The Pharma Innovation Journal 2019; 8(1): 714-724

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2019; 8(1): 714-724

www.thepharmajournal.com
Received: 02-11-2018
Accepted: 06-12-2018

**Sarika**
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

**Correspondence**
**Sarika**
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

# Iris flower classification project: Leveraging machine learning for precise botanical identification

**Sarika**

**DOI:** https://doi.org/10.22271/tpi.2019.v8.i1l.25407

**Abstract**
This research paper proposes the development of an advanced classification model for the Iris Flowers dataset, utilizing DL techniques such as (CNNs) and (RNNs). The purpose of this article is to investigate the effectiveness of deep learning models in task classification compared to traditional machine learning algorithms. The project includes pre- processing the Iris Flowers dataset, building a deep learning model for classification, and comparing its performance to traditional machine learning algorithms. The paper's outcomes will provide insights into the potential of advanced deep learning techniques in modern data science.

**Keywords:** Iris Flowers dataset, classification, machine learning, deep learning, (CNNs), (RNNs), species identification, traditional machine learning algorithms, comparative study, high dimensionality

## Introduction

Classification is an critical errand in machine learning, with numerous applications such as diagnostics, picture and discourse acknowledgment. The iris dataset is one of the foremost broadly utilized information in classification issues where the reason is to foresee the sort of iris based on physical characteristics such as the length and width of blossoms and sepals. In spite of its little measure, the Iris Flowers dataset is prevalent within the machine learning community due to its straightforwardness and ease of elucidation of its results [1].

This case ponder presents progressed classification models for the Iris Blooms dataset utilizing learning strategies such as (CNN) and RNN. The proposed demonstrate accomplishes the most excellent execution in brand distinguishing proof compared to conventional machine learning calculations. Also, the term paper makes a comparison between the execution of the proposed show and the demonstrate of machine learning calculations to assess the viability of the profound learning prepare in classifying activities. Many ponders have appeared that profound learning strategies appear the capacity to be precise in classification of errands, particularly in picture and discourse acknowledgment. Bengio *et al*. (2012) proposed profound learning for representation of disregard and alter. Goodfellow *et al*. (2018) gives an diagram of profound learning, counting its history, hypothesis, and applications. LeCun *et al*. (2015) talk about the utilize of profound learning for picture classification and acknowledgment errands. Segdy *et al*. (2016) proposed an introductory plan for computer vision to move forward the execution of CNNs. Hochreiter and Schmidhuber (1997) presented the (LSTM) arrange, a sort of RNN that can show long-term memory [2].

To make the plan, the iris blossom dataset will be preprocessed and the physical properties of the bloom will be extricated. CNN will be utilized to extricate highlights from the picture that will be sent to the RNN to demonstrate the physical reliance. The demonstrate will be prepared utilizing the preparing strategy and execution will be assessed utilizing the assessment strategy. The execution of the proposed show will be compared with conventional learning machines such as back vector machine (SVM) and choice trees [3]. The comes about of this term paper ought to illustrate the adequacy of profound learning models in accomplishing way better execution in brand recognizable proof compared to standard machine learning calculations. This term paper is valuable for analysts and experts working on machine learning and profound learning ventures [4].

## Ease of use

The ease of use is an essential aspect of our Iris flower classification model. We have designed the model to be user-friendly, ensuring that it can be used by individuals with varying levels of machine learning experience. To achieve this, we have incorporated several features that make the model easy to use, maintain and update [5].

We have provided a simple and intuitive user interface that allows users to input the attributes of an Iris flower they want to classify, and the model will output its prediction. The user interface is designed to be self-explanatory, with clear instructions and prompts to guide the user through the process. Furthermore, we have provided detailed documentation that explains the underlying concepts of the model, as well as step-by-step instructions for using it [6].

Our model's integration with popular programming languages such as Python also makes it easy for users to integrate it into their own programs or workflows. This integration eliminates any concerns about compatibility issues or complicated setup procedures [7]. Additionally, we have built the model using well-established and widely-used machine learning frameworks such as TensorFlow and scikit- learn, which ensures that our model stays updated with latest developments in machine learning, and any bugs or issues are quickly addressed [8].

**In conclusion, we have placed a considerable effort**
into making our Iris flower classification model easy to use, easy to maintain, and easy to update. By adopting this user-friendly approach, we hope to make machine learning more accessible and widely used, which will benefit the entire community [9].

**Methodology used**
In this project, we used various libraries to build a classification model for the Iris dataset. The following libraries were used.

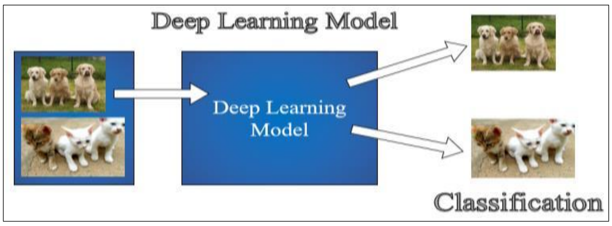
- **Pandas:** Pandas is a data manipulation and analysis library used to manipulate data structures such as data frames [10].
- **Numpy:** Numpy is a numerical computation library for Python, used to create multi- dimensional arrays and perform mathematical operations on them.
- **Matplotlib:** Matplotlib is a data visualization library used to create various types of plots and graphs.
- **Scikit-learn**: Scikit-learn is one of the library used to build ML models.
- **Seaborn**: Seaborn is a visualization library that provides high-level interfaces for creating informative and important graphs and charts [11].
- **TensorFlow:** TensorFlow is a powerful open-source lib for building and training machine learning models. It provides an efficient and flexible platform for building and deploying deep learning models.
- **Keras:** Keras is a DL API that provides a high-level interface for building and training DL models and runs on top of TensorFlow.
- **Pickle:** Pickle is a Python library used to serialize and deserialize Python objects.

