Comparative Study of Classification Performance Accuracies of Machine Learning Algorithms

Amrita Bhatnagar[1], Kavita Bisht[2], Riya Srivastava[3], Bhawna Mehra[4], Raj Prakash Mittal[5]

[1-5]*Computer Science Department, Inderprastha Engineering College, Ghaziabad, India*

---***---

Abstract- *In this paper, we will explore and try to solve the bird species classification problem. We have use the publicly available bird datasets, which provides us with the bird images to classify the species. The difficulty of this task arises from the inter-class similarities of species of birds that often fool even the expert bird watchers. This dataset has been widely used to develop approaches towards fine-grained classification, leaving us with the task of understanding how geometry in the datasets relates to such classification algorithms. Fine-grained classification often describes an end-to-end pipeline, from image to class prediction. The dataset, however, provide labels for image's classes of various breeds of birds. In our paper, we have use the supervised learning algorithm of Machine learning wherein we have make use of the labelled dataset to decide the class or breed of the bird. This supervised learning algorithm try to model relationships and dependencies between the target prediction output and the input features such that we can predict the output values for new data based on those relationships which it learned from the previous data sets.*

1. INTRODUCTION

BIRD behavior and population trends have become an important issue now a day. In our daily life we see different kinds of birds such as parrots, sparrows, etc... It is really a hectic task to figure out the species of the bird on the first sight. Even professional photographers/bird watchers also get confused while classifying the birds based on their species. Classification of birds by their species is valuable for biological research and environmental monitoring applications, especially in detecting and locating them. The simple way for the researches to determine the ram of human activities on birds is identify them and maintain a count of them in the specific location.

In such case, a reliable system that will provide large scale processing of information about birds and will serve as a valuable tool for researchers, governmental agencies, etc. is required. So, bird species identification/classifier plays an important role in identifying that a particular image of bird belongs to which species. Bird species identification means predicting the bird species belongs to which category by using an image. Here we build a deep learning model for classifying different birds based on the mentioned features. In our dataset we have 312 different attributes that differentiates a bird from others. We have used different algorithms for classifying such as KNN, SVM, Naive Bayes, LDA.

Our aim is to produce a best model that can accurately classify the bird species based on their divergent features i.e. by considering a collection of different attributes that describes the bird characteristics. Some of the characteristics for classifying birds are their color, beaks, wings, eye size, tail, etc. If we are successful enough on making that best model then this model can then be very useful for those ornithologists who are facing problems in bird species identification. Ornithologists require studying all the details of birds such as their existence in environment, their biology, their distribution, their ecological impact, etc. which is itself a very difficult task.

This model gives the accuracy for these algorithms by classifying them based on training and testing. An accuracy plot is drawn for every algorithm and at the end, the algorithm that classifies the birds with greater accuracy is determined. A final plot is shown which separates different algorithms form each other based on the accuracy.

These systems/models will also help the researchers to identify the species which are endangered or are extinct or are going to extinct. In this project, our main focus is to classify the birds from high resolution photographs taken from the CUB-200 Data set.

2. EXISTING SYSTEM

The deep learning technology has shown impressive performance in various vision tasks such as image classification, object detection and semantic segmentation. In particular, recent advances of deep learning techniques bring encouraging performance to fine-grained image classification which aims to distinguish subordinate-level categories, such as bird species.

The core of the deep learning technology is that the layers of the features are not designed by human engineers and instead learned from data using a general-purpose learning procedure. This task is extremely challenging due to high intra-class and low inter-class variance.

3. ALGORITHM USED

3.1 Classification using Naïve Bayes

Every pair of features being classified is independent of each other. Naive Bayes is a conditional probability model. Naive Bayes classifier is based on Bayes Theorem. Bayes Theorem finds the probability of an event occurring given the probability of another event that has already occurred.
Equation of Bayes theorem:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Figure 1. Probability function for Naive Bayes

Here, we try to find the probability of an event A, given the event B is true. Event B is termed as evidence

P(A) is the probability of the event before the evidence is seen.

P(A/B) is posterior probability of B, i.e. probability of event after evidence is seen.

3.1.1 Naive Assumption

Now we apply the Naïve assumption on Bayes theorem to obtain the Naive Bayes theorem, which is, independence among the features. We now split the evidence into independent parts. If events are independent among each other, we get the result as

$$P(y|x_1, ..., x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1)P(x_2)...P(x_n)}$$

Figure 2. Naive assumption for n events which can be expressed as

$$P(y|x_1, ..., x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

Figure 3. Simplified function for Naive assumption

where P(y) is called class probability and P(Xi/y) is called conditional probability. Variations or assumptions made to the distribution of P (xi / y) gives us the possibility to create different Naive Bayes classifiers. Now we apply these formulae manually on our bird species dataset.

