

**TITLE: PREDICTION OF REAL-TIME WEATHER WITH RECOMMENDATION
USING MACHINE LEARNING**

TABLE OF CONTENT

	PAGES
2.1 Weather Prediction	
2.1.1 Definition	3
2.1.2 Methods used in Weather Prediction	3-4
i. Climatology Method	
ii. Analog Methodology	
iii. Persistence and Trends Method	
iv. Numerical Weather Prediction Method	
2.1.3 Comparison between Methods used in Weather Prediction	5
2.1.4 Challenges in Prediction	6
i. High-Quality Data for Training	
ii. Model and Observational Data Compatibility	
iii. Observational Data Limitations	
iv. Computational Cost and Efficiency	
2.2 Real-Time Recommendation Systems in Weather Prediction	
2.2.1 Definition	7
2.2.2 Type of Recommendation Systems	7-11
i. Content-Based Filtering	
ii. Collaborative Filtering	
iii. Hybrid Filtering	
2.2.3 Challenges in Recommendation Systems	12-13
i. Cold-start	
ii. Scalability	
iii. Data sparsity	
iv. Diversity	
2.2.4 Comparison for Recommendation Systems	14
2.3 Supervised Machine Learning	
2.3.1 Definition	15
2.3.2 Types of Supervised Learning	15-20
i. Classification	
a) Naïve Bayes	
b) Random Forest	
c) Support Vector Machines (SVM)	
d) Decision Trees	
ii. Regression	

	a) Simple and multiple linear regression	
	b) Polynomial regression	
	c) Lasso regression	
2.3.3	Applications of Supervised Learning	20-21
	i. Fraud Detection	
	ii. Image, patterns and speech recognition	
	iii. NLP and sentiment analysis	
2.3.4	Comparison between Techniques in Supervised Learning	21
2.4	Similar Application	
2.4.1	Weather Prediction Using Futuristic Technologies: From Industry's Perspective	22-23
2.4.2	Deep Learning-Based Weather Prediction: A Survey	
2.4.3	Weather monitoring and forecasting system using IoT	
2.4.4	Development of flood forecasting and warning system using hybrid approach of ensemble and hydrological model for Dharoi Dam	
2.4.5	Comparison between systems	24
2.5	Summary	25
2.6	References	26-27

2.1 Weather Prediction

2.1.1 Definition

Weather forecasting is crucial in today's technologically advanced world, aiding outdoor programming, crop cultivation, and time management. Advancements in science and technology have led to more accurate predictions. Scientists use various techniques to analyse meteorological data, including temperature, precipitation, atmospheric pressure, sunshine, wind direction, clouds, humidity, and wind speed. Machine learning techniques like Decision Trees, Artificial Neural Networks, Naive Bayes Networks, Support Vector Machines, Random Forest, and Rule-based Techniques are commonly used (Patkar, 2022).

2.1.2 Methods used in Weather Prediction

According to (Chetan & Waghmare, 2022), there are various methods for predicting the weather. Weather prediction involves various strategies based on experience, available information, and forecast complexity.

i. Climatology Method

The climatology method is a method used in weather forecasting where meteorologists use historical data to calculate averages for a specific day. It is effective when weather patterns remain stable, but may not be suitable for predicting weather influenced by frequent external factors like climate change due to global warming. For example, Labor Day averages can be used to predict upcoming holidays (Chetan & Waghmare, 2022).

ii. Analog Method

The analog approach to weather prediction is challenging as it requires identifying a past day with weather similar to the current forecast. For example, a similar day occurred the previous month, with a warm day followed by a cold front, causing thunderstorms. However, slight differences between past and present weather can influence the

outcome, making the analog method less suitable (Chetan & Waghmare, 2022).

iii. Persistence and Trends Method

Because it is based on previous trends, the persistence and trends method require little to no talent to forecast the weather. In an ideal world, the environment changes slowly, yielding a forecast for tomorrow that is comparable to today, with a nod to the seasonal climate average. This strategy only requires you to be aware of current events and conditions, as well as knowledge of the region's weather averages (Chetan & Waghmare, 2022).

iv. Numerical Weather Prediction Method

Numerical weather prediction uses supercomputers and software models to predict weather conditions based on atmospheric elements like temperature, wind speed, and rainfall. Meteorologists analyse data to determine the forecast's accuracy, with errors occurring when equations are not precise. Numerical weather prediction is the most effective way of forecasting future meteorological conditions compared to other methods (Chetan & Waghmare, 2022).