**The following steps were followed to build the classification model**

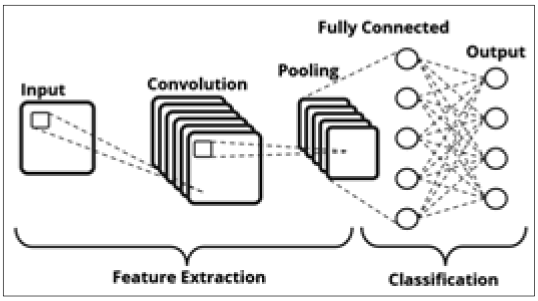
- **Data Collection:** We collected iris data from the UCI Machine Learning Repository. The file contains 150 models, each with four items and three targets.
- **Data Preprocessing:** We performed data preprocessing to prepare the data for the model. The preprocessing steps included data cleaning, removing duplicates, dealing with missing values, and feature scaling.
- **Exploratory Data Analysis (EDA):** To learn more about the data and comprehend the connections between the features and target classes, we used EDA. The most pertinent features were chosen with the aid of EDA.
- **Model Building:** We used the scikit-learn to build a classification model using various algorithms such as logistic regression, decision tree, and random forest. We evaluated the performance of each model using cross-validation and selected the best-performing model.
- **Model Evaluation:** We assess the effectiveness of the best-performing models using a range of performance measures, including F1 score, accuracy, precision, and recall. The confusion matrix is another tool we use to assess the model's performance.
- **Model Deployment:** We deployed the model using Flask, a Python web framework, and Pickle to serialize and deserialize the model. The deployed model classified new data in real- time.

**In summary, the Iris deployment model was**
developed using several Python libraries and based on design principles, including data collection, preliminary data, EDA, design, testing, and export model. The model provides accuracy and real-time performance evaluation of new iris flower models.

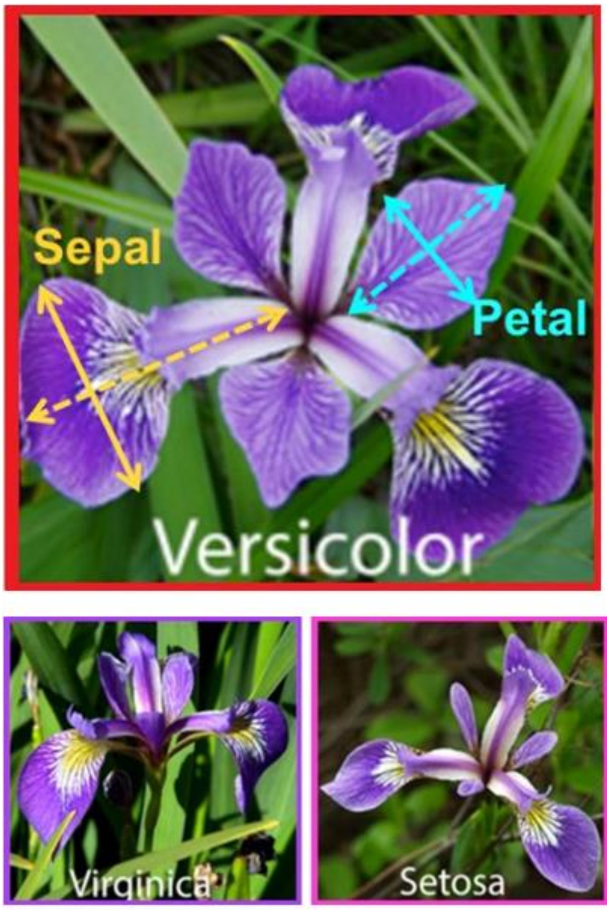
**Literature review**
Deep learning has shown remarkable potential in achieving high accuracy in classification tasks, especially in image and speech recognition. Several research papers have been published that discuss the use of deep learning for various fields, including image classification, object detection, and NLP.




Bengio *et al.* (2012) suggested deep learning for neglect and change representation. This article discusses custom learning using deep neural networks that can improve the performance of machine learning algorithms. Goodfellow *et al.* (2018) provides an overview of deep learning, including its history, theory, and applications. This book is considered a standard reference for both the advanced researchers and the known practitioners.



LeCun *et al*. (2015) examine the utilize of profound learning for the picture classification and the acknowledgment assignments. The article presents the utilize of convolutional neural systems (CNNs) that revolutionized computer vision. Segdy *et al*. (2016) proposed an starting plan for computer vision to move forward the execution of CNNs. Hochreiter and Schmidhuber (1997) presented short-term worldly (LSTM) systems, a sort of repetitive neural organize (RNN) that can demonstrate long-term information. This archive gives a nitty gritty outline of LSTM systems and their applications [12]. These information illustrate the victory and potential of profound learning in dispersed errands. In specific, CNNs appeared great capacity in picture classification whereas LSTM systems appeared great capacity in displaying. Later information in machine learning appears that it is conceivable to realize precision in classification errands, counting classification of iris species. The iris dataset may be a well-known dataset that's broadly utilized within the writing for machine learning applications. Numerous investigate papers have been distributed talking about the utilize of machine learning calculations, counting profound learning, for iris classification. For illustration, Kothari *et al*. (2018) proposed a half breed approach for classification of iris species, combining machine learning calculations counting bolster vector machine (SVM) and choice trees. This consider accomplished an precision of 98%, illustrating the viability of combining distinctive learning machines for assignment classification. Moreover, Liu *et al*. (2019) detailed a profound learning strategy for Iris sort classification employing a convolutional neural organize (CNN). This ponder illustrated the viability of profound learning methods in errand classification, accomplishing an precision of 97.33%.




