Comparative study of Jrip j48 and naive bayes algorithm in Flower specie prediction

Aakanksha Jain¹, Shivam Tiwari², Dr. M. Madhusudhana subramanyam³, Khalid Jamal⁴, Ajay Singh Yadav⁵ Sarvesh Kumar⁶*

¹Department of Cloud Technology & Information Security, Poornima University, Rajasthan, Jaipur, India
² Department of Cloud Technology & Information Security, Integral University, Lucknow, Uttar Pradesh
3 Department of Computer Science, College of Arts & Science, King Khalid University, Dhahran Al Janoub, Kingdom of Saudi Arabia
4 Department of Cloud Technology & Information Security, Integral University, Lucknow, Uttar Pradesh
5 Department of Mathematics, SRM Institute of Science and Technology, NCR Campus
6 Department of Cloud Technology & Information Security, Integral University, Lucknow, Uttar Pradesh

¹jain1994aakanksha@gmail.com
²shivam.tiwari.info@gmail.com
³mmsnaidu@yahoo.com
⁴khalid02feb@gmail.com
⁵ajay29011984@gmail.com
⁶sarvi899@gmail.com

*Corresponding author email: sarvi899@gmail.com

Abstract: Artificial intelligence and Deep learning techniques propose and provide effective mechanism for classification among several commodities like gender classification on basis of ridge count, spam mail classification and detection etc. Here we have proposed a hybrid module on basis of available AI and ML techniques by which we can achieve more than 90% accuracy for any provided test dataset. We have included over 50423 samples, out of which we use 66 percentage of data for training purpose of our model and 34% remaining sample we use as test dataset, with 8 fold we have achieve approximately 96% accuracy. For collection of dataset and processing of dataset we gone through several phases which include extraction of feature (EoF) using feature extraction technique, for cleaning of dataset we have use dimension reduction technique factor analysis(FA). In next phase for classification of flowers we have used classification techniques in which we found the accuracy score of j48 classifier is higher than Naïve Bayes and Jrip algorithm. Hence the choice of j48 classifier for classification is right.

Keywords: Bayes-classifier, Naïve-Bayes, Bayes-Net, Flower-Classifications, J48, EoF

1. Introduction: Now a day’s image classification is rapidly increasing its area of being in interest of data analyst and data scientists and the classification results and accuracy of results is giving effective mode of
prediction in various fields like healthcare, traffic monitoring, agriculture etc. When we talk about Google photos, it can be considered as best example of image processing and classification as it correctly classify and group photos of person with higher accuracy along with that it take suggestion from user about similar images. With Google photos we can easily check our images with additional search(Triguero & Vens, 2016, p. 171) (Mukane, 2013, p. 83).

In this world over millions of species of flower exist if we wanted to make a model by which we can classify the species of flower with accuracy then AI and ML algorithms comes into picture, as the algorithm with its ability to classify a class or specie of flower with provided details about that flower leads to produce results of higher accuracy.

Bloom order frameworks(Abdul Aziz et al., 2019a, p. 25) (Liu, 2014, p. 285) have a place with a shape coordinating issue, which is a major issue in PC vision and example acknowledgment. It characterized as the foundation of a similitude measure among shapes and its utilization for shape examination (Mukane, 2013, p. 83) (Abdul Aziz et al., 2019b, p. 30). The significant hypothetical fascinating inspiration for picture grouping comes from that shape coordinating is instinctively precise for people and need an associate which isn’t fathomed at this point in its full consensus. Shape coordinating framework incorporates object acknowledgment and location, picture attachment, and substance based recovery of pictures (Mukane, 2013, p. 83).

2. Algorithm Used:

2.1 J48: J48 algorithm comes under Weka platform in a inbuilt manner which is basically implementation of C4.5 algorithm which generally used for making decision in a hierarchical manner, are known as decision trees, and when we combine two or more algorithm to make decisions then that combination is known as random forest(Koul & Singhania, 2020, p. 1726). C4.5 algorithm was a advancement of original ID.3 algorithm whose limitations later on overcome in C4.5 algorithm. Whatever decisions made by these algorithms (in a hierarchy) (Mukane, 2013, p. 83) (Koul & Singhania, 2020, p. 1726) are generally used for classification purpose. Here in our proposed model we want to classify the species of provided bunch of dataset of flowers so choosing this algorithm was wise option [28 - 35].