2.1.3 Comparison between Methods used in Weather Prediction

Method	Description	Advantages	Disadvantages
Climatology Method	A simple method based on calculating averages of meteorological data over several years to forecast weather for a specific day.	<ul style="list-style-type: none">- Easy to implement- Useful when weather patterns are consistent	<ul style="list-style-type: none">- Not effective when external factors like climate change have a frequent influence
Analog Method	Involves identifying a past day with weather like the current forecast to make predictions.	<ul style="list-style-type: none">- Can provide insights based on historical patterns- Helps in understanding similar weather events	<ul style="list-style-type: none">- Difficult to find exact analogues- Relies heavily on historical data
Persistence and Trends Method	Relies on previous trends and slow environmental changes to forecast weather with minimal expertise.	<ul style="list-style-type: none">- Requires minimal skill to forecast weather- Considers current conditions and seasonal averages	<ul style="list-style-type: none">- Limited accuracy in rapidly changing weather conditions- May not account for sudden shifts in weather patterns
Numerical Weather Prediction Method	Uses computers and forecasting models to predict weather based on various atmospheric elements like temperature, wind speed, and pressure systems.	<ul style="list-style-type: none">- Most effective method for forecasting future meteorological conditions- Utilizes advanced technology	<ul style="list-style-type: none">- Errors can occur due to imprecise equations- Relies on the accuracy of computer algorithms

Table 2.1 Comparison between Methods used in Weather Prediction

Table 2.1 compares four weather prediction methods: Climatology, Analog, Persistence and Trends, and Numerical Weather Prediction. Each method has advantages and disadvantages. Climatology is simple for stable weather patterns, Analog relies on historical data but struggles with exact analogues, Persistence and Trends forecasts recent trends but is less accurate in rapidly changing conditions.

2.1.4 Challenges in Prediction

i. High-Quality Data for Training

Lack of accurate data is one of the main obstacles to using Hard AI for weather prediction. Large, reliable datasets are needed for training machine learning systems, but the availability of extensive satellite measurements for weather prediction is relatively new, having started in 1959. This limits the amount of data in training datasets to about 10,000 days (Chantry et al., 2021).

ii. Model and Observational Data Compatibility

There is some uncertainty about the compatibility of model-generated data and observational datasets for pre-training purposes. Given the potential differences between the two forms of data, it is uncertain whether pre-training on model data before fine-tuning with observational data would be advantageous (Chantry et al., 2021).

iii. Observational Data Limitations

Using observational data, particularly from satellites to train parametrization schemes is difficult since it requires data from all vertical levels of the model. This is frequently only possible with data assimilation systems, where signal-to-noise ratios can be troublesome (Chantry et al., 2021).

iv. Computational Cost and Efficiency

Soft AI techniques seek to reduce computational costs by substituting standard parametrization schemes with AI schemes that achieve comparable accuracy while incurring lower computational costs. This saved processing resources can then be used to increase the resolution or complexity of numerical models (Chantry et al., 2021).

2.2 Real-Time Recommendation Systems in Weather Prediction

2.2.1 Definition

Real-time computing enables instant information exchange and response, matching human experience speed, minimizing latency and retaining a sense of immediacy (Noah, 2024). Although true immediacy is theoretically impossible, real-time systems strive to approach it by sending output quickly after receiving input, allowing users to interact in a timely and efficient manner (Bhargav, 2024). Meanwhile, Recommendation systems use past data and machine learning algorithms to predict user interests, improving user experience in e-commerce, streaming services, and social media. They also assist farmers in selecting crops and fertilizers based on location and weather information (Fayyaz et al., 2020).

2.2.2 Type of Recommendation Systems

Based on Figure 2.1, several techniques have been proposed for the development of Recommendation Systems including Content-Based, Collaborative Filtering and Hybrid Filtering.

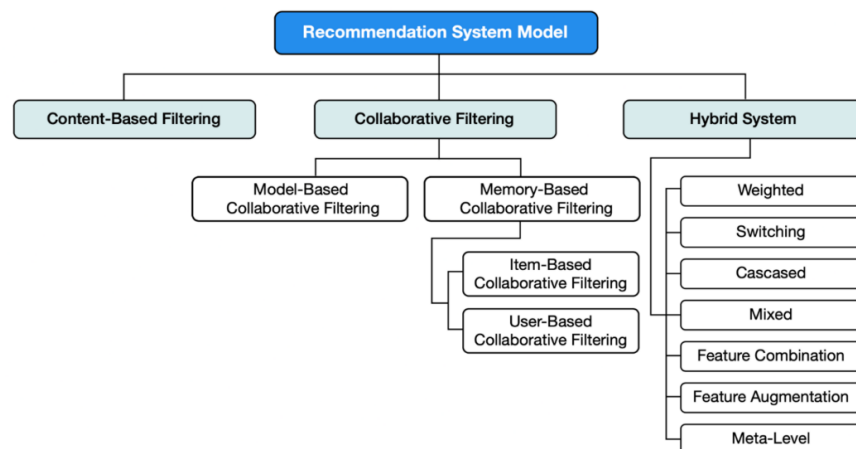


Figure 2.1 Overview of recommendation models.

i. Content-Based Filtering

Content-based recommendation systems suggest items based on user preferences, using techniques like text mining, semantic analysis, neural networks, Naive Bayes, and SVM. These systems are useful for suggesting text, music, movies, e-commerce products, and educational courses, identifying user preferences (Ko et al., 2022). The process involves analysing the contents, categories, and features of items liked or purchased by the user to recommend similar items (D. Patel et al., 2023) .

