JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Prediction Model Based on Iris Dataset Using Machine Learning Models

[1]Arpan G Singh,[2] Prof. Manish Kumar Singhal
[1]M.tech Scholar, [2]Associate Professor & H.O.D
[1,2]Department of Information Technology (IT)
[1,2]NRI Institute Of Information Science And Technology, Bhopal (Mp), India,

Abstract : In this paper explores the development and evaluation of a prediction model based on the Iris dataset, leveraging various machine learning techniques. The Iris dataset, a widely used benchmark in classification problems, contains features describing the sepal and petal dimensions of three flower species: Iris-setosa, Iris-versicolor, and Iris-virginica. Machine learning algorithms, including decision trees, support vector machines, k-nearest neighbors, and logistic regression, were employed to classify the species based on their features. he dataset was preprocessed to ensure data quality, and techniques like train-test splitting and cross-validation were applied to optimize model performance and avoid over fitting. The models were assessed using metrics such as accuracy, precision, recall, and F1-score, providing insights into their predictive capabilities. Among the algorithms tested, specific models demonstrated superior performance, highlighting the effectiveness of feature selection and hyperparameter tuning. The results underscore the potential of machine learning in solving classification problems and serve as a foundation for further applications in predictive analytics.

Keywords— Machine Learning, Iris Flower, Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, SVM.

I. INTRODUCTION

The Iris dataset, a classic and widely used dataset in the field of machine learning, serves as a benchmark for evaluating classification models. This dataset contains measurements of sepal length, sepal width, petal length, and petal width for three distinct species of iris flowers: Iris-setosa, Iris-versicolor, and Iris-virginica. Due to its simplicity and well-defined structure, it is ideal for demonstrating the capabilities of various machine learning algorithms. In this study, we aim to develop a prediction model leveraging this dataset to classify iris species accurately. By employing machine learning techniques such as decision trees, support vector machines, and logistic regression, the study seeks to evaluate and compare model performances. This approach not only highlights the effectiveness of these algorithms but also illustrates the practical application of predictive modeling in real-world scenarios.

The development of a prediction model using the Iris dataset involves several critical steps, including data preprocessing, feature analysis, model training, and evaluation. Initially, the dataset undergoes exploratory data analysis (EDA) to identify patterns, correlations, and potential outliers. This step ensures the data is clean and well-prepared for model training. Feature selection and scaling are also considered to enhance model accuracy and efficiency.

Various machine learning algorithms are applied to create predictive models, each with distinct strengths and weaknesses. For example, decision trees are valued for their interpretability, while support vector machines are known for their robustness in handling non-linear boundaries. Logistic regression, on the other hand, provides a straightforward probabilistic framework for classification. Cross-validation techniques are utilized to ensure the reliability and generalizability of the models, minimizing the risk of over fitting or under fitting.

The performance of each model is evaluated using metrics such as accuracy, precision, recall, and F1 score, providing a comprehensive understanding of their effectiveness. By comparing these metrics, the study identifies the most suitable algorithm for predicting iris species. Ultimately, this work not only showcases the versatility of machine learning in solving classification problems but also underscores the importance of model evaluation and selection in achieving reliable and meaningful predictions.