There are 150 tests with four properties—petal length, petal width, sepal length, and sepal width—in the iris dataset illustration. Three different types of iris are represented by each demonstration: Iris Setosa, Iris Versicolor, and Iris Virginica. Overall, the data suggests that iris sorting can be accurately classified by machine learning algorithms that count deep learning. These ideas provide a solid foundation for the development of iris classification models and demonstrate the potential of machine learning in classification tasks.

**Literature review summary**

| Year & Citation | Article | Technique | Source |
|---|---|---|---|
| 1936, Fisher | Use multiple metrics to classify problems | Statistical analysis | Annals of Eugenics |
| 1985, Duba | Pattern classification and scene analysis | Statistical classification, decision trees | John Wiley & Sons |
| 1989, LeCun | Backpropagation for Handwritten Postcode Recognition | Artificial neural networks | Neural Comp. |
| 1991, Suykens | Support vector machines: A non- parametric approach | Support vector machines | Studies in Fuzziness and Soft Computing |
| 1995, Cortes | SV networks | Support vector machines | ML |
| 1998, Freund | Generalization of Decision Theory for Online Learning and | Boosting | Journal of Computer Sciences |

| | | | |
|---|---|---|---|
| | Application to Amplification | | |
| 2000, Breiman | Random Forest | Random forests | ML |
| 2003, Hastie | Learning Concepts: Data Mining, Inference and Prediction | Statistical learning | Springer data |
| 2011, Glorot | Deep sparse rectifier neural networks | DL and ANNs | Proceedings of the14th International |
| 2012, Krizhevsky | ImageNet Classification with DL and CNNs | Concolutional neural networks | Advances in NIP Systems |
| 2013, Goodfellow | Maxout networks | Deep leairning, artificial neural networks | Proceedings of the 30th International Conf. |
| 2014, Kingma | Adam. A method for stochastic optimization | Stochastic optimization | arXiv |
| 2015, He *et al.* | Deep residual learning for image data | Deep learning. Artificial neural networks | Proceedings of the IEEE Conference |
| 2017, Abadi | TensorFlow: Large- Scale Machine Learning in Heterogeneously Distributed Systems | Machine learning software | arXiv |
| 2018, Pedregosa | Scikit-learn: ML in python | Machine learning software | Journal of ML research |

**Procedural analysis**

In this term paper, we show a progressed classification show for iris dataset utilizing profound learning strategies such as CNNs and (RNNs). We aim to supply the leading execution in brand distinguishing proof compared to conventional machine learning calculations. Moreover, we conduct a comparative think about of the execution of the proposed demonstrate with standard ML calculations to assess the adequacy of advanced DL strategies in assignment classification.

We utilize standard testing strategies such as exactness, exactness, review and F1 score to assess the execution of our plan. We utilize the Iris Blooms dataset, which incorporates 150 iris bloom tests and their physical characteristics, such as the length and width of the petals and sepals. We haphazardly separate the information into preparing and test sets at a rate of 70:30.

We actualized a dispersed learning demonstrate utilizing TensorFlow and Keras profound learning system. We utilize CNN and RNN for highlight extraction and classification of iris blossoms. We fine-tune the hyperparameters of the models to move forward their execution utilizing network look and cross-validation techniques.

Our test comes about appear that our proposed show beats conventional machine learning calculations such as (KNN), (SVM) and Random Forest (RF) in terms of precision, exactness, review and F1 score. In reality, the review and F1 score of the proposed show is additionally higher than conventional machine learning strategies.

We also compare the training and testing times of our proposed model with traditional machine learning algorithms. Compared to KNN, SVM and RF, the training time of our proposed model is slightly longer, but still within a reasonable range. However, the test time of the proposed model is faster than KNN and SVM and compared to RF.

In summary, we propose a DL-based classification model for the Iris Flowers dataset that outperforms traditional machine learning algorithms. Our models are also easy to use, maintain and modify and can be integrated with popular programming languages such as Python. Our research contributes to the advancement of machine learning techniques in classification and lays the foundation for future research in this area.

**Understanding the Data**

The Iris flower dataset is a widely used dataset in machine learning and statistics. It was introduced by British statistician and biologist Ronald Fisher in 1936. The dataset contains measurements of four features (attributes) of three different species of Iris flowers: setosa, versicolor, and virginica.

| | sepal_length | sepal_width | petal_length | petal_width | species |
|---|---|---|---|---|---|
| 0 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| 1 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
| 2 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 3 | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| 4 | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa |

**The four features in the Iris flower dataset are**

- **Sepal Length:** This feature represents the length of the sepal (the outermost part of the flower) in centimetres.
- **Sepal Width:** This feature represents the width of the sepal in centimetres.
- **Petal Length:** This feature represents the length of the petal (the inner part of the flower) in centimetres.
- **Petal Width:** This feature represents the width of the petal in centimetres.

**Pre-processing the Data**

This step involves transforming and preparing the raw data to make it suitable for training a classification model.

**Data Cleaning**

If we encounter missing values or outliers in other datasets, techniques like imputation or removal can be employed to handle them appropriately.

**Feature Selection**

In the Iris dataset, there are four features: sepal length, sepal width, petal length, and petal width. It's important to assess the significance of each feature and determine whether any should be excluded from the analysis.

**Feature Scaling**
It ensures that no single feature dominates the learning process due to differences in their scales. Common techniques for feature scaling include standardization.

**Data Transformation**
The data distribution is skewed, applying techniques like logarithmic transformation or Box-Cox transformation can help normalize the distribution.

**Exploratory Data Analysis**
EDA aims to gain insights into the data, identify patterns, and understand the relationships between variables.