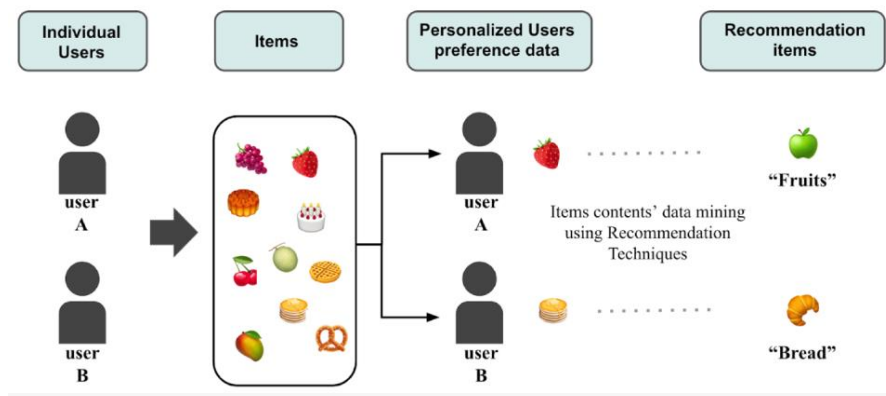


Figure 2.2 Recommendation principle of Contents-Based Filtering Model.

ii. Collaborative Filtering

Collaborative filtering, a 1990s information filtering technique, is a standard strategy in recommender systems, utilizing evaluation data to create a database of user preferences, with memory-based and model-based techniques being two types (Ko et al., 2022). Memory-based algorithms recommend items based on user-item relationships, while model-based algorithms use prediction models trained on user preference data, such as Matrix Factorization, Fuzzy Systems, Clustering, Bayesian Networks, and Deep Learning Networks (D. Patel et al., 2023). User-based collaborative filtering recommends

items based on user preferences, while item-based filtering identifies similarities between user-chosen items (Ko et al., 2022). Figure 2.3 is a visual summary of the item recommendation principle of the Collaborative Filtering Model.

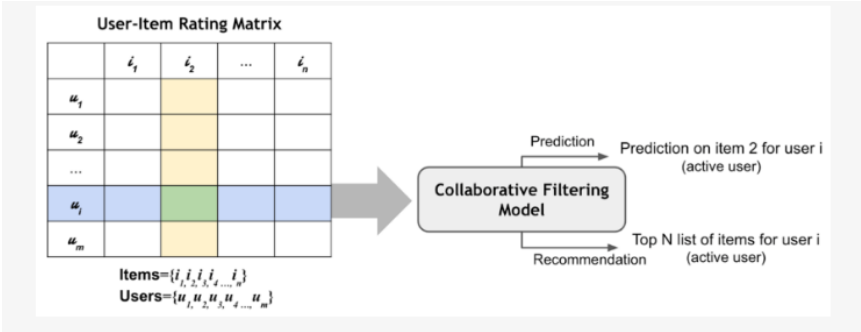


Figure 2.3 Recommendation principle of Collaborative Filtering Model.

iii. Hybrid Filtering

According to (Ko et al., 2022), both Content-Based Filtering and Collaborative Filtering have limitations: the former relies on user item metadata, while the latter depends on user item rating data. To address these limitations and enhance recommendation performance, a Hybrid recommendation model was proposed. Table 2.2 in the study outlines these seven methods of the Hybrid recommendation model.

Hybrid Method	Description
Weighted Hybridization	A method in which the weight is gradually adjusted according to the degree to which the user's evaluation of an item coincides with the evaluation predicted by the recommendation system.
Switching Hybridization	A method of changing the recommendation model used depending on the situation.
Cascaded Hybridization	After using one of the recommendation system models to create a candidate set with a similar taste to the user, the method combines the previously used recommendation system model with another model to sort the candidate set in the order of items most suited to the user's taste.
Mixed Hybridization	When many recommendations are made at the same time, Content-Based Filtering can recommend items based on the description of the items without user evaluation, but there is a start-up problem in that it cannot recommend new items with insufficient information. In order to solve this problem, the Mixed Hybridization method recommends items to the user by integrating the user's past history data that is collected when the recommendation system service is started.
Feature-Combination	A collaborative filtering model is used for featured data and example data for items, and a Content-Based Filtering model is used for augmented data.
Feature-Augmentation	A Hybrid method in which one Recommendation System Model is used to classify an item's preference score or item, and the generated information is integrated into the next Recommendation System Model.
Meta-Level	A method of using the entire model of one recommendation system as the input data in the model of another recommendation system. Since the user's taste is compressed and expressed using Meta-Level, it is easier to operate the Collaborative Mechanism than when raw rating data are used as single-input data.

Table 2.2 Seven types and concepts according to the method combining the filtering technique of the Hybrid system.