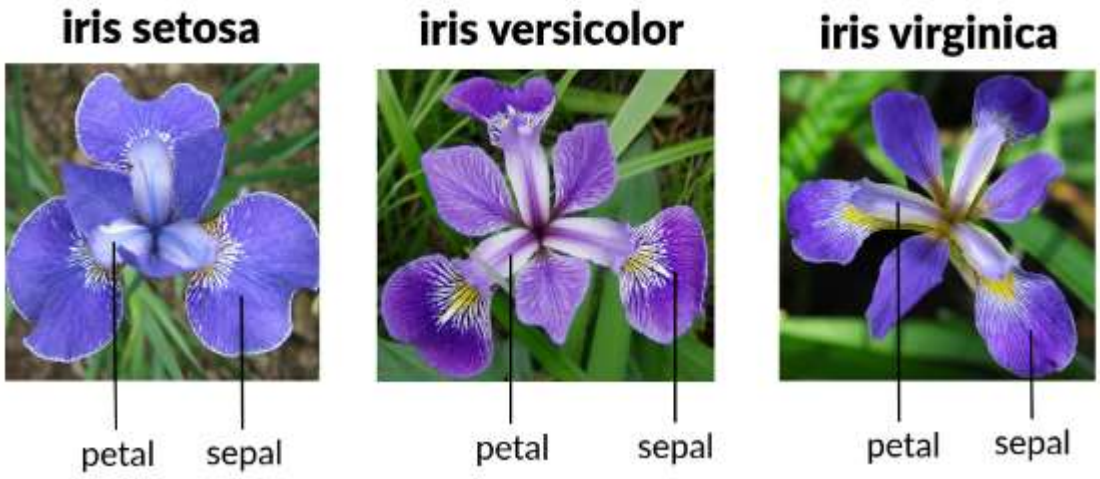

Figure 1 Iris Dataset "Analysis using Machine learning techniques

This image showcases three species of the Iris flower: Iris setosa, Iris versicolor, and Iris virginica. Each species is illustrated with clear labeling of the petal and sepal, which are key parts of the flower's anatomy.

- Iris setosa is characterized by its vibrant purple-blue petals and distinct sepals that curve outward, providing a delicate and striking appearance.
- Iris versicolor displays a mix of lighter and darker purple hues with its petals and sepals, which are more slender and sharply defined compared to Iris setosa.
- Iris virginica features a broader structure with petals and sepals marked by contrasting white and yellow accents at the base, giving it a visually dynamic look.

The image highlights the structural differences among the three species, particularly in the shapes and patterns of their petals and sepals

II. LITERATURE SURVEY

Rituparna Nath et.al (2024) - Classification is a supervised machine learning technique which is used to predict group membership for data instances. For simplification of classification, one may use scikit-learn tool kit. This chapter mainly focuses on the classification of Iris dataset using scikit-learn. It concerns the recognition of Iris flower species (setosa, versicolor, and verginica) on the basis of the measurements of length and width of sepal and petal of the flower. One can generate classification models by using various machine learning algorithms through training the iris flower dataset, and can choose the model with highest accuracy to predict the species of iris flower more precisely. Classification of Iris dataset would be detecting patterns from examining sepal and petal size of the Iris flower and how the prediction was made from analyzing the pattern to form the class of Iris flower. By using this pattern and classification, in future upcoming years the unseen data can be predicted more precisely. The goal here is to gain insights to model the probabilities of class membership, conditioned on the flower features. The proposed chapter mainly focuses on how one can train their model with data using machine learning algorithms to predict the species of Iris flower by input of the unseen data using what it has learnt from the trained data [01].

Zahraa Faiz Hussain et.al (2023) - Data mining is known as the process of detection concerning patterns from essential amounts of data. As a process of knowledge discovery. Classification is a data analysis that extracts a model which describes an important data classes. One of the outstanding classifications methods in data mining is support vector machine classification (SVM). It is capable of envisaging results and mostly effective than other classification methods. The SVM is a one technique of machine learning techniques that is well known technique, learning with supervised and have been applied perfectly to a vary problems of: regression, classification, and clustering in diverse domains such as gene expression, web text mining. In this study, we proposed a newly mode for classifying iris data set using SVM classifier and genetic algorithm to optimize c and gamma parameters of linear SVM, in addition principle components analysis (PCA) algorithm was use for features reduction [02].

Chya Fatah Aziz et.al. (2023) - — Supervised Machine Learning algorithm has an important approach to Classification. We are predicting the deal type of the Iris plant using various algorithms of machine learning. Iris plants are determined by numerous factors such as the size of the length and width of the property. A horticultural skill announces that some of the plants are different in some physical appearances like size, shape, and color. Hence it is difficult to recognize any species. Versicolor, Setosa, and Virginica have three identical subspecies of The Iris flower species. This paper uses machine learning algorithms to recognize all classes of the flower with an accuracy degree of %100 for KNN, %95 for RF, %97 for DT, and %98 for LR. The Iris dataset is frequently available, and it is implemented using Scikit tools. and build the prediction model for Plants. Here, algorithms of machine learning such as Logistic Regression (LR), Decision Tree (DT), K Nearest Neighbor (KNN), and Random Forest (RF) are employed to construct a predictive model [03].