3.1.2 Gaussian Naive Bayes classifier

Here, the continuous values associated with each feature are assumed to be distributed according to gaussian distribution. Gaussian distribution is also called normal distribution. The shape of normal distribution curve is bell-shaped, and which is symmetric about the mean of the feature values.

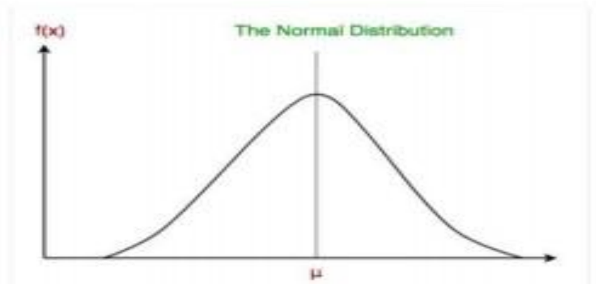


Figure 4. Normal distribution curve for Gaussian Naive Bayes classifier

Conditional probability is given by

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

Figure 5. Conditional probability function for Gaussian Naive Bayes classifier

3.2 K-Nearest Neighbour (K-NN)

Nearest-neighbor classifiers are built on training by metaphor, i.e. a testing tuple is compared with a given training tuple that are identical to it. The training tuples are described by n attributes. Every tuple is represented by a point in an n-dimensional space. Thus, every training tuple is stored in an n-dimensional pattern space.. When an unknown tuple is given, the K-NN classifier searches for the pattern space for K training tuples that are near to the unknown tuple. These k training tuples are the k “nearest neighbors” of the unknown tuple. When an large training sets are given K- Nearest Neighbor is labor intensive. It is widely used in the field of pattern recognition. In addition to this it also has many applications in intrusion detection and data mining. Let n be the training data samples and p be the unknown point. – we store n training samples in an array a []. Each element of an array represents a training tuple (x, y), – Now calculate Euclidean distance for every training data sample. Euclidean distance (a [], p).

Euclidean distance is calculated by:

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2}.$$

Figure 6. Euclidean distance for calculating nearest tuple value

The smallest distance of k is put into S.Return the majority labels among S. We are given some training data, it classifies the classes into attributes of different groups. Consider the below figure which consists of two different classes. Now we are given with another set of classes i.e. testing data and the points are allocated by the given testing data. The unclassified tuples are uncolored. Nearest-neighbor classifiers are very slow during the classification of testing tuples. Let is assume D as the training database of |D| tuples and k = 1, then number of comparisons required to classify the given testing tuple are O(|D|). The number of comparisons can be reduced to O(log(|D|)) by sorting the tuples into search trees. The running time can be reduced to O (1) by parallel implementation, independent of |D|. Other techniques to speed up classification time include the use of partial distances and editing the storing tuples. In editing method, the training tuples that are proven useless are removed. Editing method is also called as pruning or condensing. In partial distances the distance between the n-tuples is calculated, if the value exceeds a threshold, the computation is stopped, and it is returned to the next step.

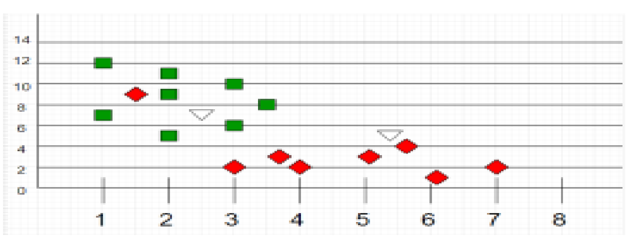


Figure 7. Represents two different classes

3.3 SVM

Support Vector Machine(SVM) is a Machine Learning algorithm that can be used for both regression and classification. Support Vector Machines mainly consists of decision planes which define decision boundaries. A decision plane separates different objects of different classes.

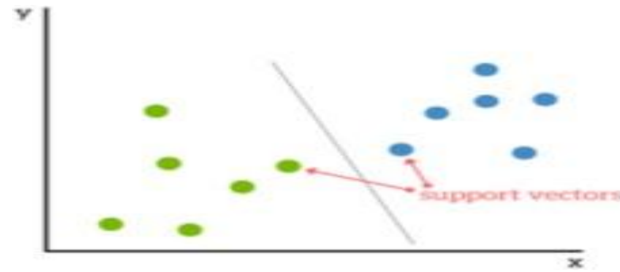


Figure 9. Represents support vectors.

