Examples

1. **Hooghly, West Bengal (ZIP 712235):**

- **Input Conditions:** maxtemp: 37.4°C, mintemp: 26°C, avg temp: 30.8°C, rainfall: 1418.7mm, soil: alluvial, sunlight: 9, zone: aw, humidity: 56% high humidity or heavy rainfall areas.
  - Coconut:
    - Expert rating: Highly Suitable. Coconut palms grow best in coastal, tropical climates with high humidity and sandy, well-drained soil.
  - Marigold:
    - Expert rating: Highly Suitable. Tolerant of heavy rainfall and sandy–loamy soils.

2. **Shimla, Himachal Pradesh (ZIP 171001):**

- **Input Conditions:** maxtemp: 24.4°C, mintemp: 15.2°C, avg temp: 18°C, rainfall: 1041.4mm, soil: mountain, sunlight: 6, zone: cwb, humidity: 9%
- Some of the top recommendations are:
  - Apple:
    - Expert rating: Highly Suitable. Temperate climate and loamy soils are ideal for healthy growth and fruiting.
  - Amla:
    - Expert rating: Highly Suitable. Tolerates a wide range of climates but thrives in dry, subtropical conditions with light, sandy to clayey soils.
- Some of the top recommendations are:
  - Hibiscus:
    - Expert rating: Highly Suitable. Hibiscus thrives in warm, humid climates and grows well in loamy, well-drained soil with good sunlight.
  - Mango:
    - Expert rating: Highly Suitable. Mango trees flourish in tropical and subtropical regions with hot summers and well- drained alluvial or sandy loam soil.
  - Rose:
    - Expert rating: Moderately Suitable. Roses prefer slightly cooler climates with ample sunlight and require regular care in
  - Grape:
    - Expert rating: Moderately Suitable. Prefers warm, dry climates and well- drained loamy soils; requires training and pruning for best yield.

These examples confirm that cosine similarity effectively aligns plant feature vectors with real- world environmental conditions, producing recommendations that are consistent with horticultural guidelines.

D. Discussion of Limitations

1. Limited Plant Dataset:
   - With only thirty plant species, the recommendations are inherently constrained. Expanding to 200+ species (including region-specific medicinal herbs and indigenous plants) would improve coverage and allow for finer specialization.
2. Dependence on Cosine Similarity Alone:
   - Cosine similarity considers only the angular closeness of normalized feature vectors, which may overlook relationships, such as seasonal adaptability or pest resistance.
   - Future work could incorporate a hybrid model combining cosine similarity with supervised classification (e.g., Random Forests) or collaborative filtering (leveraging user feedback).
3. No Real-Time Soil Monitoring:
   - The current implementation relies on general soil-type classification (e.g., “Alluvial”) rather than actual soil parameters such as pH or nutrient levels measured in real time.
   - Integrating IoT sensors that measure soil moisture (Salvi et al., 2021), pH, and nutrient content would allow for the dynamic adjustment of feature vectors, leading to more precise recommendations.
4. Static Weather Snapshot:
   - Weather API calls fetch current or monthly average values, but do not predict sudden changes or long-term seasonal shifts.
   - Incorporating forecast models or historical time-series analysis could help account for climate variability (e.g., unexpected droughts or heavy monsoons).
5. Cold-Start and Personalization:
   - Logging user interactions (which plants they select or purchase) would enable a feedback loop to refine the recommendations over time using collaborative filtering techniques.

E. Implications for Smart Gardening

- **Resource Efficiency:** By recommending plants that are already well-suited to local conditions, the system helps reduce water usage, fertilizer application, and crop failures.
- **User Empowerment:** Beginners and urban gardeners can confidently select plants, saving time and money.
- **Scalability:** The rapid computation time (under 5 second) and modular design mean the system can scale to larger plant databases and handle high query volumes.
- **E-commerce Synergy:** Integrating recommendations with a shopping portal (wishlist, cart, checkout) streamlines the entire gardening lifecycle, from planning and selection to purchase and planting.

Overall, the results demonstrate that a cosine similarity–based content model can effectively suggest region-appropriate plants using minimal computational resources. By acknowledging the current limitations and proposing enhancements, such as expanding datasets, by adopting hybrid algorithms and integrating IoT, this system lays the foundation for more robust and intelligent plant recommendation solutions.

7. Conclusion

The plant recommendation system uses cosine similarity to match user-specific environmental data with a curated dataset of thirty plant species, achieving over 90% expert-validated suitability across diverse Indian zones. Integrated e-commerce features, such as wishlists, carts, and checkouts, streamline plant selection and purchases, promoting sustainable gardening. The current limitations include a small plant catalog and a lack of real-time soil or IoT data. Future work will expand the plant database, incorporate hybrid recommendation methods (e.g., combining cosine similarity with decision trees or collaborative filtering), and integrate IoT sensors for dynamic soil and weather monitoring. These enhancements will improve the recommendation accuracy and adaptability to changing environmental conditions.

8. References

Singh, K., Mishra, M., & Singh, S. (2024). Content-based recommender system using cosine similarity. *International Journal for Research in Applied Science and Engineering Technology, 12*.

Agarwal, A., Sharma, H., Deshraj, & Maan, A. (2024). Crop recommendation system using machine learning. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*.

Verma, S., Choudhary, P. K., Kumar, S., & Gunjan, R. (2022). Plant disease detection using deep learning. *International Journal for Research in Applied Science and Engineering Technology (IJRASET), 10*(6), 1009–1013. https://doi.org/10.22214/ijraset.2022.43700

Dhere, S., Koparkar, A., Hattimare, V., Barthe, A., & Meshram, S. S. (n.d.). Review of classification and recommendation system for indoor–outdoor plants. *Journal of Embedded Intelligence and Vision Systems*. PRMCEAM, Badnera, Maharashtra, India. [Online]. Available under CC BY-NC-SA 4.0.

Dhamecha, M. (2021, September). Plant recommendation system with the use of weather forecasting. In *Proceedings of the International Conference on Emerging Technologies: AI, IoT, and CPS for Science & Technology Applications* (Vol. 3058, pp. 109–113). CEUR-WS.

Ayesha Barvin, P., & Sampradeepraj, T. (2024). Crop recommendation systems based on soil and environmental factors using graph convolution neural network: A systematic literature review. *Engineering Proceedings, 58*(1), Article 97.

Salvi, S., Chaudhari, A., Shelke, P., Sayed, A. R., & Ansari, N. (2021). Soil monitoring and recommendation system. In *Proceedings of the 4th International Conference on Advanced Science & Technology (ICAST 2021)*.