The hybrid recommendation model combines Content-Based Filtering and Collaborative Filtering models to address sparsity issues. It supplements taste data by storing user evaluation data in a matrix and learning side information data with insufficient user preference information using an auto-encoder. The model's item recommendation principle is illustrated in Figure 2.4.

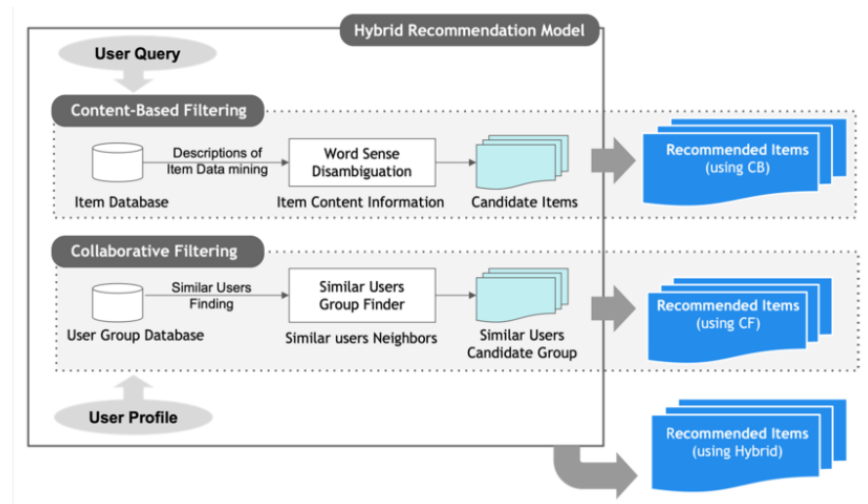


Figure 2.4 Recommendation principle of Hybrid recommendation model.

2.2.3 Challenges in Recommendation Systems

A recommendation engine might be a blessing for the modern world, but if users can easily manipulate it, it can also become an absolute misery. Consequently, the following discusses several issues that recommendation systems face:

i. Cold-start

A cold start occurs when a system lacks necessary metadata for optimal functioning, affecting both new and existing users. Techniques for cold-start recommendations include the Bayes classifier, with the naive Bayes model being the most precise. Heuristics and projection in Weighted Alternating Least Squares (WALS) can help solve this problem, allowing user embeddings without complete retraining (Fayyaz et al., 2020).

ii. Scalability

Due to their rapid expansion, scalability issues have become a major problem. While large-scale applications require fast answers, modern RS approaches suffer from performance problems when dealing with big data sets of users. Scalability is a major concern for platforms with millions of users and items since nearest-neighbour filtering algorithms demand more processing capacity due to the significant growth in users or products (Fayyaz et al., 2020).

iii. Data sparsity

Data sparsity in the Recommendation System (RS) is caused by users obtaining just a limited number of objects, resulting in empty or unknown ratings. This may lead to irrational suggestions being made for people who don't rate or comment. Modelling user preferences based on behaviours and reliable social relationships is one way to reduce data sparsity. Trust has been widely employed to increase the

robustness of RSs. Trust is defined as the belief in the ability of people to deliver accurate ratings. Trust networks and trust charts, which have lowered the mean error of predicted accuracy, can be used to determine trustworthiness. The merge strategy has been implemented to improve the overall forecast accuracy of RS by incorporating trusted neighbours of active users (Fayyaz et al., 2020).

iv. Diversity

Recommendation algorithms often provide similar or diverse ideas but often rely on overlapping similarities rather than differences, leading to a limited selection and the omission of specialized items. Increasing diversity can improve accuracy, but focusing solely on it can lead to over-concentration. Surprising and personalization measures can assess a recommendation system's diversity, while customization measures measure the distinctiveness of users' lists. (Fayyaz et al., 2020).

2.2.4 Comparison for Recommendation Systems

Name	Benefits	Disadvantages	Application
Collaborative	Ratings from users are taken into consideration	Low performance for first users.	Information portals
	Not tied to the subject of the service	A significant amount of data involving user ratings is required	Small online stores
Content-based	Functions immediately, especially on platforms for novice users	Linked to the content of the service	Blogs
	Works correctly, even with a small amount of data	Not based on the wishes of users	Music or movie
Hybrid	High productivity	Hard to maintain	Large online stores
	No problems with previous approaches	Complex development process	Complicated systems with many of users

Table 2.3 Comparison of Recommendation Systems

Table 2.3 compares recommendation systems in weather prediction, addressing challenges like cold start, scalability, data sparsity, and diversity. This will help understand the strengths and limitations of each approach, leading to more effective and efficient systems for weather prediction applications.

2.3 Supervised Learning

2.3.1 Definition

SL stands for supervised learning, which uses labelled data sets and supervision. Label data sets are those that already have the intended response known (Al-Sahaf, H., 2019). Figure 2.5 provides an excellent illustration of a labelled data collection and a supervised learning procedure: model of supervised learning. The data sets collected are the results of using the input data. Since the input data is referred to as label data, the labelled data is known as apple. The processing of supervised learning algorithms serves as the model. As apples are the model's predicted fruit, the result is accurate (Gupta et al., 2022).