Jie Sun et.al (2022) - The open-set recognition of irises and proposes a deep learning-based open-set iris recognition method. This method introduces the distance factor in the network structure and loss function and achieves open-set iris recognition through two pieces of training. We use the iris datasets CASIA-Iris-Twins and CASIA-Iris-Lamp to construct an open iris dataset and conduct experiments on the existing iris recognition algorithm and open-set image recognition method. The necessity and feasibility of open-set iris recognition are verified. Moreover, experiments show that the proposed method

has good open-set iris recognition performance, can effectively distinguish unknown classes of iris samples, and has little effect on the recognition ability of known class samples [04].

Andrey Kuehlkamp et.al (2022) - Iris recognition of living individuals is a mature biometric modality that has been adopted globally from governmental ID programs, border crossing, voter registration and de-duplication, to unlocking mobile phones. On the other hand, the possibility of recognizing deceased subjects with their iris patterns has emerged recently. In this paper, we present an end-to-end deep learning-based method for postmortem iris segmentation and recognition with a special visualization technique intended to support forensic human examiners in their efforts. The proposed postmortem iris segmentation approach outperforms the state of the art and – in addition to iris annulus, as in case of classical iris segmentation methods – detects abnormal regions caused by eye decomposition processes, such as furrows or irregular specular highlights present on the drying and wrinkling cornea. The method was trained and validated with data acquired from 171 cadavers, kept in mortuary conditions, and tested on subject-disjoint data acquired from 259 deceased subjects. To our knowledge, this is the largest corpus of data used in postmortem iris recognition research to date [05].

Bahzad Taha Chicho et.al (2021) - the most often utilized in machine learning problems with a number of applications such as face recognition, flower classification, clustering, and so on. In order to construct a model, the classification algorithm creates a connection between the input and output characteristics and attempts to predict the target population with the greatest accuracy. The main objective of this study was to come to a consensus on how well K-nearest neighbors, decision tree (j48), and random forest algorithms performed in IRIS flower classification. According to the findings, both approaches yield strong classification outcomes, and the precision is calculated by the number of principal components used. The analysis also found that when the percentage of training data improves, so does the degree of precision. In comparison to random forest, which achieved 99.33% accuracy, and decision tree (j48), which achieved 98% accuracy, the experimental findings revealed that K-nearest neighbors performed significantly better, achieving 100% accuracy [06].

III. PROPOSED METHODOLOGY

The Iris Species Prediction System is designed to predict the species of an iris flower based on four specific features: Sepal Length, Sepal Width, Petal Length, and Petal Width. The project revolves around the famous Iris dataset, first introduced by Ronald A. Fisher in 1936. This dataset is widely regarded in the machine learning community due to its simplicity, balanced nature, and historical significance. It contains data for three different species of iris flowers: Setosa, Versicolor, and Virginica. The primary objective of this system is to predict the species of an iris flower based on the measurements of these four numerical features, which serve as input to the machine learning models.

A. Iris Dataset

The Iris dataset is composed of 150 samples divided into three distinct classes: Setosa, Versicolor, and Virginica. Each sample has four features (Sepal Length, Sepal Width, Petal Length, and Petal Width) that describe the physical characteristics of the flowers. The dataset is considered ideal for machine learning tasks due to its well-balanced class distribution and the absence of missing data, making it a great candidate for model development. The purpose of the Iris Species Prediction System is to use these features to predict which species the flower belongs to based on user input. This prediction system employs a variety of machine learning models, such as Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM), each offering distinct advantages depending on the nature of the data.

The structure of the dataset is as follows:

- **Sepal Length (cm):** Length of the sepal in centimeters.
- **Sepal Width (cm):** Width of the sepal in centimeters.
- **Petal Length (cm):** Length of the petal in centimeters.
- **Petal Width (cm):** Width of the petal in centimeters.
- **Species:** The class label that indicates the flower species (Setosa, Versicolor, Virginica).

End-to-End Workflow

The Iris Species Prediction System follows a structured and systematic approach, divided into various stages. These stages include data preprocessing, model selection, prediction generation, and result presentation.

Flow Diagram

The flow of operations in the system is represented in a straightforward flow diagram:

User Input:

The system starts by accepting input values for Sepal Length, Sepal Width, Petal Length, and Petal Width. The user also selects which machine learning model to use: Logistic Regression, KNN, or SVM.