2.2 Naïve Bayes: Another algorithm which is popular in term of statistics and widely used for classification purpose is Naïve Bayes Koul, S., & Singhania, U. (2020). This algorithm is a implementation of original Bayes’ theorem (S.Sarkate & B. Khanale, 2014, p. 16) with assuming assumption among features of provided or selected dataset. In statistics Naïve Bayes classifier known as ‘probabilistic classifier’. The assumptions which are made while implementations of naïve bayes(S.Sarkate & B. Khanale, 2014, p. 16) consider the independency among features that mean the presence of a feature in class in not related to presence of other feature Patel, I., Patel, S., & Patel, A. (2018). (Wäldchen et al., 2018). Naïve bayes is popular because of easy implementation and ability to perform well and give considerable accuracy level on high sophisticated datasets as well(Chitra et al., 2016, p. 18) (Pornpanomchai et al., 2011, p. 350)

2.3 Jrip: JRIP once in a while called as RIPPER is one of well-known classifier calculation (Pornpanomchaisri et al., 2011, p. 350)(Sai Kumar et al., 2011, p. 49). In JRIP examples of the dataset are assessed in expanding request, for given dataset of danger a bunch of rules are produced. JRIP (RIPPER) (Mohanty & Bag, 2017, p. 520) calculation treats each dataset of given information base and produces a bunch of rules including all the qualities of the class. At that point next class will get assessed and does likewise measure as past class, this cycle proceeds until all the classes have been covered(Najjar & Zagrouba, 2016, p. 53023).
3. Related work:

Saitoh and Kaneko proposed a programmed acknowledgment framework (Wäldchen et al., 2018) for wild blossoms. They utilize both the bloom and leaves picture to perceive the blossom name. To begin with, the bloom and leaves are divided. At that point, their highlights are separated to speak to the wild utilizing a bunching strategy. Their acknowledgment is accomplished utilizing a piecewise direct discriminant work (Pornpanomchai et al., 2011, p. 350)(Sai Kumar et al., 2011, p. 49). Nilsback and Zisserman mean to upgrade characterization execution on a comparable class’s enormous dataset by removing blends(Zaitoun & Aqel, 2015, p. 801) of highlights. Such blends of highlights can improve arrangement execution on an enormous dataset of comparable classes. These highlights are shading, SIFT for both the frontal area district and limit, and Histogram of Gradients. Their framework fragments blossoms picture, at that point extricates the highlights blends which are utilized as contribution to Support Vector Machine (SVM) classifier to arrange bloom picture (Abdul Aziz et al., 2019a, p. 25). Another perceiving rose framework is introduced in (Abdul Aziz et al., 2019a, p. 25)(Walowe Mwadulo, 2016, p. 396). For bloom division, a client chooses the fitting bouncing window that holds the blossom area through an intuitive interface. At that point, the blossom is portioned from the chose windows. They separate tone and shape highlights of the entire blossom district and the pistil/stamen zone to speak to the blossom highlights. At that point, they perceive blossom by looking at the distances between the info bloom picture and all bloom pictures existing in the information base to come to the most like the info blossom picture (Abdul Aziz et al., 2019a, p. 25)(Patel et al., 2018, p. 1112). A Division Phase A significant issue in a bloom order framework is the manner by which to remove the blossom locale from a characteristic complex foundation with great accuracy(Wäldchen et al., 2018) (Dutt, 2018, p. 116). Blossom locale can be divided dependent on utilizing shading highlights, since its pictures comprise of a huge green territory and bloom zone. The green region speaks to the leaves encompassing the bloom and blossom territory speaks to the bloom area which is portrayed by its tone (Jain, 2017, p. 89). Shading fragmented can be accomplished utilizing the distinction/distance between two tones. In this work, the pictures are changed to Lab shading space. At that point, a casing which incorporates the bloom is chosen by client. The Delta E is utilized to compute the distance between each pixel in the chose outline and the normal LAB tone (Jain, 2017, p. 49 ). From that point forward, OTSU limit is applied on every Lab tone to fragmented bloom. OTSU calculation attempts to locate an ideal division between classes by registering a worldwide edge for a picture (Sharma, 2017, p. 83)

4. Experimental Results

Our experimental approach has been carried out to evaluate the proposed Hybrid module of 2 types of classifiers to have comparative analysis of two kinds. The proposed system was implemented using WEKA 3.8.3 on Windows 8.1 operation system.

4.1 Dataset
The hybrid combination is evaluated using around 50423 samples which represented 3 species. The used dataset in this experiment has been taken from GitHub in which we have added additional entries via extraction of features for each and every species and created around 50423 samples after that we have used 66% of data as training data for our proposed model and remaining data which was chosen at complete random as test data.