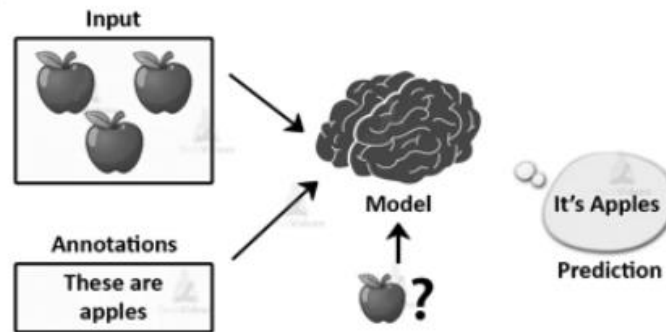


Figure 2.5 Supervised Learning Model

2.3.2 Types of Supervised Learning

Supervised Learning works on two different categories Regression and Classification.

i. Classification

In machine learning it refers to a predictive modelling task in which a class label is predicted for a given example. It is a mathematical mapping of a function (f) as a target, label, or category from input variables (X) to output variables (Y). Structured or unstructured data can be used to predict the class of the provided data points. For instance, email service providers' spam detection features, such as

"spam" and "not spam," may present classification challenges. Most famous supervised machine learning algorithms have been discussed here (Sarker, 2021).

- a) *Naïve Bayes (NB)*: The Naïve Bayes algorithm, based on Bayes' theorem, is a widely used classifier for real-world scenarios like spam filtering and categorization. It assumes independent feature pairs, efficiently categorizing noisy data and building reliable prediction models. However, its performance may be affected by its feature independence assumptions, requiring less training data for rapid parameter estimation (Sarker, 2021).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability

Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Figure 2.6 Overview of Naïve Bayes structure

(Mahesh, 2019).

- b) *Random Forest (RF)*: The random forest classifier is a widely used ensemble classification method in machine learning and data science. It uses "parallel ensembling" to fit multiple decision tree classifiers on different data sets, reducing overfitting and improving control and prediction accuracy. RF learning models with multiple decision trees are more accurate than those based on a single decision tree. This method combines random feature selection with bootstrap aggregation to create a sequence of decision trees with controlled variance, making it suitable for regression and classification challenges (Sarker, 2021).

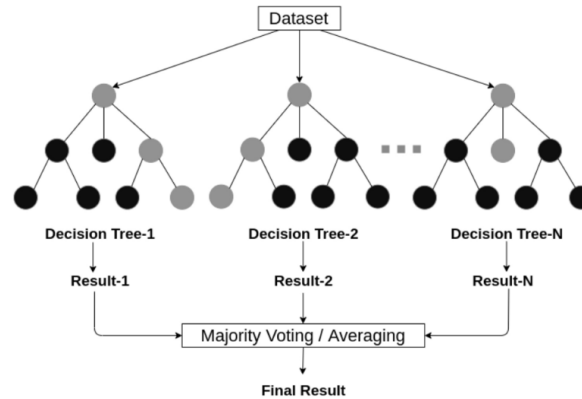


Figure 2.7 An example of a random forest structure considering multiple decision trees (Sarker, 2021).

- c) *Support Vector Machines (SVM)*: SVM are nonparametric classifiers based on statistical learning theory, used to classify data with uncertain relationships between variables. They can also categorize nonlinear and multi-class data. The SVM classifier's principle is to define the hyperplane that best distinguishes between two classes, maximizing the distance between support vectors, and using this hyperplane to generate an optimal decision function(Kavzoglu et al., 2020).
- d) *Decision Tree (DT)*: A decision tree is a graph that represents options and outcomes in the shape of a tree. The graph's nodes represent events or choices, while its edges reflect decision rules or conditions. Each tree has nodes and branches. Each node represents an attribute in a classification group, and each branch indicates a possible value for the node (Mahesh, 2019).

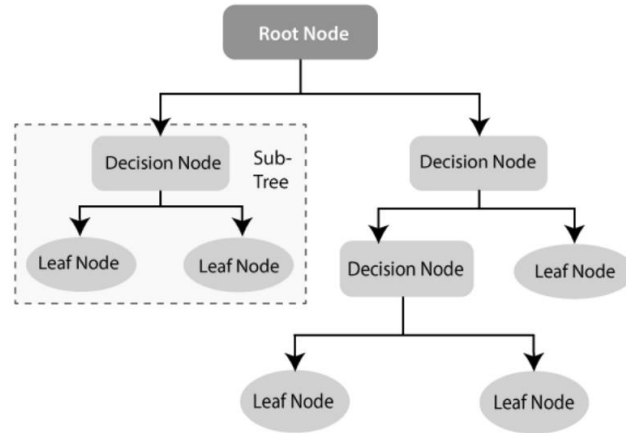


Figure 2.8 An example of a decision tree structure (Sarker, 2021).