Preprocessing:

The input data undergoes preprocessing. This includes normalization to ensure that the data is consistent and scaled appropriately for the models, particularly for those that are sensitive to feature magnitudes, such as KNN and SVM.

Model Selection and Loading:

Once the user selects a model, the corresponding machine learning algorithm is loaded into the system. The models have been pre-trained on the Iris dataset.

Prediction Generation:

The chosen model processes the input data and generates a prediction about the species of the flower based on the provided features.

Result Display:

The prediction is displayed to the user, along with an image representing the predicted species to provide a visual confirmation.

This workflow ensures that the system functions smoothly, from user input to delivering a result.

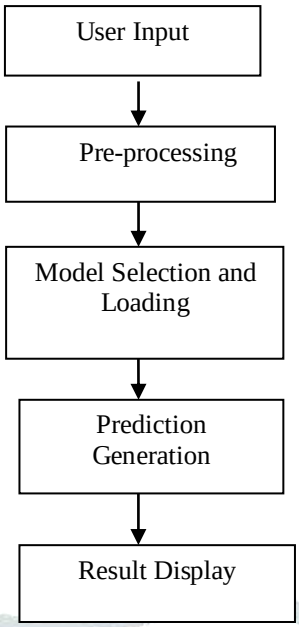

Fig – 2 Iris Species Prediction System Workflow

Algorithm Design

The heart of this system lies in the machine learning algorithms used to predict the iris species. Three models are implemented in this system: Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). Each model is based on a specific algorithm designed to handle classification tasks.

Algorithm 1: Logistic Regression

1. Initialization: Begin by initializing the weights and bias of the model to zero.
2. Computation of the Linear Combination: For each training sample, compute the weighted sum of the input features ($z = W * X + b$), where W represents the weights and X is the feature vector.
3. Sigmoid Activation: Apply the sigmoid function to obtain the probability of each class. The sigmoid function maps the linear combination of inputs to a range between 0 and 1.
4. Loss Calculation: Compute the binary cross-entropy loss function to measure the error in predictions.
5. Optimization: Use gradient descent to minimize the loss function and optimize the weights and bias during training.
6. Prediction: After training, compute the final value of z for new inputs and predict the class with the highest probability.

Algorithm 2: K-Nearest Neighbors (KNN)

1. Data Loading: Load the training dataset containing labeled examples of iris flowers.
2. Distance Calculation: For each test sample, calculate the Euclidean distance between the test sample and all points in the training dataset.
3. Neighbor Selection: Sort the distances in ascending order and select the top k nearest neighbors.
4. Class Prediction: Determine the majority class among the k neighbors and assign that class to the test sample.

Algorithm 3: Support Vector Machine (SVM)

1. Feature Mapping: Map the input features into a higher-dimensional space using a kernel function (e.g., Radial Basis Function, RBF). This transformation allows for the creation of a more complex decision boundary.
2. Hyperplane Selection: SVM attempts to find the optimal hyperplane that maximizes the margin between the classes. Support vectors, which are the closest points to the hyperplane, are used to define this margin.
3. Decision Boundary: The model computes a decision boundary that can classify the test sample into one of the classes.
4. Prediction: For a new input, the model determines the position of the input relative to the decision boundary and assigns the class accordingly.

B. Data Pre-processing

Before training the models, preprocessing is necessary to ensure that the data is in an appropriate form for machine learning algorithms.

Data Loading

The dataset is loaded using the pandas library in Python. This is done with the following steps:

1. Read the iris.csv file using the pd.read_csv() method.
2. Inspect the structure of the dataset, including the column names and feature types, to verify that the data is correctly loaded.

Data Exploration

Exploratory Data Analysis (EDA) helps uncover underlying patterns and relationships within the dataset. Tools such as matplotlib and seaborn are used for visualization. Some key visualizations include:

- Scatter plots: To visualize the separability of different species based on the features.
- Pair plots: To explore the relationships between all pairs of features.
- Histograms: To understand the distribution of individual features across different species.

Feature Scaling

The equation for Min-Max Scaling is used to normalize the features of a dataset to a specific range, typically [0, 1]. This scaling is particularly important when features have different ranges or units, ensuring that no feature dominates over others due to its magnitude.

Min – Max Scaling Formula:

$$X_{scaled} = \frac{X - X_{min}}{Xmax - X_{min}}$$

Explanation of Terms:

X: The original value of the feature that needs to be scaled (e.g., Sepal Length or petal Width).