The features which we have considered in this dataset are:
- **Sample-width**: We have taken basic measurement in which flower of each species may have
- **Sample-height**: We have taken basic measurement in which flower of each species may have
- **Petal-width**: We have taken basic measurement in which flower of each species may have
- **Petal Height**: We have taken basic measurement in which flower of each species may have
- **Specie**: Specie is our class variable or we say target variable. Target variable is always a dependant variable which depends on independent variable which are our other features which we consider as variable which play significant role in prediction of target variable

| TPR  | FPR  | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class         |
|------|------|-----------|--------|-----------|-----|----------|----------|---------------|
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000| 1.000    | 1.000    | setosa-flower |
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000| 1.000    | 1.000    | versicolor-flower |
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000| 1.000    | 1.000    | virginica-flower |
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000| 1.000    | 1.000    | Weighted Avg.  |

Table 1: Jrip Classifier class accuracy:
Table 2: J48 Classifier class accuracy:

| TPR  | FPR  | Precision | Recall | F-Measure | MCC   | ROC Area | PRC Area | Class    |
|------|------|-----------|--------|-----------|-------|----------|----------|----------|
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000 | 1.000    | 1.000    | setosa-flower |
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000 | 1.000    | 1.000    | versicolor-flower |
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000 | 1.000    | 1.000    | virginica-flower |
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000 | 1.000    | 1.000    | Weighted Avg.   |

Table 3: Naïve Bayes Classifier class accuracy:

| TPR  | FPR  | Precision | Recall | F-Measure | MCC   | ROC Area | PRC Area | Class    |
|------|------|-----------|--------|-----------|-------|----------|----------|----------|
| 1.000| 0.000| 1.000     | 1.000  | 1.000     | 1.000 | 1.000    | 1.000    | setosa-flower |
| 0.973| 0.033| 0.965     | 0.973  | 0.969     | 0.939 | 0.997    | 0.996    | versicolor-flower |
| 0.928| 0.017| 0.943     | 0.928  | 0.936     | 0.916 | 0.995    | 0.987    | virginica-flower |
| 0.970| 0.020| 0.970     | 0.970  | 0.970     | 0.951 | 0.997    | 0.995    | Weighted Avg.   |

Figure 4: J48 tree Classification output-I
Figure 5: J48 tree Classification output-II

Figure 6: J48 decision Tree
Figure 7: Jrip Classification output-I

--- Summary ---

| Correctly Classified Instances | 17144 | 100 % |
| Correctly Classifed Cases | 0 | 0 % |
| Correlation Coefficient | 1 |
| Mean absolute error | 0 |
| Root mean squared error | 0 |
| Relative absolute error | 0 % |
| Root relative squared error | 0 % |

--- Detailed Accuracy By Class ---

| Class            | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROCArea | PRC Area | Class |
|------------------|---------|---------|-----------|--------|-----------|-----|---------|----------|-------|
| setosa-flower    | 1.000   | 0.000   | 1.000     | 1.000  | 1.000     | 1.000| 1.000   | 1.000   | setosa-flower |
| versicolor-flower| 1.000   | 0.000   | 1.000     | 1.000  | 1.000     | 1.000| 1.000   | 1.000   | versicolor-flower |
| virginica-flower | 1.000   | 0.000   | 1.000     | 1.000  | 1.000     | 1.000| 1.000   | 1.000   | virginica-flower |

--- Confusion Matrix ---

| a | b | c |
|---|---|---|
| 4770 | 0 | 0 |
| 0 | 5232 | 0 |
| 0 | 0 | 4082 |

--- Weighted Avg. ---

| Class            | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROCArea | PRC Area | Class |
|------------------|---------|---------|-----------|--------|-----------|-----|---------|----------|-------|
| setosa-flower    | 1.000   | 0.000   | 1.000     | 1.000  | 1.000     | 1.000| 1.000   | 1.000   | setosa-flower |
| versicolor-flower| 1.000   | 0.000   | 1.000     | 1.000  | 1.000     | 1.000| 1.000   | 1.000   | versicolor-flower |
| virginica-flower | 1.000   | 0.000   | 1.000     | 1.000  | 1.000     | 1.000| 1.000   | 1.000   | virginica-flower |

Figure 8: Jrip Classification output-II
Figure 9: Naive Bayes Classification output-I

Figure 10: Naive Bayes Classification output-II
5 Conclusions And Future Work
A hybrid approach is proposed which is capable to identify the species of flower which is provided as data to be tested with approx. 96 present accuracy. The proposed approach incorporates three stages: division, highlight extraction, and arrangement stages. Division stage plans to improve the exactness by isolating the blossom shape from the picture normal foundation. After that with chosen classification algorithm we have applied each algorithm on prepared dataset and with the comparative analysis we have found that The tree classifier and rule based classification algorithm shows best accuracy measure in prediction of correct specie i.e. Setosa, Virginica, vericolor with the accuracy of 100%, whereas selected Naïve Bayes algorithm shows good results but in comparison of rule based and tree classifier it shows 96% accuracy. This hybrid model of combination of three algorithms will be suitable of any kind of flower specie and with some modification it can exceed its limitation to category other than flower also. This hybrid model shows nearly maximum possible accuracy on such a large dataset and with number of folds it won’t take much time in calculation of correct specie, with these characteristics of hybrid module we are proposing this Hybrid combination of algorithm for data analysis and prediction purpose for future also.