ii. Regression

Regression is a machine learning technique that generates accurate predictions from input variables, such as continuous outcome variables (Y) dependent on predictor variables (X). It aims to develop a mathematical model explaining the Y function of the x variables, using sophisticated linear regression algorithms. (Ayaz Mirani et al., 2021).

a) *Simple and multiple linear regression:* Linear regression is a popular machine learning (ML) technique that uses a linear regression line to establish a relationship between a continuous dependent variable (Y) and one or more independent variables (X) using the best-fit straight line. It is defined by the following equations:

$$y = a + bx + e \quad (1)$$

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n + e, \quad (2)$$

where e is the error term, b is the line's slope, and an is the intercept. While basic linear regression only includes one independent variable, defined in Eq. 1, multiple linear regression is an extension of simple linear regression that allows two or more

predictor variables to model a response variable, y , as a linear function specified in Eq. 2 (Sarker, 2021).

b) Polynomial regression: In polynomial regression, the link between the independent variable (x) and the dependent variable (y) is expressed as the polynomial degree of n^{th} in x , rather than as a linear relationship. The linear regression (polynomial regression of degree 1) equation, which is defined as follows, is also the source of the polynomial regression equation:

$$y = b_0 + b_1x^1 + b_2x^2 + b_3x^3 + \dots + b_nx^n + e \quad (3)$$

Here, y is the predicted/target output, b_0, b_1, \dots, b_n are the regression coefficients, x is an independent/ input variable. In simple words, we can say that if data are not distributed linearly, instead it is n^{th} degree of polynomial then we use polynomial regression to get desired output (Sarker, 2021).

c) Lasso Regression: LASSO and Ridge regression are effective methods for minimizing complexity and avoiding overfitting in learning models with multiple characteristics. LASSO penalizes the absolute value of coefficients to reduce prediction error, while Ridge regression minimizes weights without setting the coefficient value to zero. These approaches help reduce multicollinearity by removing less crucial features (Sarker, 2021).

2.3.3 Applications of Supervised Learning

i. Fraud Detection

The market for online credit card fraud detection is expected to reach \$32 billion by 2020. That is roughly more than the total profit earned by numerous multinational corporations. The quantity of payment methods available today, such as credit cards, debit cards, multiple wallets, UPI, and much more, has led to a rise in criminal activity. Fraud detection is viewed by machine learning as a classification task (Shetty et al., 2022).

ii. Image, speech, and pattern recognition

Image identification is a widely used application of machine learning, identifying objects as digital images. It's used in various applications like cancer classification, identifying faces or characters, and suggesting tags on social media. Speech recognition systems like Google Assistant use machine learning to recognize sounds and understand language. Pattern recognition is the automatic identification of patterns in data, using techniques like feature selection, clustering, classification, and sequence labelling (Sarker, 2021).

iii. NLP and sentiment analysis

Natural language processing (NLP) is a computer-based method that uses machine learning to interpret and analyse written or spoken information, including chatbots, speech recognition, document descriptions, and translation. Sentiment analysis, a branch of NLP, extracts opinions and public mood from writings like blogs, reviews, and social media, enabling enterprises to understand social sentiment and emotions (Sarker, 2021).

2.3.4 Comparison between Techniques in Supervised Learning

Techniques	Support Vector Machines (SVM)	Random Forest	Decision Tree
Accuracy	96.82% (Low)	99.32% (High)	98.18% (Medium)
Interpretability	The complicated hyperplane separation in high-dimensional spaces makes the data less interpretable. (High)	Despite being correct, the ensemble aspect of integrating numerous trees may cost some interpretability. (Low)	Extremely interpretable since they show a sequence of judgments based on attributes, which facilitates comprehension of the logic underlying the forecasts. (Medium)
Computational Complexity	Can be computationally expensive, especially with large datasets. (High)	Requires more computational resources than Decision Trees due to multiple trees. (Medium)	Relatively faster to build and evaluate compared to SVM and Random Forest. (Low)
Robustness to overfitting	Margin maximization helps in generalizing well but requires careful parameter tuning. (Medium)	Addresses overfitting through ensemble averaging and feature randomness. (High)	Prone to overfitting with complex datasets due to its hierarchical nature. (Low)

Table 2.4 Comparison between Techniques (Shende et al., 2024)

Table 2.4 compares three supervised learning techniques: Support Vector Machines (SVM), Random Forest, and Decision Trees, focusing on accuracy, interpretability, computational complexity, and robustness to overfitting. SVMs have a lower accuracy of 96.82%, while Random Forests have a higher accuracy of 99.32% and moderate interpretability. Decision Trees have a higher accuracy of 98.18% but are prone to overfitting.

2.4 Similar Application

In this part, the explanation is about the summary of existing related work, which is from previous research related to the prediction of real-time weather with recommendations using machine learning. Moreover, this project is motivated by the thesis produced by the previous research. All the research explains how the prediction is conducted using different techniques and approaches.