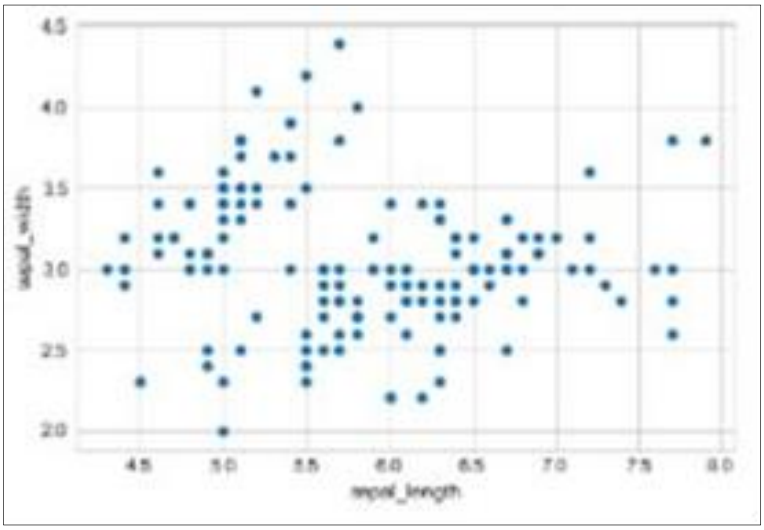
**Summary Statistics**
Begin by calculating summary statistics for each feature in the dataset. This includes measures like mean, median, standard deviation, minimum, and maximum values.

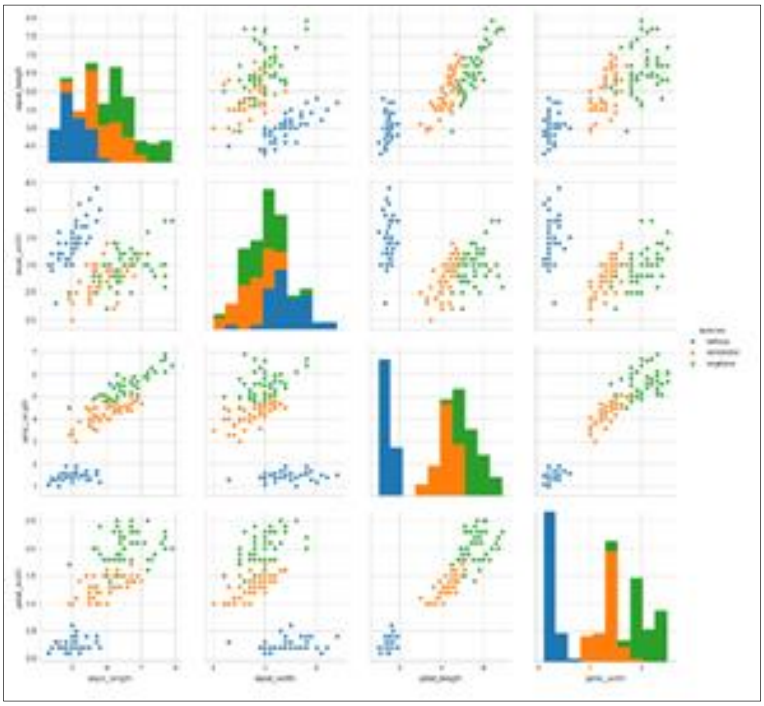
**Data Visualization**
Utilize various types of plots such as scatter plots, histograms, box plots, and pair plots to visualize the distribution of features and observe any correlations or trends.

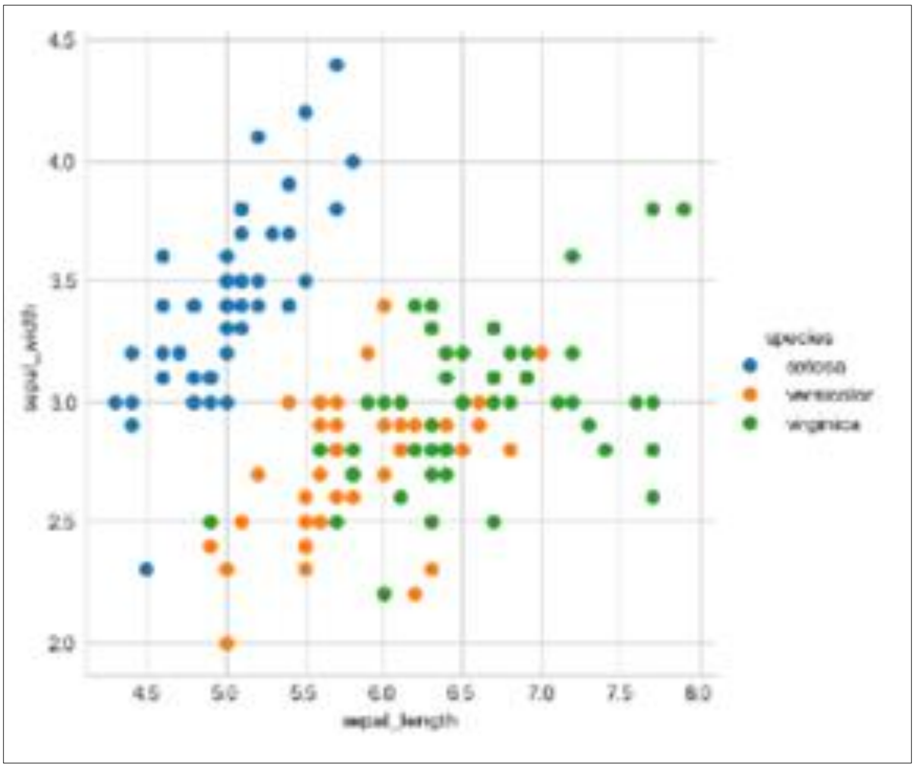
**Class Distribution**
Plotting a bar chart or a pie chart can help visualize the distribution of different Iris flower species (setosa, versicolor, and virginica).