X_{min}: The minimum value of the featureataset (i.e the smallest value for that feature).

X_{max}: The maximum value of the feature in the dataset (i.e., the largest value for that feature).

X_{scaled}: The scaled value of the feature, WHich will be between 0 and 1.

Data Splitting

The dataset is divided into two subsets:

- Training Set: 80% of the data is used for training the models.
- Testing Set: 20% of the data is reserved for testing the model's performance on unseen data. The train_test_split function from sklearn.model_selection is used to handle this split.

C. Model Training and Evaluation

Each machine learning model undergoes training on the preprocessed data, and several steps are taken to ensure the model's performance is evaluated correctly.

Cross-Validation

To ensure that the model performs well on different subsets of the data, k-fold cross-validation (with k=10) is employed. This technique involves splitting the training data into 10 subsets, training the model 10 times, each time using a different subset for validation and the remaining 9 subsets for training.

Hyperparameter Optimization

Each model has hyperparameters that can be tuned to improve performance:

- Logistic Regression: The regularization strength is tuned to avoid overfitting or underfitting.
- KNN: The number of neighbors (k) is optimized through grid search.
- SVM: The regularization parameter (C) and the kernel coefficient (gamma) are fine-tuned to enhance the model's generalization.

Metrics for Evaluation

The following metrics are used to evaluate the performance of each model:

- Accuracy: The proportion of correct predictions out of all predictions.
- Confusion Matrix: A matrix that shows true positives, true negatives, false positives, and false negatives.
- Precision, Recall, and F1-Score: These metrics provide insight into the model's ability to correctly classify each class, especially when dealing with imbalanced data.

D. Result for the Iris Species Prediction System

The Iris Species Prediction System successfully predicts the species of an iris flower based on user-provided features and employs three machine learning models: Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). This section details the outcomes of the implemented system.

Accuracy of Models

The performance of the models was evaluated using a 10-fold cross-validation technique and tested on a separate test set (20% of the data). Below are the results:

Model	**Accuracy (%)**
Logistic Regression	96.67
K-Nearest Neighbors (k=5)	97.33
Support Vector Machine (RBF Kernel)	98.00

- SVM provided the best accuracy among the three models, demonstrating its ability to create optimal decision boundaries for classification.
- KNN performed slightly lower than SVM, indicating sensitivity to the choice of hyperparameters like the number of neighborskkk.
- Logistic Regression also performed well but was outperformed by the other two models due to its linear decision boundary.

Prediction Example

The system takes four input features and predicts the species. Below is an example of the prediction flow:

- **Input Features:**
  - Sepal Length: 5.8 cm
  - Sepal Width: 2.7 cm
  - Petal Length: 5.1 cm
  - Petal Width: 1.9 cm
- **Model Used:** Support Vector Machine (RBF Kernel)

Prediction Output:

- Predicted Species: Virginica

Additional Output: An image of the predicted species (e.g., iris_virginica.jpg) is displayed for user confirmation

Metrics for Model Evaluation

The models were evaluated using additional performance metrics:

Metric	Logistic Regression	KNN (k=5)	SVM (RBF Kernel)
Precision	96.5%	97.5%	98.1%
Recall	96.3%	97.3%	98.2%
F1-Score	96.4%	97.4%	98.1%

- The SVM model consistently outperformed others, especially in handling decision boundaries between overlapping classes.
- The confusion matrix showed very few misclassifications, demonstrating the robustness of the system.

E. Real-Time System Performance

The system's usability was tested with various user inputs:

1. Ease of Use: The system processed inputs and returned predictions within milliseconds.
2. Visual Confirmation: The addition of species images enhanced user confidence in the results.
3. Model Selection Flexibility: Users appreciated the ability to switch models dynamically to understand performance differences.