The data points which are closer to the decision plane are called support vectors. A decision plane linearly separates the data. It also classifies the set of data and divides it into different classes. If the data points are far away from the plane, then they are considered as correctly classified. So, we must take care that the data points are as far as away from the decision plane. The decision plane decides the class of the data points.

Figure 11. Represents Tuples randomly arranged without decision plane

If there is no decision plane the dataset will be jumbled representing a non-linearly separable dataset. Thus, SVM considers decision planes and decision boundaries in order to separate the dataset linearly.

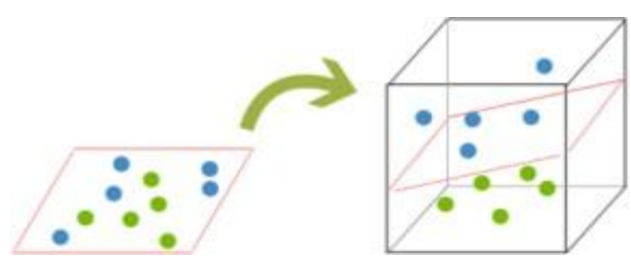


Figure 12. Represents 3-D view of tuples and decision planes in space

To classify the dataset a 3-D view is necessary, which is very simple to explain. Imagine that our data points are randomly arranged on the sheets and these sheets are raised all at a sudden moving the balls into the air. When the data points are in the air we use the sheet to separate them. The lifting of the balls means the mapping of the data into a higher dimension. The process of lifting the data points in known as kernelling.

3.4 LDA

LDA–Latent Dirichlet allocation. Suppose you are given with the following sentences:

- I consume water and food
- Birds do fly
- Birds drink water

The main function of LDA is that it automatically detects topics which are of these type from the document. When we give these kinds of sentences, LDA classifies the different colored words into different topics. For example, when given the above sentences LDA may classify the cyan words under the topic of C, that are labeled as Consume. In the same way red words are also classified under a different topic B, that are labeled as Birds. In this way LDA labels every word and categorizes them into different topics. The advantages of defining words under different topics are

We can assume the content expansion of every sentence by counting the words Sentence 1: 100% Topic C. Sentence 2: 100% Topic B. Sentence 3: 50% Topic C and 50% Topic B.
Each word can be described by determining the proportions of each word from the given sentences. If we consider the above example, C is comprised of following proportions- 30% water, 30% food, and 40% consume. And B is comprised of following proportions- 50% birds and 50% fly.

3.5 Random Forests

Random forests are one of the ensemble methods. Assume that all the classifiers in the ensemble is a decision tree classifier so that the group can be called a „forest". At each node, random selection of attributes or labels is done to generate individual decision trees and to find the split. While classifying the tuples into appropriate classes, each tree votes and the majority class is returned.

Decision tree can be constructed in 2 different ways:

- While constructing the decision tree classifier, if the selection of labels at each node is done randomly to determine the split, then the trees formed this way with random input selection is called Forest-RI.
- Another type of Random Forest, is Forest-RC which determines the split at each node by choosing attributes that are formed by linearly combining the existing labels.