2.4.1 Weather Prediction Using Futuristic Technologies: From Industry's Perspective

The study emphasizes the importance of precise weather forecasting in modern society, highlighting challenges like data scarcity, atmospheric complexity, and extreme weather. It highlights the role of advanced technologies like artificial intelligence, machine learning, big data analytics, remote sensing, and citizen science in improving weather prediction accuracy. The use of satellite data, drones, and UAVs allows real-time data collection, leading to more accurate forecasts. The study also highlights potential benefits for industries like agriculture and disaster response (Kumari & Muthulakshmi, 2023).

2.4.2 Deep Learning-Based Weather Prediction: A Survey

The study paper analyses the application of deep learning in weather forecasting, concentrating on its superiority over traditional methods in capturing spatio-temporal aspects and processing time series data for increased prediction accuracy. It illustrates the necessity of precise forecasts in numerous facets of human existence and analyses the potential uses of deep learning-based weather prediction in real-world scenarios. The study also highlights the integration of scientific knowledge with data-driven models to increase interpretability and prediction precision, indicating a trend toward theory-guided data science in meteorology (Ren et al., 2021).

2.4.3 Weather monitoring and forecasting system using IoT

The research study focuses on creating a weather monitoring system using IoT technologies to monitor various environmental factors like temperature, humidity, wind speed, light intensity, UV radiation, carbon monoxide levels, soil moisture, and rainwater levels. The system provides precise weather forecasts using an API for data analysis and prediction. It also features a mobile application for user safety. The system's potential applications include meteorological agencies, weather stations, aerospace, marine sectors, and agriculture, highlighting its cost-effectiveness and minimal maintenance requirements (Islam et al., 2023).

2.4.4 Development of flood forecasting and warning system using hybrid approach of ensemble and hydrological model for Dharoi Dam

The study aims to create a flood forecasting system using a hybrid method of ensemble and hydrological models for the Sabarmati River basin near the Dharoi Dam. The model is built using RStudio and HEC-HMS software, and TIGGE data is used to improve flood predictions and rainfall-runoff dynamics simulations. The project aims to increase lead time for flood predictions, evaluate system effectiveness in real-time forecasting, and provide scalability and adaptability for flood resistance (A. Patel & Yadav, 2023).

2.4.5 Comparison between Similar Applications

Article Title	Innovation Trends	Methodology	Advantages	Disadvantages
Weather Prediction Using Futuristic Technologies: From Industry's Perspective	Discusses the increasing role of AI in weather prediction and the prevalence of AI-centric keywords such as "machine learning" and "neural network".	Uses patent data to understand trends in weather prediction innovation.	Explores the potential of future weather prediction technologies, highlighting potential advancements and industry needs for future weather prediction systems.	May not be based on currently available technologies, making implementation difficult.
Deep Learning-Based Weather Prediction: A Survey	Highlights the increasing popularity of machine learning, particularly deep learning, in meteorology.	Reviews more than 20 methods highlighted in existing literature.	Surveys existing deep learning techniques used for weather prediction and offers insights into the current state-of-the-art methods in this field.	May be too technical for non-experts to understand and focuses on deep learning only, potentially excluding other promising approaches.
Weather monitoring and forecasting system using IoT	Discusses the benefits of IoT-based weather monitoring systems, such as real-time data collection, higher accuracy, and wider coverage.	Uses IoT devices and sensors for real-time data collection.	The study explores the use of real-world technologies (IoT) for weather prediction, suggesting that IoT-based systems could be deployed in various locations.	Relying on the accuracy and coverage of deployed IoT sensors and processing large amounts of IoT data might require significant computing power.
Development of flood forecasting and warning system using hybrid approach of ensemble and hydrological model for Dharoi Dam	Discusses the development of an atmospheric-hydrologic flood forecasting model.	Uses ensemble and hydrological models in predicting floods.	Focuses on a specific problem (flood forecasting) for a particular location (Dharoi Dam) and utilizes a hybrid model potentially leading to more accurate predictions.	The specific approach is not directly transferable to other locations or flood forecasting scenarios, and it relies on the availability of historical data for the Dharoi Dam and surrounding area.

Table 2.5 Comparison between Similar Applications

2.5 Summary

This chapter provides an extensive overview of weather forecast methodologies and the incorporation of suggestion systems through machine learning. It starts by defining weather prediction and the importance of accurate predictions, for example, in agriculture and event planning. The text discusses traditional methods of weather forecasting such as climatology, analog, persistence, and trends, as well as numerical weather prediction, which have all their own advantages and disadvantages. It focuses on the difficulties faced when doing a weather forecast, such as the quality of the of the data needed and computational efficiency. Additionally, it investigates various types of recommendation systems, including content-based, collaborative filtering, and hybrid filtering, to address problems like the cold-start problem, scalability issues, and data sparsity, among others. This study underscores the potentiality of machine learning, with special reference to deep learning, in enhancing the accuracy of weather forecasts by effectively dealing with spatio-temporal data and time series. Furthermore, it talks about applying IoT-based real-time meteorological observation along with the development of a hybrid flood forecasting system. Finally, this paper summarizes how these advanced technologies enhance weather forecasting techniques.