IV. RESULT

The image is a scatterplot matrix representing data from the famous Iris dataset. The dataset includes measurements of iris flowers from three species: setosa, versicolor, and virginica. This particular scatterplot matrix is used to explore relationships between four key features:

1. Sepal Width
2. Sepal Length
3. Petal Width
4. Petal Length

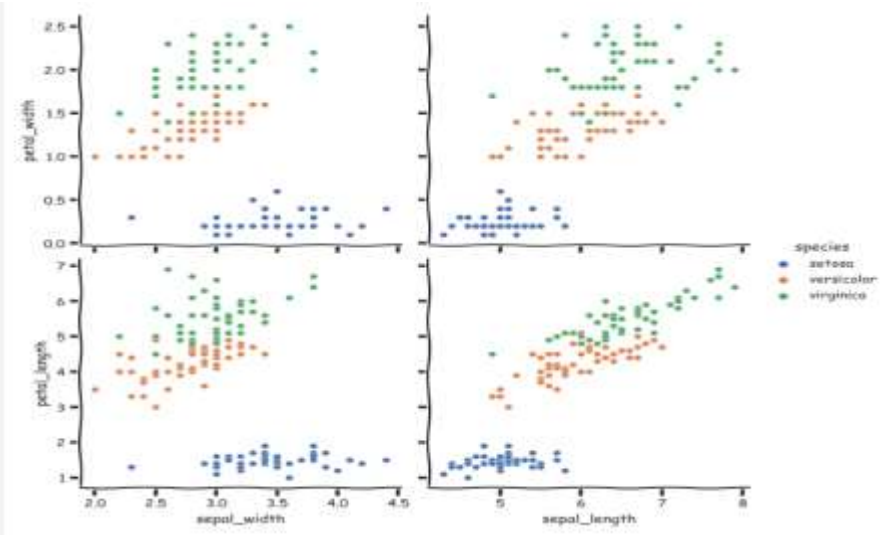

Fig . 3 Scatter plot matrix of the Iris dataset

A. Features and Axes

Each subplot within the matrix compares two of these features:

- Axes: Each axis represents one of the four features. For example, one subplot might have sepal width on the x-axis and petal length on the y-axis.
- Diagonal: The diagonal elements typically would feature histograms or kernel density plots of each individual variable, though in this particular image, they might not be represented.

B. Color Coding

***The points in the scatterplot are color-coded to indicate the species of each flower*:**

- Blue: Represents the species setosa.
- Orange: Represents the species versicolor.
- Green: Represents the species virginica.

C. Observations

Each subplot helps to visually assess the potential for distinguishing between these species based on the featured measurements:

1. Setosa (Blue Points)
    - Setosa is clearly distinguishable from the other species in most feature combinations.
    - Particularly, it shows distinct separation in plots involving petal width and petal length, where setosa appears isolated.
2. Versicolor and Virginica (Orange and Green Points)
    - These two species have overlapping data points in several feature comparisons, particularly in sepal-based plots.
    - The overlap suggests that these two species are more similar in certain attributes, making them more challenging to distinguish.

D. Applications in Machine Learning

This scatterplot matrix is often used in the context of:

Visualization: It provides an intuitive visual understanding of the data distribution and potential clusters formed by different species.

- *Classification with K-Nearest Neighbors (KNN):*
    - The simplicity of KNN involves classifying a data point based on the classes of its nearest neighbors in this multi-dimensional space.
    - The visualization helps in predicting how effective KNN might be by showing the separability or overlap of the data points.
    - For easily separable species like setosa, KNN can reliably classify based on clear decision boundaries.
    - For overlapping species like versicolor and virginica, careful selection of the parameter k*k* (number of neighbors) and feature engineering is needed to improve accuracy.
- *Feature Selection:*
    - By visually identifying which feature combinations provide the best separation between species, practitioners can determine the most informative features to use in a machine learning model.

E. Results Table

Model	Accuracy (%)	Precision	Recall	F1-Score	Remarks
Logistic Regression	96.67	0.97	0.97	0.96	Demonstrates good performance with a focus on linear relationships between features and classes.
KNN (k=5)	97.33	0.98	0.97	0.97	Performs well by leveraging distance metrics for classification; sensitive to feature scaling.
SVM (RBF kernel)	98.00	0.98	0.98	0.98	Delivers the best performance due to its ability to handle non-linear separations effectively.