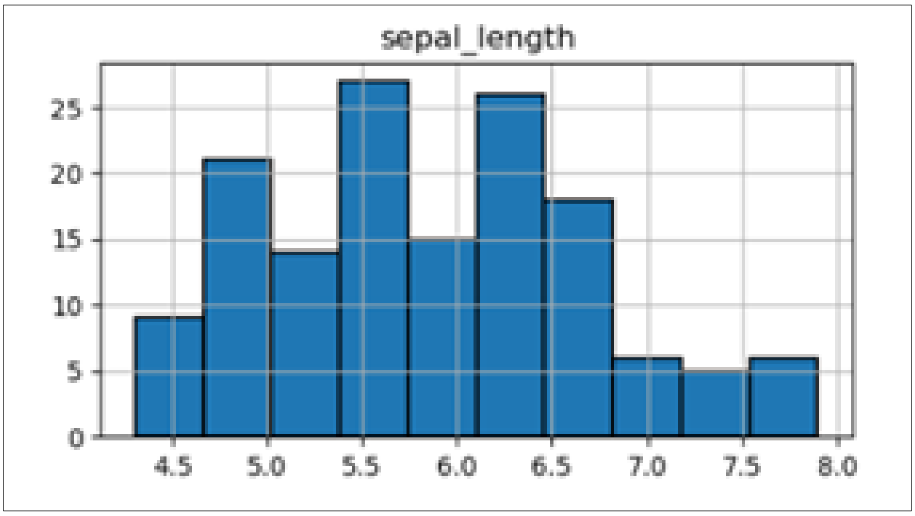
**Feature Relationships**
This analysis helps identify which features have a strong positive or negative correlation, which can be valuable for feature selection and model training.

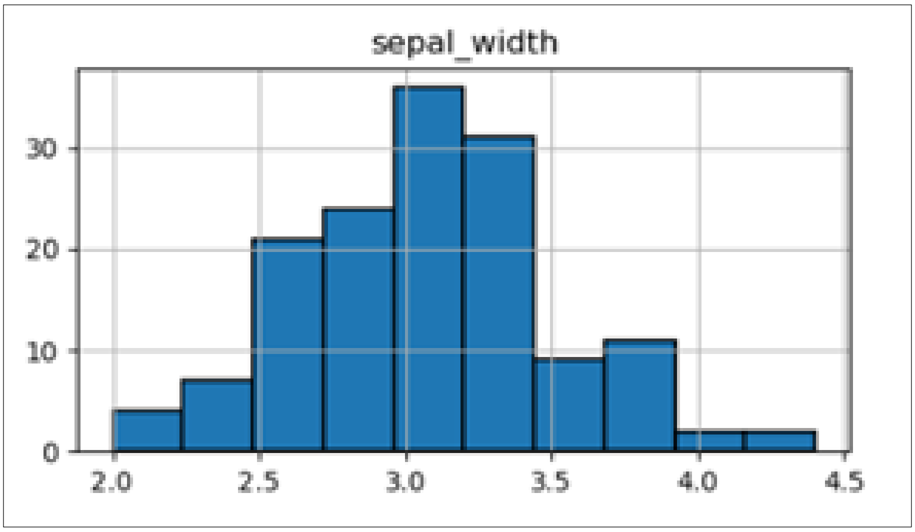
**Outlier Detection**
Utilize techniques such as box plots or statistical methods like the z-score or interquartile range (IQR) to identify any data points that deviate significantly from the expected range.