Random forests are built by aggregating N decision trees

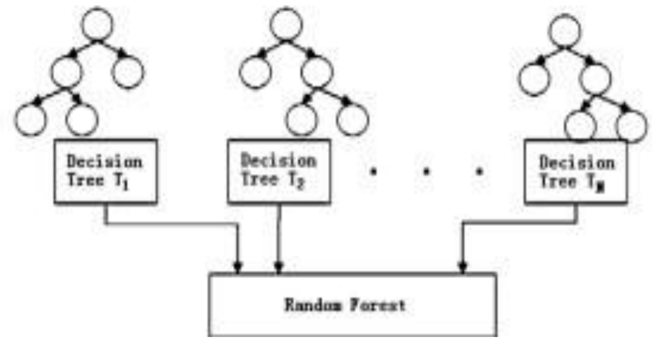


Figure 13. An image showing how a Random Forest is constructed

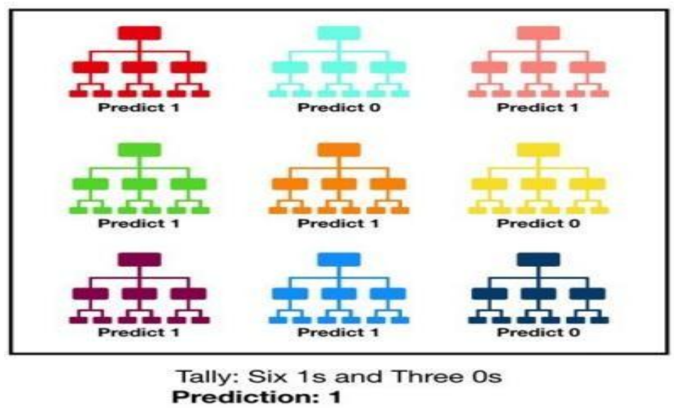


Figure 13. Prediction Diagram

3.6 Classification and regression tree (CART)

It is a term used to describe decision tree algorithms that are used for classification and regression learning tasks. Where decision tree algorithms are nothing but if-else statements that can be used to predict the results based on data. It consists of two types of decision trees: - classification(where target variable is fixed or categorical is then used to identify the "class" within which a target variable would most likely fall.) and regression (where there is a target variable and algorithm needs to find that value)

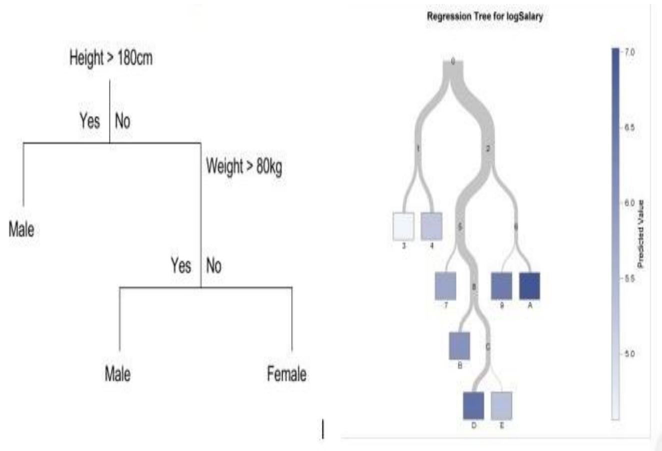


Figure 14. Classification and Regression Tree diagram

4. Results & Conclusion

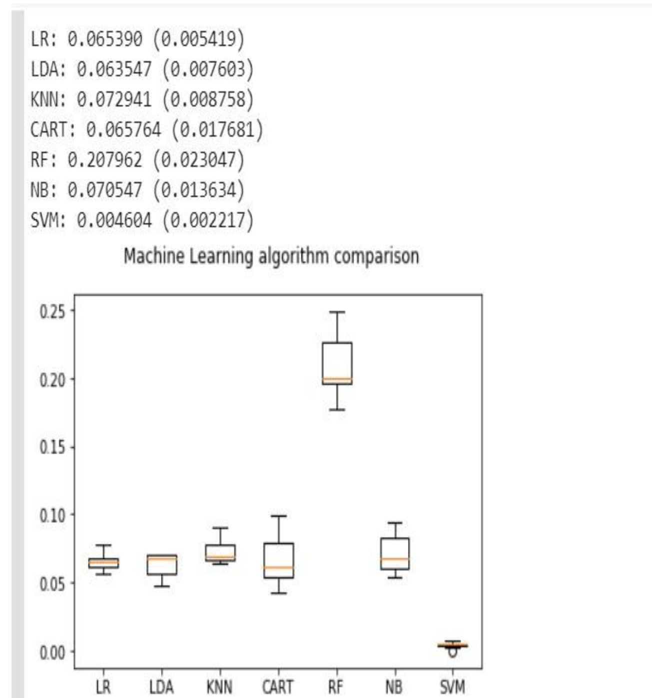
As we can see in the box plot figure, the accuracies are **not** so good. Random Forests (RF) gives the maximum accuracy of **20.7%**. This is mainly due to the number of images we use per class. We need large amounts of data to get better accuracy.

Also, some other reasons for this inaccuracies includes that, this dataset was challenging to deal with as this dataset has images of varying sizes and sometimes the resolution of the images lies in between 800x600 to 6000x4000 due to which it become difficult to process the image and convert it into required vector size.

Also, the training dataset mostly contains bird images in which bird were almost 10–20% of the whole image whereas in case of test images the bird contains 70–80% of the image. Sometimes, the model fails to detect the birds due to less number of birds in the dataset. In some classes, the birds cover not even 10% of the whole images or the color of bird and surrounding are very similar. Cases where birds are brown in color. In those cases, the model fails to localize birds due to occlusion problem or background similarity problems.