References
1. Wäldchen, J., Rzanny, M., Seeland, M., & Mäder, P. (2018). Automated plant species identification—Trends and future directions. PLOS Computational Biology, 14(4), e1005993. https://doi.org/10.1371/journal.pcbi.1005993.
2. Triguero, I., & Vens, C. (2016). Labelling strategies for hierarchical multi-label classification techniques. Pattern Recognition, 56, 170–183. https://doi.org/10.1016/j.patcog.2016.02.017
3. Nilsback Maria-Elena, A. Zisserman, in: CVPR’06 (2006): Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1447–1454.
4. Yoshioka, Y., Iwata, H., Ohsawa, R., & Ninomiya, S. (2004). Quantitative evaluation of flower colour pattern by image analysis and principal component analysis of Primula sieboldii E. Morren. Euphytica, 139(3), 179–186. https://doi.org/10.1007/s10681-004-3031-4.
5. Triguero, I., & Vens, C. (2016). Labelling strategies for hierarchical multi-label classification techniques. Pattern Recognition, 56, 170–183. https://doi.org/10.1016/j.patcog.2016.02.017.
6. Abdul Aziz, N. H., Ab. Aziz, N. A., Muhammad, B., Md Yusof, Z., Mohd Azmi, K. Z., Mohamad, M. S., & Tsuboi, Y. (2019a). A Tutorial on Population-based Simulated Kalman Filter. Mekatronika, 1(2), 23–32. https://doi.org/10.15282/mekatronika.v1i2.489.
7. Liu, R. (2014). Image Restoration Based on L1 + L1 Model. International Journal of Signal Processing, Image Processing and Pattern Recognition, 7(5), 273–286. https://doi.org/10.14257/ijisip.2014.7.5.24
8. Koul, S., & Singhana, U. (2020). Flower Species Detection Over Large Dataset Using Convolution Neural Networks and Image Processing Techniques. Volume 5 - 2020, Issue 8 - August, 5(8), 1722–1727. https://doi.org/10.38124/ijisr20aug006
9. Mukane, S. M. (2013). Flower Classification Using Neural Network Based Image Processing. IOSR Journal of Electronics and Communication Engineering, 7(3), 80–85. https://doi.org/10.9790/2834-0738085
10. Patel, I., Patel, S., & Patel, A. (2018). Analysis of Various Image Preprocessing Techniques for Denoising of Flower Images. International Journal of Computer Sciences and Engineering, 6(5), 1111–1117. https://doi.org/10.26438/ijcse/v6i5.11111117
11. Jain A.(2017). Detection of Malicious Executables Using Rule Based Classification Algorithm, Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 35–38 DOI: 10.15439/2018KM04 ISSN 2300-5963 ACSIS, Vol. 14”.
12. Jain A .(2017), “Clustered Comparative Analysis of Security Sensor Discrimination Data, Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 29–33DOI: 10.15439/2018KM21 ISSN 2300-5963 ACSIS, Vol. 14.
13. Sharma, S. (2017). A Review on Cloud Computing Security and Privacy in Service Oriented Architecture (SOA). International Journal of Machine Learning and Networked Collaborative Engineering, 01(02), 81–92. https://doi.org/10.30991/ijmlnce.2017v01i02.004.
14. Dutt, V. (2018). USAGE OF THE BIG DATA IDEA IN ASSOCIATIONS POTENTIAL OUTCOMES, OBSTRUCTIONS AND DIFFICULTIES. International Journal of Machine Learning and Networked Collaborative Engineering, 02(01), 114–128. https://doi.org/10.30991/ijmlnce.2018v02i01.005
15. S. Sarkate, R., & B. Khanale, P. (2014). Domain Specific Knowledge based Machine Learning for Flower Classification using Soft Computing. International Journal of Computer Applications, 104(1), 14–17. https://doi.org/10.5120/18166-9027
16. Chitra, H. S. H., Suguna, S., & Sujatha, S. N. (2016). A Survey on Image Analysis Techniques in Agricultural Product. Indian Journal of Science and Technology, 9(12), 14–28. https://doi.org/10.17485/ijst/2016/v9i12/77171
17. Flower Recognition System Based on Image Segmentation Algorithms. (2020). International Journal of Innovative Technology and Exploring Engineering, 9(3), 685–688. https://doi.org/10.35940/ijitee.c8374.019320
18. Najjar, A., & Zagrouba, E. (2016). Adaptive descriptor based on the geometric consistency of local image features: application to flower image classification. Journal of Electronic Imaging, 25(5), 053023. https://doi.org/10.1117/1.jei.25.5.053023
19. Zaitoun, N. M., & Aqel, M. J. (2015). Survey on Image Segmentation Techniques. Procedia Computer Science, 65, 797–806. https://doi.org/10.1016/j.procs.2015.09.027
20. Abdul Aziz, N. H., Ab. Aziz, N. A., Muhammad, B., Md Yusof, Z., Mohd Azmi, K. Z., Mohamad, M. S., & Tsuboi, Y. (2019a). A Tutorial on Population-based Simulated Kalman Filter. Mekatronika, 1(2), 23–32. https://doi.org/10.15282/mekatronika.v1i2.4894
21. Samina Khalid, Tehmina Khalil, Shamila Nasreen, “A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning”, Science and Information Conference, London, UK, August 27-29, 2014, pp. 1-8.
22. Walowe Mwadulo, M. (2016). A Review on Feature Selection Methods For Classification Tasks. International Journal of Computer Applications Technology and Research, 5(6), 395–402. https://doi.org/10.7753/ijcatr0506.1013
23. Inthiyaz, S., Madhav, B. T. P., Kishore, P. V. V., & Keedwell, E. (2017). Flower segmentation with level sets evolution controlled by colour, texture and shape features. Cogent Engineering, 4(1), 1323572. https://doi.org/10.1080/23311916.2017.1323572
24. Mohanty, A. K., & Bag, A. (2017). Image Mining for Flower Classification by Genetic Association Rule Mining Using GLCM features. International Journal of Advanced Engineering, Management and Science, 3(5), 516–521. https://doi.org/10.24001/ijaems.3.5.17.
25. Flower Identification and Classification using Computer Vision and Machine Learning Techniques. (2019). International Journal of Engineering and Advanced Technology, 8(6), 277–285. https://doi.org/10.35940/ijeat.e7555.088619.
26. Sai Kumar, T., Chandra M, M., & Murthy P, S. (2011). COLOUR BASED IMAGE SEGMENTATION USING FUZZY C-MEANS CLUSTERING. International Journal on Intelligent Electronic Systems, 5(2), 47–51. https://doi.org/10.18000/ijies.30099
27. Pornpanomchai, C., Pitchayuk Supapattranon, C. K., & Siriwisesokul, N. (2011). Leaf and Flower Recognition System (e-Botanist). International Journal of Engineering and Technology, 3(4), 347–351. https://doi.org/10.7763/ijet.2011.v3.251