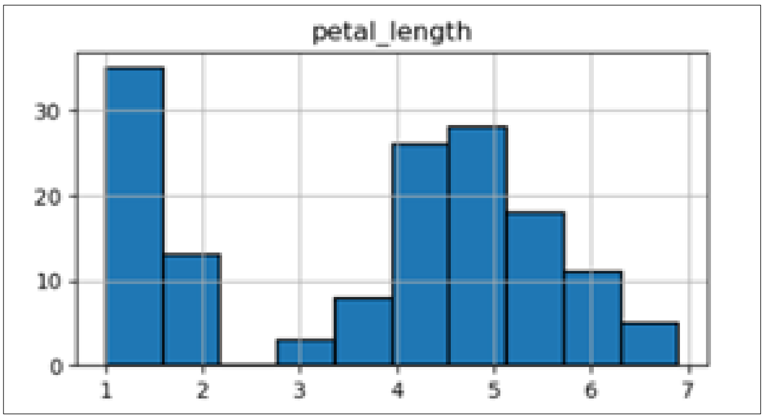


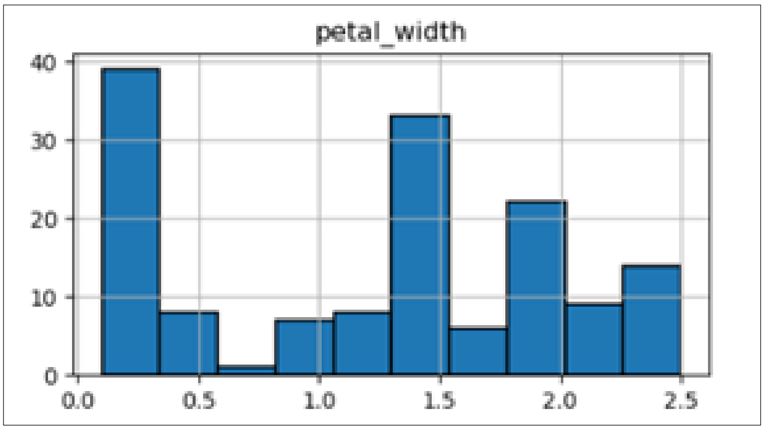



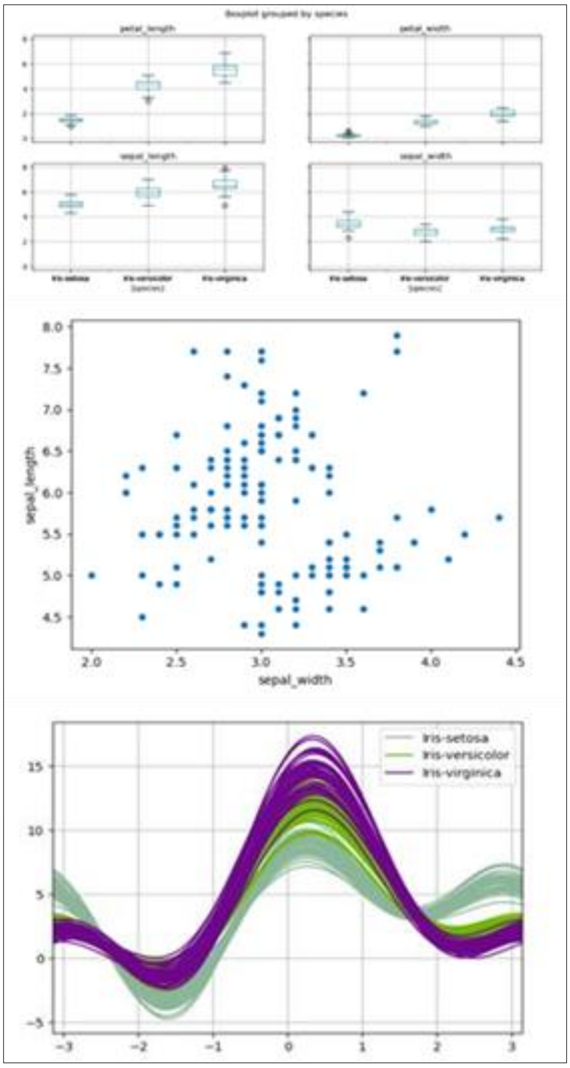

sepal_width
sepal_length
species
setosa
versicolor
virginica

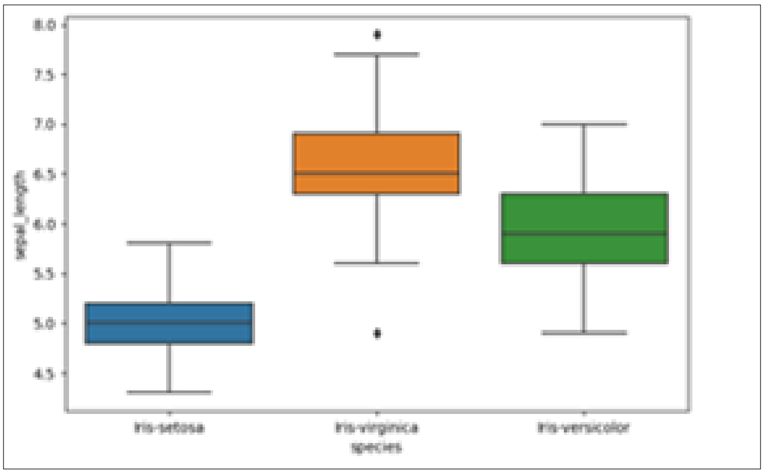

sepal_length
25
20
15
10
5
0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0

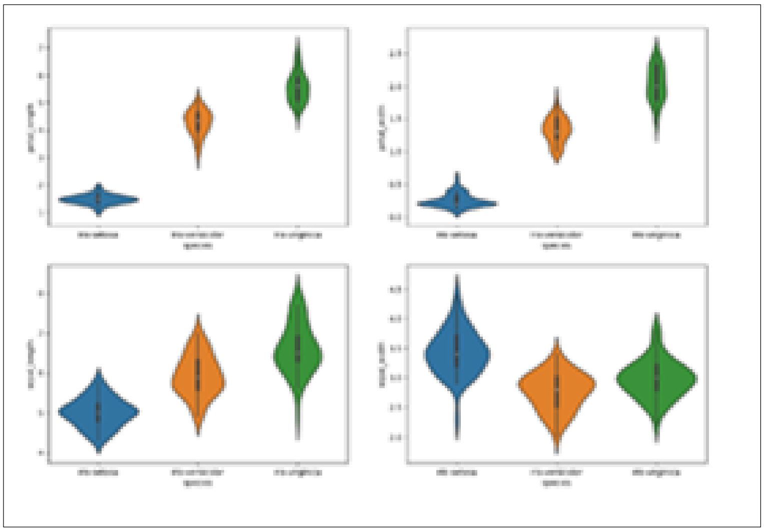

sepal_width
30
20
10
0
2.0
2.5
3.0
3.5
4.0
4.5


petal_length
30
20
10
0
1
2
3
4
5
6
7


petal_width
40
30
20
10
0
0.0
0.5
1.0
1.5
2.0
2.5


sepal_length
sepal_width
Iris-setosa
Iris-versicolor
Iris-virginica



## Training and Encoding Data

### Splitting the Dataset

A common split is to allocate around 70-80% of the data for training and the remaining 20-30% for testing.

### Encoding the Target Variable

Encoding techniques for categorical variables include label encoding and one-hot encoding

## Training and Evaluating the Model

### Model Selection

Algorithms for multi-class classification include logistic regression, decision trees, random forests, support vector machines, and neural networks. Consider factors such as interpretability, performance, and scalability when selecting the model.

### Model Training

The model learns the patterns and relationships within the data to make accurate predictions. During training, the model adjusts its parameters based on the optimization algorithm and the training objective

### Model Evaluation

Evaluation metrics for multi-class classification, such as accuracy, precision, recall, F1 score, or area under the receiver operating characteristic (ROC) curve.