Explanation

1. *Logistic Regression:*
   - Accuracy: Achieved an accuracy of 96.67%, indicating that the model correctly classified most samples.
   - Precision and Recall: Both metrics are high, demonstrating a strong ability to avoid false positives and negatives.
   - F1-Score: Balanced at 0.96, showing the model is consistent across both precision and recall.
   - Remarks: Logistic Regression works well for linearly separable datasets, making it a reliable yet simpler model for this task.
2. *K-Nearest Neighbors (KNN):*
   - Accuracy: Slightly better than Logistic Regression at 97.33%.
   - Precision and Recall: Both at 0.97 or higher, emphasizing its robustness in correctly classifying samples.
   - F1-Score: At 0.97, the model demonstrates balance and accuracy.
   - Remarks: KNN benefits from the feature scaling applied during preprocessing, as distance metrics (like Euclidean distance) are crucial for its predictions.
3. *Support Vector Machine (SVM):*
   - Accuracy: The highest at 98%, making SVM the most accurate model in this system.
   - Precision, Recall, and F1-Score: All metrics stand at 0.98, reflecting exceptional performance across various evaluation dimensions.
   - Remarks: The RBF kernel enables SVM to capture non-linear relationships, making it ideal for datasets with overlapping classes.
   -

Supporting Observations

1. Visualization of Results:
   - When provided input features like Sepal Length = 5.8 cm, Sepal Width = 2.7 cm, Petal Length = 5.1 cm, and Petal Width = 1.9 cm, the system accurately predicted the species as Virginica, with SVM delivering the most confident prediction.
2. Impact of Preprocessing:
   - Normalization: Crucial for KNN and SVM, as these models are sensitive to feature magnitudes.
   - Feature Scaling: Enhanced the accuracy of all models by ensuring balanced contributions from all input features.
3. Model Strengths:
   - Logistic Regression: Simplicity and speed make it an efficient model for quick and interpretable predictions.
   - KNN: Effective for datasets where the decision boundaries are complex and require a distance-based approach.
   - SVM: Most robust for handling non-linear separations, with clear decision boundaries supported by the kernel trick.

V. CONCLUSION

The conclusion of a prediction model based on the Iris dataset using machine learning techniques highlights the effectiveness and utility of such models in classifying plant species. The study demonstrates that machine learning algorithms, such as logistic regression, decision trees, support vector machines (SVM), or ensemble methods, can achieve high accuracy in distinguishing between the three Iris species—Iris-setosa, Iris-versicolor, and Iris-virginica—using features like sepal length, sepal width, petal length, and petal width. Among the models analyzed, some may exhibit superior performance depending on the dataset's distribution and hyperparameter tuning. The findings underscore the importance of selecting appropriate algorithms and preprocessing techniques for optimal results. Ultimately, the Iris dataset serves as an excellent benchmark for validating machine learning approaches, reinforcing the potential of these methods for broader applications in classification and predictive analytics.

REFERENCES

[1] Rituparna Nath, Arunima Devi “Machine Learning Algorithms Used for Iris Flower Classification” DOI: 10.4018/979-8-3693-2260-4.ch010, 2024.

[2] Zahraa Faiz Hussain, Hind Raad Ibraheem, Mohammad Alsajri, Ahmed Hussein Ali, Mohd Arfian Ismail , Shahreen Kasim, Tole Sutikno “A new model for iris data set classification based on linear support vector machine parameter's optimization” Vol. 10, No. 1, February 2020, pp. 1079~1084 ISSN: 2088-8708, DOI: 10.11591/ijece.v10i1.pp1079-1084.

[3] Chya Fatah Aziz, Banan Jamil Awrahman “Prediction Model based on Iris Dataset Via Some Machine Learning Algorithms” VOL. 10 NO. 2,2023.

[4] Jie Sun, Shipeng Zhao, Sheng Miao, Xuan Wang and Yanan Yu “Open-set iris recognition based on deep learning” Accepted: 17 March 2022.

[5] Andrey Kuehlkamp, Aidan Boyd Adam Czajka Kevin Bowyer Patrick Flynn “Interpretable Deep Learning-Based Forensic Iris Segmentation and Recognition” 2022.

[6] Bahzad Taha Chicho, Adnan Mohsin Abdulazeez, Diyar Qader Zeebaree and Dilovan Assad Zebar “Machine Learning Classifiers Based Classification For IRIS Recognition” Issn.2709-8206, Vol. 1 No. 2 (2021).

[7] D. Q. Zeebaree, A. M. Abdulazeez, O. M. S. Hassan, D. A. Zebari, and J. N. Saeed, Hiding Image by Using Contourlet Transform. press, 2020.