Finally, we train each of our machine learning model and check the cross- validation results. Here, we have used only our training data which further means that these algorithms are unable to train fully on the given images and that's why now-a-days deep learning algorithms are more in use for image-based detection rather than machine learning algorithms due to their better performance in feature extraction from the images.

Also, Classical ML algorithms often require complex feature engineering. Usually, a deep dive exploratory data analysis is first performed on the dataset. A dimensionality reduction might then be done for easier processing. Finally, the best features must be "carefully" selected to pass over to the ML algorithm. There's no need for this when using a deep network as one can just pass the data directly to the network and usually achieve good performance right off the bat. This totally eliminates the big and challenging feature engineering stage of the whole process. So, this suggest to use deep learning over classical machine learning algorithms more.



5. References

[1] bo zhao1,2 jiashi feng2 xiao wu1 shuicheng yan2 1school of information science and technology, southwest jiaotong university, chengdu 613000, china 2department of electrical and computer engineering, national university of singapore, 117583, singapore. https://link.springer.com/article/10.1007/s11633-017-1053 -3

[2] megvii research nanjing, megvii technology, nanjing, china 2national key laboratory for novel software technology, nanjing university, nanjing, china 3graduate school of ips, waseda university, fukuoka, japan. https://arxiv.org/pdf/1907.03069.pdf.

[3] steve branson et al. "bird species categorization using pose normalized deep convolu- tional nets". in: corr abs/1406.2952 (2014).

[4] f. pedregosa et al. “scikit-learn: machine learning in python”. in: journal of machine learning research 12 (2011), pp. 2825–2830.

[5] wah et al. the caltech-ucsd birds-200-2011 dataset. tech. rep. cns-tr- 2011-001. california institute of technology, (2011).

[6] h.kopka and p.w. daly, a guide to latex, 3rd ed harlow, england: addison-wesley, 1999.

[7] j. han, m. kamber and j. pei, data mining, 3rd ed. amsterdam: elsevier/morgan kaufmann, 2012.

[8] Shai Shalev-Shwartz and Shai Ben-David (2014). Understanding Machine Learning From Theory to Algorithms.

[9] Taiwo, O. A. (2010). Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-034-6, InTech, University of Portsmouth United Kingdom. Pp 3 – 31. Available at InTech open website: http://www.intechopen.com/books/new-advances-inmachine-learning/types-of-machine-learning-algorithms

[10] Tapas Kanungo, D. M. (2002). A local search approximation algorithm for k-means clustering. Proceedings of the eighteenth annual symposium on Computational geometry. Barcelona, Spain: ACM Press.

[11] Cheng, J., Greiner, R., Kelly, J., Bell, D.& Liu, W. (2002). Learning Bayesian networks from data: An informationtheory based approach. Artificial Intelligence Volume 137. Copyright © 2002. Published by Elsevier Science B.V. All rights reserved pp. 43 – 90. Available at science Direct: http://www.sciencedirect.com/science/article/pii/S0 0043702 02001911

[12] Hormozi, H., Hormozi, E. & Nohooji, H. R. (2012). The Classification of the Applicable Machine Learning Methods in Robot Manipulators. International Journal of Machine Learning and Computing (IJMLC), Vol. 2, No. 5, 2012 doi: 10.7763/IJMLC.2012.V2.189pp. 560 – 563. Available at IJMLC website: http://www.ijmlc.org/papers/189-C00244-001.pdf

[13] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica 31 (2007). Pp. 249 – 268. Retrieved from IJS website: http://wen.ijs.si/ojs2.4.3/index.php/informatica/article/download/148/140.

[14] Logistic Regression pp. 223 – 237. Available at: https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12 .pdf

[15] Musa, A.B. Comparative study on classification performance between support vector machine and logistic regression. Int. J. Mach. Learn. & Cyber. **4,** 13–24 (2013). https://doi.org/10.1007/s13042-012-0068-x

BIOGRAPHIES

Amrita Bhatnagar
Assistant Professor
Computer Science Dept
Inderprastha Engineering College

Kavita Bisht
Final Year Student
Computer Science Dept
Inderprastha Engineering College

Bhawna Mehra
Final Year Student
Computer Science Dept
Inderprastha Engineering College

Raj Prakash Mittal
Final Year Student
Computer Science Dept
Inderprastha Engineering College

Riya Srivastava
Final Year Student
Computer Science Dept
Inderprastha Engineering College