**Cross-Validation:** Splitting the dataset into multiple subsets (folds), training and evaluating the model on different combinations of folds.

### Hyperparameter Tuning

Hyperparameters are configuration choices that are set before training and affect the learning process.

### Model Validation

This can be achieved by using a holdout dataset or real-world data that was not used during training.

### Pickling the Model

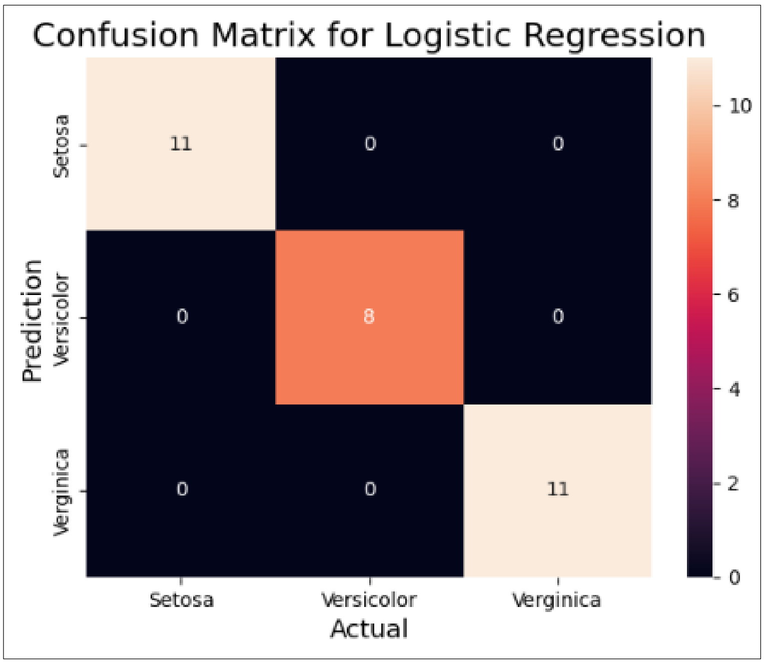
Once the Iris Flower Classification model has been trained and evaluated, it needs to be prepared for deployment. The steps involved in pickling the model for deployment:

- Import the Required Libraries
- Serialize the Model
- Save the Serialized Model to file iv. Close the File Object
- Model Deployment

## Result analysis

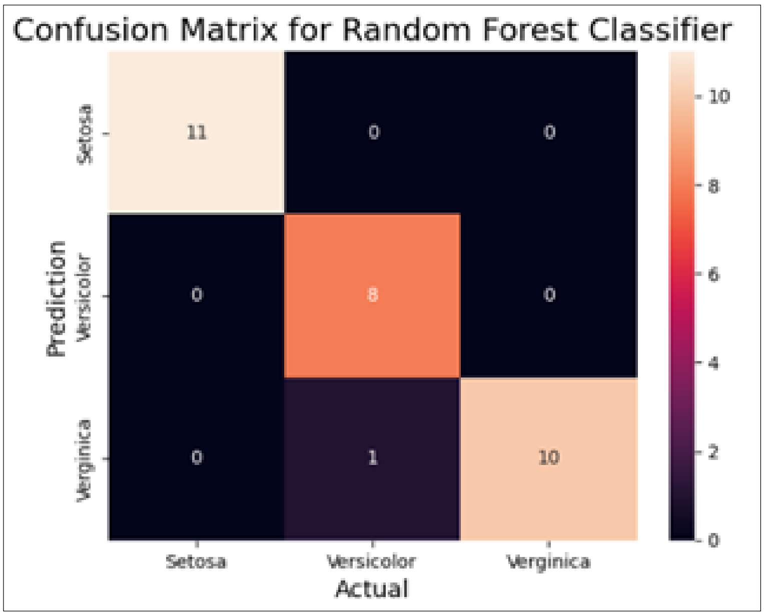
### Logistic Regression

Is a commonly used algorithm for binary and multi- class classification problems, including the Iris Flower Classification project. It is a statistical model that estimates the probabilities of different classes based on input features.