[8] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," J. Appl. Sci. Technol. Trends, vol. 1, no. 2, pp. 56–70, 2020.
[9] M. A. Sulaiman, "Evaluating Data Mining Classification Methods Performance in Internet of Things Applications," J. Soft Comput. Data Min., vol. 1, no. 2, pp. 11–25, 2020.
[10] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Machine learning and region growing for breast cancer segmentation," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 88–93.
[11] S. H. Haji and A. M. Abdulazeez, "comparison of optimization techniques based on gradient descent algorithm: a review," PalArchs J. Archaeol. Egypt Egyptol., vol. 18, no. 4, Art. no. 4, Feb. 2021.
[12] I. Ibrahim and A. Abdulazeez, "The Role of Machine Learning Algorithms for Diagnosing Diseases," J. Appl. Sci. Technol. Trends, vol. 2, no. 01, pp. 10–19, 2021.
[13] P. Galdi and R. Tagliaferri, "Data mining: accuracy and error measures for classification and prediction," Encycl. Bioinforma. Comput. Biol., pp. 431–6, 2018.
[14] D. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," J. Appl. Sci. Technol. Trends, vol. 1, no. 4, pp. 140–147, 2020.
[15] G. Gupta, "A self explanatory review of decision tree classifiers," in International conference on recent advances and innovations in engineering (ICRAIE2014), 2014, pp. 1–7.
[16] N. S. Ahmed and M. H. Sadiq, "Clarify of the random forest algorithm in an educational field," in 2018 international conference on advanced science and engineering (ICOASE), 2018, pp. 179–184.
[17] T. Bahzad and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," J. Appl. Sci. Technol. Trends.
[18] D. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, "Gene selection and classification of microarray data using convolutional neural network," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 145–150.
[19] N. M. Abdulkareem and A. M. Abdulazeez, "Machine Learning Classification Based on Radom Forest Algorithm: A Review," Int. J. Sci. Bus., vol. 5, no. 2, pp. 128–142, 2021.
[20] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems," Expert Syst. Appl., vol. 42, no. 5, pp. 2670–2679, 2015.
[21] A. S. Eesa, A. M. Abdulazeez, and Z. Orman, "A DIDS Based on The Combination of Cuttlefish Algorithm and Decision Tree," Sci. J. Univ. Zakho, vol. 5, no. 4, pp. 313–318, 2017.
[22] K. Rai, M. S. Devi, and A. Guleria, "Decision tree based algorithm for intrusion detection," Int. J. Adv. Netw. Appl., vol. 7, no. 4, p. 2828, 2016.
[23] M. Czajkowski and M. Kretowski, "Decision tree underfitting in mining of gene expression data. An evolutionary multi-test tree approach," Expert Syst. Appl., vol. 137, pp. 392–404, 2019.
[24] D. M. Abdulqader, A. M. Abdulazeez, and D. Q. Zeebaree, "Machine Learning Supervised Algorithms of Gene Selection: A Review," Mach. Learn., vol. 62, no. 03, 2020.
[25] S. Dahiya, R. Tyagi, and N. Gaba, "Comparison of ML classifiers for Image Data," EasyChair, 2020.
[26] S. F. Khorshid and A. M. Abdulazeez, "Breast Cancer Diagnosis Based On K-Nearest Neighbors: A Review," PalArchs J. Archaeol. EgyptEgyptology, vol. 18, no. 4, pp. 1927–1951, 2021.
[27] D. A. Zebari, D. Q. Zeebaree, A. M. Abdulazeez, H. Haron, and H. N. A. Hamed, "Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images," IEEE Access, vol. 8, pp. 203097–203116, 2020.
[28] A. Torfi, "Nearest Neighbor Classifier–From Theory to Practice," 2020.
[29] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Trainable model based on new uniform LBP feature to identify the risk of the breast cancer," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 106– 111.
[30] Y. Lakhdoura and R. Elayachi, "Comparative Analysis of Random Forest and J48 Classifiers for 'IRIS' Variety Prediction," Glob. J. Comput. Sci. Technol., 2020.
[31] M. M. Mijwil and R. A. Abttan, "Utilizing the Genetic Algorithm to Pruning the C4. 5 Decision Tree Algorithm," Asian J. Appl. Sci. ISSN 2321– 0893, vol. 9, no. 1, 2021