2.6 References

- Ayaz Mirani, A., Muhammad Suleman Memon, E., Qabulio, M., Suleman Memon, M., Chohan, R., & Ali Wagan, A. (2021). Machine learning in agriculture: A review. *Researchgate.Net*. https://www.researchgate.net/profile/Engr-Dr-Muhammad-Suleman-Memon/publication/351766285_Machine_Learning_In_Agriculture_A_Review/links/60a83aae45851522bc0a5479/Machine-Learning-In-Agriculture-A-Review.pdf
- Bhargav, N. (2024, March 18). *Real-Time Systems / Baeldung on Computer Science*. <https://www.baeldung.com/cs/real-time-systems>
- Chantry, M., Christensen, H., Dueben, P., & Palmer, T. (2021). Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI. *Philosophical Transactions of the Royal Society A*, 379(2194). <https://doi.org/10.1098/RSTA.2020.0083>
- Chetan, D. D., & Waghmare, C. (2022). A Literature Review on Improvement of Weather Prediction By Using Machine Learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.4140237>
- Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Applied Sciences* 2020, Vol. 10, Page 7748, 10(21), 7748. <https://doi.org/10.3390/APP10217748>
- Gupta, V., Mishra, V. K., Singhal, P., & Kumar, A. (2022). An Overview of Supervised Machine Learning Algorithm. *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022*, 87–92. <https://doi.org/10.1109/SMART55829.2022.10047618>
- Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. F. (2023). Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. *Journal of Agriculture and Food Research*, 14, 100880. <https://doi.org/10.1016/J.JAFR.2023.100880>
- Kavzoglu, Taskin & Bilücan, Furkan & Teke, & Alihan. (2020). *COMPARISON OF SUPPORT VECTOR MACHINES, RANDOM FOREST AND DECISION TREE METHODS FOR CLASSIFICATION OF SENTINEL - 2A IMAGE USING DIFFERENT BAND COMBINATIONS*. https://www.researchgate.net/publication/346776010_COMPARISON_OF_SUPPORT_VECTOR_MACHINES_RANDOM_FOREST_AND_DECISION_TREE_METHODS_FOR_CLASSIFICATION_OF_SENTINEL_-_2A_IMAGE_USING_DIFFERENT_BAND_COMBINATIONS

- Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics* 2022, Vol. 11, Page 141, 11(1), 141. <https://doi.org/10.3390/ELECTRONICS11010141>
- Kumari, S., & Muthulakshmi, P. (2023). Weather Prediction Using Futuristic Technologies: From Industry's Perspective. *ACM International Conference Proceeding Series*, 245–251. <https://doi.org/10.1145/3607947.3607994>
- Mahesh, B. (2019, January). (PDF) *Machine Learning Algorithms -A Review*. ResearchGate. https://www.researchgate.net/publication/344717762_Machine_Learning_Algorithms_-_A_Review
- Noah. (2024, May 26). *What is real time (real-time)? Definition, Computing & Data – Analyst Answers*. <https://analystanswers.com/what-is-real-time-real-time-definition-computing-data/>
- Patel, A., & Yadav, S. M. (2023). Development of flood forecasting and warning system using hybrid approach of ensemble and hydrological model for Dharoi Dam. *Water Practice and Technology*, 18(11), 2862–2883. <https://doi.org/10.2166/WPT.2023.178/1311410/WPT2023178.PDF>
- Patel, D., Patel, F., & Chauhan, U. (2023). Recommendation Systems: Types, Applications, and Challenges. *International Journal of Computing and Digital Systems*, 13(1), 851–868. <https://doi.org/10.12785/IJCDS/130168>
- Patkar, U. (2022, August). *WEATHER PREDICTION USING MACHINE LEARNING*. ResearchGate. https://www.researchgate.net/publication/362517661_WEATHER_PREDICTION_USING_MACHINE_LEARNING
- Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., & Wang, X. (2021). Deep Learning-Based Weather Prediction: A Survey. *Big Data Research*, 23. <https://doi.org/10.1016/j.bdr.2020.100178>
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 1–21. <https://doi.org/10.1007/S42979-021-00592-X/FIGURES/11>
- Shende, A., Yeole, S., Kanakia, H. T., & Salmani, S. (2024). Comparative Analysis of SVM, Decision Tree, Random Forest for Crop Recommendation. *2nd International Conference on Intelligent Data Communication Technologies and Internet of Things, IDCIoT 2024*, 1018–1022. <https://doi.org/10.1109/IDCIOT59759.2024.10467414>
- Shetty, S. H., Shetty, S., Singh, C., & Rao, A. (2022). Supervised machine learning: Algorithms and applications. *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools, and Applications*, 1–16. <https://doi.org/10.1002/9781119821908.CH1>