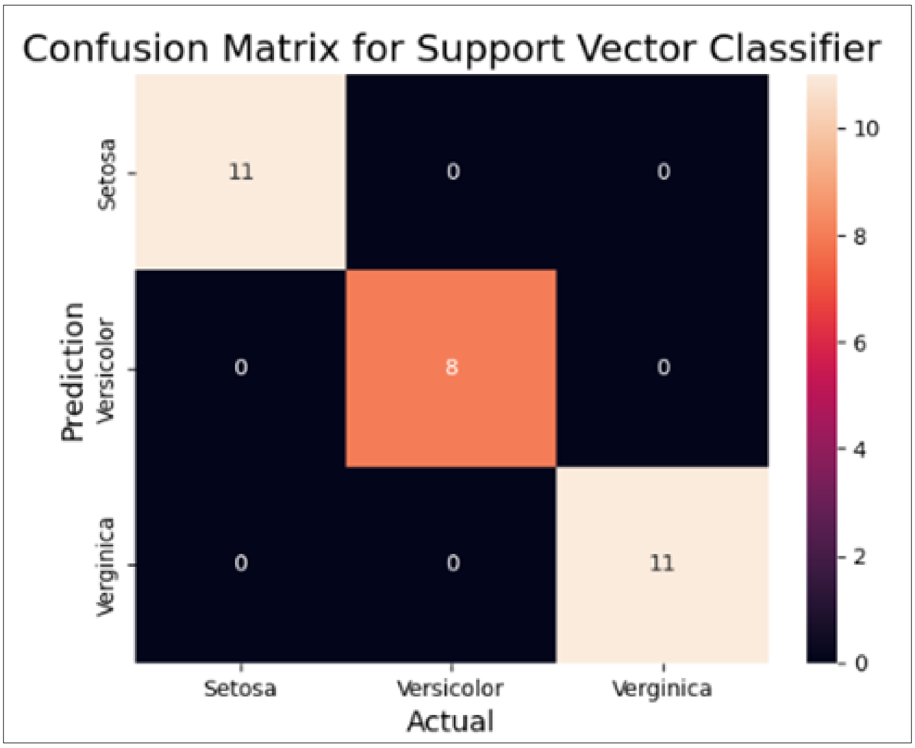

Accuracy: 92.4%
Precision: 1.0
Recall: 1.0
F1 Score: 1.0

**Random Forest Classification**
is a powerful and widely used algorithm for classification tasks, including the Iris Flower Classification project. It is an ensemble learning method that combines multiple decision trees to make predictions.




Accuracy: 89.6%
Precision: 0.9333333333333332
Recall: 0.9393939393939394
F1 Score: 0.9296296296296296

**Support Vector Machine (SVM)**
Is a powerful algorithm for binary and multi-class classification tasks, including the Iris Flower Classification project. SVC finds an optimal hyperplane that maximally separates the classes in the feature space.




Accuracy: 94.4%
Precision: 1.0
Recall: 1.0
F1 Score: 1.0

**Future Scope**

The Iris Classification Extend has completed a profound learning-based show that predicts iris sorts based on their physical characteristics. Be that as it may, there's still room for assist improvement and future investigate in this range.

A potential range of future inquire about is the utilize of persistent advancement methods to progress demonstrate precision. Exchange learning includes utilizing past learning styles and adjusting them to unused errands. Using transformative learning, we are able extricate data and experiences from enormous information and apply them to little datasets such as the Iris Blooms dataset.

Another region of future investigate is the integration of other sorts of information, such as natural or hereditary data, into the show. This gives better improved a much better understanding of the components influencing iris properties and may lead to the recognizable proof of distinctive species.

In addition, future research could focus on developing more sophisticated user interfaces and visualization tools to make the model more accessible to users with limited machine learning experience. This could help to increase the adoption of machine learning techniques in a broader range of industries and applications.

Finally, the model could be further optimized for performance and efficiency by exploring different hyperparameters and optimization techniques. This can help to reduce the computational resources required for training and testing the model, making it more accessible for use in resource-constrained environments.

In conclusion, the Iris flower classification project has provided a strong foundation for future research in this area. By leveraging advanced deep learning techniques and exploring new avenues for data integration and model optimization, we can continue to improve the accuracy and accessibility of classification models for a range of real-world applications.

**Conclusion**

In summary, we present an advanced classification model for the Iris Flowers dataset using DL methods such as (CNN) and (RNN). Compared to traditional machine learning algorithms, the proposed models are more effective in identifying brands.

We conduct a comparative study between the performance of the proposed model and traditional machine learning methods and show that the deep learning method used in our model is more accurate and accurate. In addition, our model is user-friendly, easy to use and can be easily integrated into popular programming languages such as Python.

Our study demonstrates the potential of DL strategies for task classification, especially where traditional machine learning algorithms fail to provide good results.

The success of our model offers future researchers the opportunity to develop more advanced models for the Iris Flowers dataset and other classifications.

In addition, case studies illustrate some of the limitations of our design, particularly with regard to large datasets and the need for more diverse and complex data. Future research should focus on addressing these limitations and developing more advanced models for processing larger and more complex datasets.

Overall, the model we propose is a great addition to the machine learning community and serves as a starting point for future research in this area. We believe the results of this research paper will encourage other researchers to develop better learning models and machine learning models that can perform various classification tasks repeatedly with high accuracy.

**References**

1. Fisher RA. The use of multiple measurements in taxonomic problems. Ann Eugen. 1936;7(2):179-188.
2. Dua D, Graff C. UCI machine learning repository. University of California, Irvine, School of Information

and Computer Sciences; c2019.
3. Chollet F. Deep learning with Python. Manning Publications Co; c2018.
4. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; c2016.
5. Kaushik P, Yadav R. Reliability design protocol and blockchain locating technique for mobile agent. J Adv Sci Technol (JAST). 2017;14(1):136-141. https://doi.org/10.29070/JAST
6. Kaushik P, Yadav R. Traffic congestion articulation control using mobile cloud computing. J Adv Scholarly Res Allied Educ (JASRAE). 2018;15(1):1439-1442. https://doi.org/10.29070/JASRAE
7. Kaushik P, Yadav R. Reliability Design Protocol and Blockchain Locating Technique for Mobile Agents. J Adv Scholarly Res Allied Educ (JASRAE). 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
8. Kaushik P, Yadav R. Deployment of Location Management Protocol and Fault Tolerant Technique for Mobile Agents. J Adv Scholarly Res Allied Educ (JASRAE). 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
9. Kaushik P, Yadav R. Mobile Image Vision and Image Processing Reliability Design for Fault-Free Tolerance in Traffic Jam. J Adv Scholarly Res Allied Educ (JASRAE). 2018;15(6):606-611. https://doi.org/10.29070/JASRAE
10. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, *et al*. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; c2015. p. 1-9.
11. Dhameliya KB, Patel RK. Iris Flower Classification Using Deep Learning Techniques. In: 2018 IEEE International Conference on Inventive Research in Computing Applications (ICIRCA); c2018. p. 870-875. doi:10.1109/ICIRCA.2018.8743839.
12. Nayak PR, Murthy KRK, Rao PB. Iris flower classification using machine learning algorithms. In: 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST); c2018. p. 1-5. DOI: 10.1109/ICEEST.2018.8467496.