28. Jindal, M., Bajal, E., Chakraborty, A., Singh, P., Diwakar, M., & Kumar, N. (2020). A novel multi-focus image fusion paradigm: A hybrid approach. Materials Today: Proceedings.

29. Dhaundiyal, R., Tripathi, A., Joshi, K., Diwakar, M., & Singh, P. (2020, April). Clustering based Multi-modality Medical Image Fusion. In Journal of Physics: Conference Series (Vol. 1478, No. 1, p. 012024). IOP Publishing.

30. Singh, P., & Shree, R. (2017). A New Computationally Improved Homomorphic Despeckling Technique of SAR Images. International Journal of Advanced Research in Computer Science, 8(3).

31. Kumar, P., & Diwakar, M. A novel approach for multimodality medical image fusion over secure environment. Transactions on Emerging Telecommunications Technologies, e3985.

32. Tyagi, T., Gupta, P., & Singh, P. (2020, January). A Hybrid Multi-focus Image Fusion Technique using SWT and PCA. In 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 491-497). IEEE.

33. Kumar, M., & Diwakar, M. (2019). A new exponentially directional weighted function based CT image denoising using total variation. Journal of King Saud University-Computer and Information Sciences, 31(1), 113-124.

34. Singh, P., & Shankar, A. (2021). A novel optical image denoising technique using convolutional neural network and anisotropic diffusion for real-time surveillance applications. Journal of Real-Time Image Processing, 1-18.

35. Singh, P., Diwakar, M., Shankar, A., Shree, R., & Kumar, M. (2021). A Review on SAR Image and its Despeckling. Archives of Computational Methods in Engineering, 1-21.